SigTur/E-Destination: A System for the Management of Complex Tourist Regions

Joan Borràs\textsuperscript{c}, Joan de la Flor\textsuperscript{d}, Yolanda Pérez\textsuperscript{a}, Antonio Moreno\textsuperscript{b}, Aida Valls\textsuperscript{b}, David Isern\textsuperscript{b}, Alicia Orellana\textsuperscript{c}, Antonio Russo\textsuperscript{a}, and Salvador Anton-Clavé\textsuperscript{a}

\textsuperscript{a}Universitat Rovira i Virgili, Departament de Geografia, Tarragona, Spain
\textsuperscript{b}Universitat Rovira i Virgili, Departament d'Enginyeria Informàtica i Matemàtiques, Tarragona, Spain
\textsuperscript{c}Science and Technology Park for Tourism and Leisure, Vila-Seca, Spain

\{joan.borras, joan.delaflor, alicia.orellana\}@pct-turisme.cat
\{myolanda.perez, antonio.moreno, aida.valls, david.isern, antonio.russo, salvador.anton\}@urv.cat

Abstract

The development of digital and web technologies opened new horizons regarding the generation of personalized contents and the management of visitor services. The objective of SIGTur/E-Destination is to provide to the tourist and travel sector electronic tools for the sustainable management of destinations, also possibly leading to an increase in the effectiveness of visitor service firms. The SIGTur/E-Destination system includes a warehouse of geo-referenced information that updates the available information on the tourism activities in the Costa Daurada and Terres de l'Ebre in the south of Catalonia (Spain). This catalogue of information on resources, attractions, products, establishments and packages has to make possible the improvement of the interaction of between public and private agents and visitors. A recommender system uses this catalogue to offer personalized information to the tourists.

Keywords: Travel recommender systems; geographic information systems; Artificial Intelligence

1 Background

The spatial behaviour of visitors into a destination, and specifically their selection of visits once in a destination; the routing between them and the selection of means of transport; the time and attention dedicated to each of them; purchases; the choice of “gateways” and nodal points where to rest or stay overnight is strongly influenced by time budgets (Shoval & Isaacson, 2009; Oppermann, 1995; McKercher & Lau 2008). These factors, among others, come to determine the pace of development of a destination (Shoval, 2000), the satisfaction of visitors (Verbeke & Lievois, 2004), and the structure of the market (Russo, 2002).

Innovative solutions to this problem rely in planning and management initiatives that minimise the information deficiencies and asymmetries influencing the spatial behaviours of visitors. When tourists have full information about the location, access,
and quality of the attractions and complementary services in a destination region, they are more likely to organise their stay in a way that matches their preference schemes. On the other hand, the clever design, organisation and communication of opportunities in the region may bring to a more balanced tourism activity, spatially, thematically and financially, with important returns in terms of sustainable development.

However, once accrued this knowledge, it is then necessary that destination management organisations provide the necessary infrastructure to facilitate the spatial activity of visitors. For this reason, recommender systems are emerging as important elements in the development and management strategies of destination regions and cities, with increasing degrees of sophistication. The widespread usage of web environments in travel planning, as well as the consolidation of web 2.0 technologies, is opening the way to a new generation of tourism recommender systems with the potential to enrich tourism experiences to a substantial degree. This paper provides an illustration of the design and potential effects of one such recommender system of the “new generation”, to be applied in the management of a very complex destination region: Costa Daurada, in the south of Catalonia, one of the most developed coastal destinations in Spain (Anton-Clavé, 2010). In this region mostly focusing on “3S” tourism; tourism activity is concentrated in a narrow stretch of coastal resorts (as is shown by various studies on tourists’ spatial behaviour in this area). Diversification and differentiation for a renewed sustainable development clearly passes through an “activation” of natural, cultural and eno-gastronomic resources found in inland areas and cities.

In a partnership with the major tourism operators and DMOs of the area, and using European Regional Development co-funding, the Science and Technology Park of Tourism and Leisure of Vila-seca, Tarragona, has designed a recommender system, SIGTUR/E-Destination, that makes the whole range of products and itineraries accessible to visitors that plan their visits in this area, as well as to those that being already there want to enjoy richer experiences. It is based on the interaction of geospatial technologies and Artificial Intelligence algorithms, adding higher-level functionalities to currently available recommender systems and providing users with a greater range of possibilities to identify leisure activities according to their profile and, beyond that, facilitate the planning of the trip and the decision-making process before and during the stay. In the remainder of this paper we proceed as follows: in the next section the main functionalities of the system are described. Next, the architecture and design of the system are detailed, emphasizing the geospatial and recommender features. We conclude with final considerations and guidelines for future research.

