SigTur/E-Destination: Ontology-based personalized recommendation of Tourism and Leisure Activities

Antonio Moreno a, Aida Valls a, David Isern a, *, Lucas Marin a, Joan Borràs b

a Departament d'Enginyeria Informàtica i Matemàtiques, Intelligent Technologies for Advanced Knowledge Acquisition (ITAKA), Universitat Rovira i Virgili, Av. Països Catalans, 26. 43007 Tarragona, Catalonia, Spain
b Science & Technology Park for Tourism and Leisure, C/Joanot Martorell, 15. 43480 Vila-seca, Catalonia, Spain

A R T I C L E   I N F O
Article history:
Received 9 November 2011
Received in revised form 20 January 2012
Accepted 20 February 2012
Available online 17 March 2012

Keywords:
Recommender system
Tourism
Ontology
Collaborative filtering
Management of user profile

A B S T R A C T
SigTur/E-Destination is a Web-based system that provides personalized recommendations of touristic activities in the region of Tarragona. The activities are properly classified and labeled according to a specific ontology, which guides the reasoning process. The recommender takes into account many different kinds of data: demographic information, travel motivations, the actions of the user on the system, the ratings provided by the user, the opinions of users with similar demographic characteristics or similar tastes, etc. The system has been fully designed and implemented in the Science and Technology Park of Tourism and Leisure. The paper presents a numerical evaluation of the correlation between the recommendations and the user’s motivations, and a qualitative evaluation performed by end users.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Nowadays, due to the astonishing volume of available information, the search and organization of the activities to be done during a holiday trip can be an arduous task. In addition, in the majority of cases, tourists are looking for information about a place in which they have not been before, which hampers the selection of the activities that fit better with one’s preferences. Another point of view to consider is the one of the local stakeholders, which have interest in promoting the tourist welfare of a particular region, especially those activities that are less popular. An efficient design, organization and communication of opportunities in the region may lead to a more balanced tourism activity (spatially, thematically and financially), with important returns in terms of sustainable development. Those are the main reasons that motivate the use of decision support tools (known as Recommender Systems) that help actors in both sides of the problem to make the optimal elections, in this case, the best tourist activities.

Recommender Systems are emerging as important tools in the development and management strategies of destination regions and cities. These systems are able to deal with increasing degrees of sophistication in the definition of the alternatives available to the user and in the management of user’s data. That relieves the users from having to manually evaluate their possible choices and helps to avoid judgement mistakes when comparing all the available alternatives.

In order to make a satisfactory recommendation it is important to ensure that the characteristics of the recommended activities match with the tourist’s interests (i.e., preferences). The information about the user, including his/her preferences, is usually stored in a personal data structure known as profile. The information stored in the profile is usually gathered in three ways: it can be explicitly captured by asking the user directly for it (e.g., requiring the user to fill a questionnaire), the system can try to associate the user with a predefined social group that has well-known preferences (technique known as stereotyping) or new information about the user can be obtained in an implicit way by observing his/her interaction with the system (e.g., analyze the evaluations provided by the user and recommend activities similar to the ones the user liked). Recent approaches also try to build profiles using semantic knowledge, rather than mere numerical representations (Blanco-Fernández et al., 2011). In this sense, Artificial Intelligence representation languages and inference tools are of outmost interest to improve the quality of the recommendations.

It is also important to note that the recommendation of touristic activities is highly related to the spatial distribution of activities and visitors. Therefore, it can be claimed that the
combination of Artificial Intelligence and Geographic Information Systems (GIS) in a Recommender System provides an appropriate way to deal with spatial data during the recommendation process. These technologies allow users to reduce and make more effective their travel planning time by receiving personalized assistance (Ricci et al., 2009).

This paper presents SigTur/E-Destination, a tourism recommender system developed by the Science and Technology Park of Tourism and Leisure of Vila-Seca, based on the integration of GIS and Artificial Intelligence algorithms. On one hand, the GIS enables to store a large collection of geospatial information related to Tourism and Leisure Activities, as well as to provide a map-based interface to the user with the localization of the proposed activities. On the other hand, Artificial Intelligence tools, such as ontologies, provide a semantic integration of the geospatial information within the recommender system, while content and collaborative methods help to provide personalized recommendations. SigTur/E-Destination makes the whole range of products and itineraries accessible to visitors that plan their visits to the region of “Costa Daurada and Terres de l’Ebre” (in Catalonia, Spain), as well as to those that being already there want to enjoy richer experiences. The system provides users with a great range of possibilities to identify leisure activities according to their profile and, beyond that, facilitates the planning of the trip and the decision-making process before and during the stay.