2 Functionalities of the System

Tourism recommender systems allow users to optimise their travel planning time, receiving personalized assistance. In order to achieve such personalization, semantic web technologies may be used (Berners-Lee et al., 2001). Semantic web techniques
using ontologies of the tourism domain are useful in the development of new intelligent systems. In addition, recommendations can be enhanced with algorithms that learn user profiles (Sieg et al., 2007), as well as with the traditional content and collaborative filters. However, most recommender systems have some important limitations: they are focused on accommodation facilities, or feed on recommendations from the users themselves, or do not take into account spatial parameters, such as the location of the tourist resources and where tourists stay.

On one hand, the usage of geospatial technologies allows the recommender system to display geo-referred resources on a web map-based application, filtering results on real time according to spatial parameters, such as position or distance. On the other hand, Artificial Intelligence provides the system with the capability to learn automatically from the users' behaviour, meaning that the application is able to refine the recommendations on-the-fly and improve them for future visitors, without requiring explicit collaboration from users.

![PCT SIGTUR/e-Destination](image)

**Fig. 1.** Fill-in form to initialize the user profile

This system provides a particularly powerful tool for the spatial management of tourism activity in a complex tourist region counting on a wide range of tourist products and territorial features dispersed in space. In such a context, the ultimate objective of tourism policy would be that of diversifying tourist activity by presenting alternatives that are not so easy to visualize and access as the core product.

The users are first presented with a brief questionnaire that is used to know the general preferences of the visitor, his/her demographic characteristics and the localization. This information is processed by the recommender system (as will be explained in the next section), which matches these values with the descriptions of the available activities in the area of interest. The system performs a selection of the most appropriate activities and displays a list ordered by the probability to satisfy the user's
preferences. In order to introduce a diversification mechanism, the system computes the similarity of recommended activities, rejecting those that are too similar. In addition, this procedure provides also a "surprise" factor in the recommendation. Although this decision reduces the accuracy of the process, it is estimated to improve user satisfaction (Ziegler et al., 2005).

Fig. 2. Map-based recommendations of cultural activities

Since activities in the application are geo-referenced, recommendations can be spatially dependent. That is, the user is able to specify his/her preferences regarding the location of an overnight stay in Costa Daurada and Terres de l’Ebre region, and the radius of distance that he/she is willing to move around during his/her stay. Then the distance term will also affect the final result of the recommendation. In addition, when the user’s decision of destination is still unknown, the system is able to recommend those destinations that fit better with the user’s preferences, taking into account the location of the activities. Figure 2 gives a representation of recommended activities in the proximity of Tarragona.

After the recommendation of the activities, the user is able to interact with the displayed items. He/she can request detailed information of a certain activity, as well as add it to a travel planner or express if the activity is liked or disliked. Moreover, the user can request other activities that are similar to the selected one. The implicit knowledge provided by these feedback actions of the user is taken into account in the following recommendations made to that user, so that new activities will be considered in the proximity of the specified destination matching the current map view.
3 Architecture and Design of the System

The system needs to manage information efficiently in order to achieve multiple user objectives simultaneously. Thus an adequate architecture must be defined. Figure 3 illustrates the general architecture of the system. All modules are based on Open Source technologies. The core of the architecture is the recommender system, developed in Java, which provides the interaction between modules. The web server application has been built in Java Server Faces (https://javaserverfaces.dev.java.net) with ICEfaces extension which is an Ajax framework that allows the development of rich Internet applications in Java. Since the application does not need to work with push technologies by now, the asynchronous mode is deactivated due to its additional resource requirements. Thus, the client presentation update is done synchronously by the request/response cycle. Concurrency of the system is addressed by the framework using a thread pool which provides bounded thread usage in large-scale applications. The web map server interacts with the client cartographic API, to display dynamic maps in a web.

![Diagram of the SIGtur/E-Destination system](image)

Data storage is divided in two databases, one that contains the tourist resources and another that stores the user profiles. User data are managed by PostgreSQL (http://www.postgresql.org/) and tourist resources are stored in the PostGIS (http://postgis.refractions.net/) extension. Database connections are handled by the Hibernate framework (http://www.hibernate.org) with a spatial extension that handles geographic data. Some modifications have been applied to Hibernate that improve the pool of database connections. In order to process spatial functions over tourist resources, such as computing the distance between two destinations, the JTS Topology Suite API (http://www.vividsolutions.com/jts/) is used.