A preliminary work (Borràs et al., 2011) presented the main features of the system from the user’s perspective, summarizing the interface with the user through interactive maps. In this paper, we focus the attention in the recommender system and the construction and exploitation of user profiles by means of an ontology-based model. The paper also explains the techniques used to combine the feedback from one user with the feedback collected from other users in order to make a good selection of activities. Several feedback indicators are gathered by the system and mapped into concepts of the ontology in order to learn an unsupervised way which types of activities are the most interesting for a given user. A specific domain ontology supports the recommendation and plays a key role in the recommendation process.

This paper aims to provide a new perspective on the usage and combination of ontology-based algorithms in recommender systems, focusing on the experience of the project SigTur/E-Destination. The rest of the paper is organized as follows. First, an analysis of previous related works is presented. After that the SigTur/E-Destination system is introduced, paying special attention to the management of the user profile and all the different aspects taken into account by the recommender system. A quantitative evaluation of the recommendations provided by the system for typical classes of tourists and a qualitative evaluation of the whole system made through questionnaires are presented. The paper ends with a conclusion and an outline of some paths for future work.

2. Related work

This section briefly reviews related work in the areas directly associated to ontology-based intelligent recommendation of touristic activities, namely the representation of user preferences and the use of ontologies in e-Tourism. Finally, we discuss the situation and contribution of SigTur/E-Destination in these fields. Table 1 summarizes the approaches analyzed through this section.

2.1. Representation and management of user profiles

The relevance of the advice of a recommender system depends on the accuracy of the user profile. The basic approaches to user modeling in the literature include demographic characteristics, context-aware information, and personal preferences (Montaner et al., 2003) (see columns D, CA and P on Table 1).

Demographic characteristics, such as civil status, age or studies, identify the user in a social group or domain. This kind of information is usually not enough to drive the recommendation processes. However, it can complement other information and be useful in a first stage to have a rough classification of the user that allows to start giving generic recommendations based in users with similar social characteristics, as done in Basiri et al. (2010).

Several aspects, such as the geographical location of the user, the budget of the trip or the means of travel (De Carolis et al., 2009),

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Brief description</th>
<th>Ontologies</th>
<th>Types of recommenders</th>
<th>GIS info.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content-based</td>
<td>Collaborative</td>
<td>Demographic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Chikhaoui et al., 2011)</td>
<td>Hybrid recommender system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Basiri et al., 2010)</td>
<td>Hybrid recommender system that uses predefined stereotypes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(De Carolis et al., 2009)</td>
<td>MyMap: context-based recommender in a mobile device</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Abbaspour and Samadzadegan, 2011)</td>
<td>Time-dependent urban tour planning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Montejo-Ráez et al., 2011)</td>
<td>Ottùm: Web-based activity planner</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Schiavino and Amandi, 2009)</td>
<td>Recommender of holiday packages and tours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Vansteenkoven et al., 2011)</td>
<td>City trip planner: recommends points of interest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Garcia et al., 2011; Sebastiá et al., 2009)</td>
<td>e-Tourism: activity scheduler</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Huang and Bian, 2009)</td>
<td>Personalized recommendation of tourist attraction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Büyüközkan and Ergün, 2011)</td>
<td>Case-based recommendation of trip alternatives</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Castillo et al., 2008)</td>
<td>Software assistant for tourists</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Lee et al., 2009)</td>
<td>Agent-based ontological recommendation of activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Batet et al., 2012)</td>
<td>Turist®: agent-based recommender on mobile devices</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1
Comparison of some personalized recommender systems.
can be characterized as the context of the recommendation. Letting
the recommender know some information about the user where-
abouts and current travel circumstances helps optimizing the
system advices, saving time and unnecessary displacements. This
kind of information, usually objective, is normally used in recom-
mender systems focused in creating touristic itineraries and plan-
ing some vacation time (Abbaspour and Samadzadegan, 2011;
Montejo-Ráez et al., 2011).

Recommender systems, however, need to have a model of the user
interests and, in most cases, they try to improve it dynami-
cally with the feedback given by users, so that it can be success-
fully employed to produce adequate recommendations (Coelho
et al., 2010). This is where the third kind of information, personal
preferences, comes in handy. It is frequent to distinguish between
the details about the interests over certain features or attributes of
the recommendable items, and the historical information about the
user interaction with the recommender system, in the form of
explicitly (rating of items, user comment) or implicitly (observa-
tion of user interaction) given information.

2.2. Tourism recommendation types and trust

As pointed out in Garcia et al. (2011), current Tourism
recommenders can suggest either a trip destination or a set of
activities to do once the user has arrived to a certain place.
Most of the current travel recommender systems have focused
on the former issue. The main problem faced by this kind of
systems regards the integration of heterogeneous sources of
information (like Web resources associated to flight or hotel
companies), in order to configure a trip that matches with the
user’s constraints and preferences. In many cases, information
extraction techniques and semantic technologies such as specially
tailored ontologies are used to parse, interpret and integrate
information stored in heterogeneous sources (Ambite et al.,
2002; Camacho et al., 2005). Those systems store a set of
preferences of the user that allows them to rate and rank a
collection of destinations. On the other side, there are fewer
approaches, like SigTur/E-Destination, focused on recommending
attractions that the tourist can visit at the destination. These
systems suggest places to visit in a certain geographical area
taking into account the user profile, usually providing a daily plan
(Huang and Bian, 2009). Unlike the approaches mentioned above,
attraction recommendations are typically performed on-line at
the destination place according to several dynamic criteria (e.g.,
timetable, location, etc.). In this case, the main problem consists
on considering and integrating these criteria to propose a list of
recommendations (Garcia et al., 2011; Sebastáı̈ et al., 2009).

Recommender systems may also be classified in three basic
types, depending on the recommendation strategy: content-based,
collaborative and demographic approaches (Montaner et al., 2003).
In collaborative filtering approaches, recommendations are made
by matching users that have similar preferences and suggesting
items they like (e.g., Castillo et al., 2008; Lee et al., 2009). This
strategy requires several ratings from the users before the system
may begin to make useful recommendations. This is known as the
cold-start problem (Basiri et al., 2010). Content-based approaches
(Pazzani and Billsus, 2007) recommend items by considering the
features of the activities that users have enjoyed previously. Only
items closely related to those the user liked in the past are
recommended. To perform this kind of recommendation, it is
necessary to build a user profile that stores the degree of interest
on each of the different criteria that describes an activity.
Demographic recommender categorize users according to stereo-
typical classes and base the recommendations on general features
related to those classes. In order to overcome the limitations of
individual recommendation schemas, some authors propose
hybrid approaches (Schiaffino and Amandi, 2009; Burke, 2000;
Salter and Antonopoulus, 2006; Christakou et al., 2007), which
combine some of the above mentioned types.

Another aspect worth considering is the degree of confidence
on the recommendations shown to the user. It is easy for people
to trust on recommendations made by a close friend or a relative,
but most users show some level of mistrust on recommender
systems (Hinze and Quan, 2009; Quan and Hinze, 2008).

2.3. Use of semantic information: Tourism ontologies

Ontologies define areas of common understanding between
multiple actors, easing their interoperability and permitting a
high-level communication (Berners-Lee et al., 2001). Their basic
components are concepts and relationships between them.
In the last decades, the Tourism sector has developed catalogs
and taxonomies to facilitate information management. Lately, an
effort to generate global standards has been made in order to
facilitate the exchange of data between Tourism agents. This is
the case of the Thesaurus on Tourism and Leisure Activities1
defined by the World Tourism Organization (WTO).
Recently, different Tourism ontologies have been developed.
Some of them have reached a considerable level of consolidation,
allowing the representation of not only generic aspects, but also
specific sub-domains that describe detailed scenarios (such as
regional ontologies). Harmonize2 was one of the first ontologies
that aimed to face the interoperability problems of Tourism,
 focusing on data exchange between organizations. It covered four
main topics of the Tourism domain: attractions, events, food and
drink, and accommodation. Afterwards, Mondeca (Prantner et al.,
2007), developed an ontology with around 1000 concepts, most of
them contained in the Thesaurus on Tourism and Leisure Activities
developed by the WTO.
Another ontology, QALL-ME (Ou et al., 2008), emerged in order
to establish a shared structure for multimodal and multilingual
Tourism question answering. The DERI e-tourism ontology (Hepp
et al., 2006), developed in the OnTour project, covered three main
issues: accommodation, activities and infrastructures. Some
classes of this ontology were used as a test-bed for an automatic
system of ontology population (Ruı̈z-Martı́nez et al., 2011). Finally,
the cDOTT ontology (Barta et al., 2009) (The Core Domain
Ontology for Travel and Tourism), developed in 2009, was based
on the Harmonize ontology. Its main idea was to define a
common ontology for the tourism sector in order to support the
interoperability of the agents in low-level operations.