3.1 Geospatial features and GIS database

As seen in Figure 3, there are different geospatial elements that take part on the SIGTUR/E-Destination system. This section details the role of each geospatial feature within the system.
Geoserver (http://geoserver.org/) is an open-source server written in Java for publishing spatial data from any source using the Open Geospatial Consortium standards (http://www.opengeospatial.org/) and building web map-based applications. In the case of the SIGTUR/E-Destination system, Geoserver has been configured to be connected to the GIS database in order to publish the spatial data from tourist resources as separate layers on the map. This server has a user friendly interface to define the publishing parameters for each layer before being displayed on the map. OpenLayers (http://openlayers.org/) is an open-source Javascript library that provides an API to obtain base maps from external sources, such as Google, Yahoo or Bing Maps. This API can reproject maps on-the-fly, enabling an easy overlay of spatial data from different sources and projections. In the initial stages of the project we considered the options of using topographic base maps, served by the Catalan spatial data infrastructure (http://www.geoportal-idec.cat/geoportal/eng/) or using the most recent and popular web-mapping platforms, such as Google Maps. Although this second alternative is much less accurate and subject to cartographic errors, the final decision was to use it, since the potential user is a tourist without a strong technical profile, who is more familiar with Google Maps mash-up applications. PostGIS is a spatial database extension for PostgreSQL that enables PostgreSQL to be used as a backend spatial database for Geographical Information Systems (GIS).

A GIS is defined as an information system used to input, store, retrieve, manipulate, analyze, and output geographically referenced data or geospatial data in order to support decision making for planning and management (Goossen et al., 2009). In this sense, the GIS database of the SIGTUR/E-Destination system stores the tourist resources (activities and services) from the region of Costa Daurada and Terres de l’Ebre to be used by the recommender system. It has been designed following the same structure and categories of the domain ontology, in order to enhance the system consistency. Consequently, each table - that represents a map layer - of the GIS database corresponds to a node at the highest level of the hierarchy stored in the domain ontology. Currently, the GIS database contains the main tourist resources of the region (over six hundred so far). Nevertheless, there is still a considerable work on adding new resources and updating the existing ones. Anyway, the GIS database has been designed in order to support easily these future tasks related to additions and updates. Besides, the structure of the tables is as similar as possible, containing most of the same fields, which facilitates the management of the database and massive operations. Just in the cases in which tourist resources require specific information additional fields have been added. In order to permit the interaction between the GIS database and the domain ontology each table in the GIS database includes a field that stores the tags with the visitor activity specifications, enabling the system to yield recommendations properly. In addition, the database contains a table that defines the relationships between all tourist activities, as a result of geo-processing operations between layers, so that the system can recommend tourist activities based on similarity.

Regarding the spatial data, they have been pooled from a wide variety of sources, what means that an effort has been required to convert different formats (Spreadsheets, CAD, shapefiles, etc) and in some cases geo-coding them. In the cases
in which spatial data on tourist resources were not available, they have been generated by the authors. Once the data have been homogenized and categorized according to the different activities, they have been uploaded to the GIS database, making them ready to use by the recommender system. It must be said that the use of a GIS software platform has facilitated all these operations.

3.2 Recommender system

In the development of recommender systems it is usual to employ hybrid methods that combine content-based and collaborative techniques, in order to overcome the main drawbacks of the individual methods and thus improve the accuracy and user satisfaction of the recommendation. Therefore, a hybrid recommender system is proposed, which uses two types of techniques: content filters and collaborative recommendations. One of the main features of this system is that the proposed methods are based on semantic domain knowledge represented in an ontology. Ontologies define areas of common understanding between multiple actors, easing their interoperability and permitting a high-level communication (Berners-Lee et al., 2001). Ontologies include concepts and relationships between them. Incorporating semantic information to the recommender systems (Gawinecki et al., 2005; Lee et al., 2009) normally overcomes the main problems of traditional methods. Such information provides to the system reasoning capabilities and the possibility of improving the results of Machine Learning algorithms (Mobasher, 2007). The ontology has been constructed by a team of tourism experts to model the knowledge about visitor activities in a sufficiently general way as to permit repetitions of its use. It is also quite easily extendable to add new activities if required.