2.4. Discussion

In comparison with the related work presented in this section,
the SigTur/E-Destination system presents some novel character-
istics. One of its distinguishing features is the integration of
several types of information and recommendation techniques,
as will be detailed in the rest of the paper. Concerning the
information used by the recommender, it takes into account
demographic data, details that define the context of the travel
(e.g., composition of the travel group), geographical aspects,
information provided explicitly by the user (e.g., main travel
motivations) and implicit feedback deduced from the interaction
of the user with the system.

---

1 See: http://www.wtoelibrary.org/content/m7434p/fulltext.pdf (last access August 6th, 2011).

---

A. Moreno et al. / Engineering Applications of Artificial Intelligence 26 (2013) 633–651
SigTur/E-Destination employs many recommendation techniques, from the use of stereotypes [standard tourist segments] to content-based and collaborative filtering techniques. The Artificial Intelligence tools used in SigTur/E-Destination include automatic clustering algorithms, ontology management, and the definition of new similarity measures between users, based on complex aggregation operators.

An important aspect of SigTur/E-Destination is the use of a domain ontology to guide the recommendation process, which permits to make inferences about the correspondence between the characteristics of an activity and a certain user profile. SigTur/E-Destination makes a knowledge-level analysis of the user preferences, including processes that make bottom-up and top-down propagation of the preferences over the concepts of the ontology. The system also associates a certain degree of confidence to each specific recommendation. This information is very useful in order to take the final decision of which activities to show to the user.

Finally, the project includes GIS tools to store the main tourism and leisure resources with geospatial information, which is used to recommend the activities and to show the results in a user-friendly map-based Web application. The GIS database was designed according to territorial singularities of the “Camp de Tarragona and Terres de l’Ebre” area. It was decided to build a specific ontology that fits the specificities of this territory. The design of this new domain ontology followed the main concepts of the thesaurus of the WTO. The level of detail in each part of the ontology depends on the set of activities available in this particular area. For example, there is a deep level of detail about concepts related with “Wine” due to the importance of Enotourism in the region.

3. Design and development of the recommender system

This section describes the architecture of SigTur/E-Destination and its main components, along with the explicit and implicit information gathered from the users. One of the basic knowledge structures of the system is a Tourism ontology, whose design and composition are also commented below. Afterwards, Section 4 details the recommendation techniques used by the system, and how it combines content-based and collaborative approaches.

3.1. Architecture of the SigTur/E-Destination system

Fig. 1 depicts the general architecture of SigTur/E-Destination. All modules, which have been developed using Open Source technologies, are organized in a traditional client-server structure. The most novel aspect of the system is the careful combination of different technologies, which has led to the development of an application that uses advanced Artificial Intelligence techniques in an efficient way, presenting a low execution time. These techniques are totally hidden from the users, who only interact with a user-friendly Web application that shows information on maps and lists that are very easy to manage.

Users access the system through a Web browser. The Web server application has been built in Java Server Faces with the 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. The concurrency of the system (as multiple users may be accessing the application at the same time) is addressed by the framework using a thread pool which provides bounded thread usage in large-scale applications. The representation of the geographical resources in maps has been achieved using the Google Maps API. This API allows creating maps embedded in the application and using other services such as Street View, geocoding and the calculation of routes between two given points.

The core of the architecture is the recommender system, developed in Java, which handles the interaction between all the modules. Moreover, it manages the user profile dynamically updating its state after each user action. This allows the recommender system to take into account the behavior of the user and provide more accurate results.

Data are stored in two databases. One of them contains the tourist resources, including all the geographical information needed to show them in maps. The other one stores the user profiles. User data are managed by PostgreSQL and tourist resources are stored in the PostGIS extension. Database connections are handled by the Hibernate framework 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 has been used. Databases are not only used to manage data but also to optimize search functions. SQL scripts have been developed to execute data mining techniques in an efficient way, hence providing time responses that are lower than other methods such as collaborative filters. Moreover, spatial PostGIS functions have been used to filter geo-referenced items in order to optimize data queries.