3.3 Tourism domain ontology

A large ontology in the tourism domain has been built in order to describe the tourist activities in a hierarchy. The ontology represents up to 184 connected concepts and five hierarchy levels in the tourism domain with multiple ancestors. Figure 4 illustrates a portion of the ontology. The ontology is used to explicitly classify the activities to recommend within a core set of predefined concepts, providing an explicit meaning that is used by the intelligent system. The main objective of the ontology is to model the degree of interest of the user for each concept. This degree is evaluated by explicit and implicit information, as well as through the collaboration with other users. Once the user's interest of each node is well defined, the degree of interest of each activity is computed, given that these activities are associated via at least one concept in the hierarchy. As it will be seen, this approach overcomes the main drawbacks suffered by the non-semantic methods.

The domain ontology has been developed with the Protégé editor (http://protege.stanford.edu) using the OWL language (http://www.w3.org/TR/owl-features). Jena (http://jena.sourceforge.net) is the semantic web framework used in the system that provides a management and a rule-based inference engine for ontologies. For each user, the ontology is loaded from the file as memory model, since it is not required to model persistent ontologies and it is the most efficient way to compute inference rules. However, the semantic information related to each user needs to be
stored in a database. Therefore, the ontology concepts are previously migrated from the ontology to the user database without requiring their hierarchy level and node connections, thus reducing memory resources.

![Diagram of tourism domain ontology]

**Fig. 4.** Portion of the tourism domain ontology

### 3.4 Content filters

Content filters (Pazzani & Bilnius, 2007) are a kind of systems based on a direct matching between the characteristics of the activities to be recommended and the interests of the user in each of those features. To perform this kind of recommendation, it is required to build a user profile that stores the degree of interest (i.e. a score) on each of the different criteria that describe an activity. Such information can be extracted by fill-in forms but, as the set of characteristics can be quite large, it is known that this process is not adequate because long questionnaires result in inconveniences for participants in a survey.

Our goal was to develop a system that may consider many different types of activities with a large set of distinct features. The domain ontology presented above (Figure 4) allows the system to know the taxonomic and semantic relations between the different elements: in this case, the activities and their semantic descriptions. The content acquired from the user is the degree of interest in different tourism motivations. This degree of interest is mapped into the developed ontology, which can be navigated to find more general or more specific types of activities that can be of interest for a particular user. With this approach, user preferences can be generic terms, such as culture, leisure, or sports, avoiding the need of requiring too many details from the tourist that is using the system. Figure 1 depicted the fill-in form of the initial user profile acquisition. Activity categorization is given by more detailed concepts, such as
historical museum, cathedrals, theme parks, and so forth. Therefore, the ontology hierarchy permits to compute similarities between users’ preferences and activity descriptions. Such similarity is computed by the aggregation of activity concepts that are related to the users’ preferences using the parent-child relationship of the ontology.

Once the user receives activity recommendations on the map, he/she is able to interact with the recommendation list. The explicit data are the specified user’s liking on recommended activities and the implicit data are given by the user’s behaviour, such as adding activities to a travel planner, seeking detailed activity information or requiring other activities similar to a recently recommended one. Since activities are mapped to concepts in the domain ontology, the user’s actions are also applied to those concepts. This method allows the application to acquire a more detailed concept of interest. In addition, a spreading activation algorithm (Sieg et al., 2007) has been developed using the domain ontology, that sends for each node the weight obtained by explicit and implicit actions over activities to their neighbour nodes.

3.5 Collaborative filters

The idea of collaborative filtering (Kruszyk et. al., 2007) is to make recommendations based on what similar users have visited and their level of satisfaction. Similarity between users is normally computed by matching user ratings. This method requires that each user rates a set of items to predict accurate recommendations. However, the probability of users rating the same items in large data bases is relatively small. To overcome this drawback, ontologies can be applied in order to have a hierarchical and semantic structure of the activities, which permits to make inferences at different levels of generality (Fink & Kobsa, 2002). For example, the appreciation of one user for Tarragona National Archaeological Museum gives evidence that he/she probably likes both the rated museum and other “archaeological museums”. This gives the ability to calculate the similarity between users that have rated either the same activities or similar activities. In this system, in addition to the preferred interests of the users, demographic data are considered. The knowledge of travel accommodation, group composition and country of origin can be used to classify the users in common tourist typologies. Thus, the categorization of the visitors is based on motivational and demographic data as well as the users’ ratings of activities.

The categorization is performed with an automatic clustering process that generates a set of common tourist types, according to both tourist motivations and demographic characteristics. Machine Learning techniques have been used to build automatically a classification of the users. In particular, the $c$-means clustering algorithm has been used (Brouwer, 2007). Clustering processes are based on the measurement of the similarity between two users. To calculate this similarity a novel method combining different aggregation operators has been used. In particular, the information regarding the user preferences on different types of activities is aggregated using the OWA operator (Torra & Narukawa, 2007). Then, the global evaluation of the similarity with respect to the interests is combined with the comparison of the demographic features using the LSP operator (Dujmovic, 2005). Those operators are particularly
interesting because they permit to specify different policies during the integration of the information. So, one can decide which features are mandatory, which ones are optional, and the degree of simultaneity required to make the global similarity evaluation.

In the initial stages of application of the system, when no rating information of the current user is available, general knowledge based on the characteristics of visitors of Costa Daurada and Terres de l’Ebre is used. The system has been initially enriched with common tourist activity preferences obtained from a survey of 30,000 questionnaires conducted in Costa Daurada and Terres de l’Ebre between 2001 and 2009. Up to 100 different tourist type classes were extracted using the most common combinations of tourist categorization. When the system is in the exploitation stage, the categorization based on the users’ ratings obtained with our system will acquire more relevance, giving a more accurate recommendation to the user. In addition, not only the users’ rating - which some users avoid - is acquired, but also the observed user behaviour that can be used to derive knowledge about the appreciation of specific activities. In order to reduce the computation time required to calculate the similarity between users, a pre-computing clustering process is performed during times of low activity in the server.

3.6 Integration of content-based and collaborative methods

The combination of content-based and collaborative methodologies is achieved through the aggregation of their weight values for each concept of the domain ontology. The aggregation values change over time giving a different level of relevance. The relevance of collaborative methods depends on the level of similarity between the current user and the most similar class. In addition, the distance between the activities location and the specified destination is also considered in the final recommendation.

Clustering methodologies are applied on the survey data to extract generalization of tourists’ interests while our user application database remains poor in information. More relevance is given to collaborative filters as the number of users of the application increases, thus yielding more accurate recommendations with respect to the surveys. During the user life cycle, the collaborative filters based on ratings and inference domain methods will augment in relevance as user actions increase, giving less importance to initial motivations and demographic data that are used in the initial stages.

A basic aspect of the system is the collection of explicit and implicit feedback in order to propagate the weight of user preferences in upward and sideward directions. Imagine for example a tourist that expresses high interest in museums and monuments rather than in theme parks and shopping. Then, the inference system will deduce that the tourist is more interested in culture than in leisure. Since the last statement is not always true, the final recommendation is given by the combination of all the methods previously commented, giving rise to a hybrid recommender system that provides the main advantages of each methodology.
4 Final Considerations and Future Research

By increasing the familiarity and accessibility of a complex destination region to at least a part of the visitors, the SigTur/E-Destination system described in this paper allows a sustainable management of tourist flows. In addition, it is bound to bring about an improvement in visitor satisfaction and, as a consequence, the profitability of attractions. On the other hand, the services and digital contents are considered important resources to achieve a diversification and differentiation of the product. SigTur/E-destination configures, then, a laboratory that helps companies and local institutions to implement advanced systems of contents and proposals for the current and potential visitors based on their profiles. Moreover, it is possible to use SigTur/E-Destination to simulate the result of determinate decision-making processes, predictions, and scenarios for tourism and territorial development.

The possibility to easily update the information on assets and products (for instance about opening times, prices and eventually reservations) in a collaborative way at the “back-office” level is a key feature of this system that will be developed in further stages, taking full advantage of the role of our Science and Technology Park of Tourism and Leisure as a knowledge hub and service provider for the professional tourist networks of this region. This is especially important when considering the ephemeral and volatile nature of tourist information that is normally found in destination websites, however frequently updated, and the high actualization costs faced by Destination Management Organizations. Another development that is foreseen for this system is its adaptation for mobile devices, which will provide travel guide recommendations in real time. Moreover, it will include context-awareness, based on GPS, weather or traffic information among others, to provide more accurate recommendations. Next, the semantic side of the system could be enhanced using not only ontologies, but folksonomies created collaboratively by users annotating contents. This can provide a more detailed and dynamic categorization of the catalogue activities; however, intrusive information could also damage item categorization and finally the quality of the obtained recommendations. For this reason its treatment deserves further research.

Finally, an immediate development would be to define new methods to evaluate the recommender system. Normally, the accuracy of the recommendation is evaluated. Nevertheless, it does not assure user satisfaction on the recommendations. In this sense, next January it is planned to present the application in the International Tourism Trade Fair FITUR in order to be tested by potential visitors and obtain their feedback regarding the user experience.

References


**Acknowledgement**

This project was funded by the FEDER European Regional Funds and the Government of the Province of Tarragona.