Fig. 1. Architecture of SigTur/E-Destination.
3.2. Tourism 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 203 connected concepts in 5 hierarchy levels. As it is depicted in Fig. 2, the ontology is structured around eight main concepts that constitute the first level of the hierarchy: “Events”, “Nature”, “Culture”, “Leisure”, “Sports”, “Towns”, “Routes” and “ViewsPoints”. The last three classes are considered transversal concepts, since they share children nodes with other main classes, e.g., “Routes” and “Nature” are both superclasses of the “NatureRoutes” class. The rest of the concepts in the ontology are connected via is a (subclass) relationships with these main classes. The ontology is not a pure taxonomy, as it contains multi-inheritance between concepts, e.g., EthnographicMuseum is a subclass of both Museum and Traditional.

This ontology has been developed using the Thesaurus of the World Tourism Organization1 as a reference guide to represent the touristic and leisure activities in the “Costa Daurada and Terres de l’Ebre” region. The decisions about which concepts and relationships should be represented have been taken by a committee of experts in the tourism domain from the Scientific and Technological Park of Tourism and Leisure. The level of detail in each part of the ontology depends on the set of activities available in our particular geographical area of interest. For example, there is a deep level of detail about concepts related with “Wine” due to the importance of enotourism in the region. In any case, the ontology could be easily extended with more concepts if it was necessary. For instance, this ontology could be customized to another region where winter sports were relevant, by adding a new concept called “WinterSports” (with its appropriate subclasses) and putting it as a subclass of the “NonAquaticSports” concept.

The ontology is used to explicitly classify the activities to recommend among a predefined set of distinctive main concepts, which are used by the intelligent recommender system in its reasoning processes, as will be explained in the following section. Each activity is tagged with one or more ontology concepts, which are usually low level nodes in the hierarchy. For instance, the Roman Amphitheater of Tarragona is tagged with the following concepts: “HistoryMuseums”, “Roman”, “HumanHeritage”, “Romanesque” and “Amphitheater”. The ontology only contains classes that permit to describe types of activities. It does not include instances to represent activities, since the number of activities may change dynamically at run-time. Hence, activities are stored in a database that is maintained via a Web content manager (the GIS database shown in Fig. 1).

For each user session the ontology classes are loaded into memory, so that the recommender system may associate a preference degree to each of the classes, depending on the explicit and implicit information provided by the current user. These preferences are the key information to decide which activities to recommend to the user.

The domain ontology has been developed with the Protégé9 editor and represented in the OWL language. Jena10 is the Semantic Web framework used in SigTur/E-Destination. It provides tools to manage the ontology and to apply inference mechanisms based on rules.

3.3. Creation and management of the user profile

The SigTur/E-Destination recommender system manages a user profile that is composed by two parts: (1) a static part represented as a vector with demographic and travel information and (2) a dynamic part represented with a particular version of the Tourism ontology. Considering that the ontology contains classes that represent types of activities, each class permits to store the degree of interest on the corresponding type of activity for a given user. For instance, if the current user of the system likes to visit museums and is especially interested in wines, the concepts “Culture”, “Museums”, and particularly “WineMuseums” will have a higher degree of preference than others. This part of the profile is updated when new knowledge is obtained from the user.

---

9 This work uses the stable release 3.4.7 of Protégé editor available at http://protege.stanford.edu (last access August 6th, 2011).
10 For more information: http://jena.sourceforge.net (last access August 6th, 2011).
The degree of interest in each concept is calculated taking into account the interest in more general and more specific types of activities, as will be explained in the following section. In order to infer the user preferences, the application acquires both explicit and implicit information from the user. The former includes the specification of the travel motivations and the rating values given by explicit evaluations of visited events. The latter is obtained from the observation of the actions of the user on the system, such as requesting more information from a certain activity or adding an event to the travel planner. The next subsections give more details on these two types of information.

3.3.1. Explicit information

The first task of a user in the system is to complete a form, which is used to create the initial profile. The main goal is to obtain as much information as possible with a small number of questions. The Tourism partners of the SigTur/E-Destination project elaborated a survey questionnaire to discover the most common travel motivations of the tourists that visit the “Costa Daurada and Terres de l’Ebre” region. From a statistical analysis of thousands of surveys, it was discovered that the main motivations (sorted in order of importance) were the following: beach, shopping, relaxation, theme/leisure parks, culture, nature, gastronomy, sports, and shows/events. Each of these motivations corresponds to a concept stored as a class in the Tourist ontology. Fig. 3 shows the interface used by the tourist to enter the degree of relevance (0–100%) of each of these motivations. These values are stored into the ontology of the user as an initialization of his/her profile.

The data needed to initialize the demographic and travel information of the user is obtained also with a form presented to the user at the beginning (see form in Fig. 4). These data include information about the country of origin of the user, other people with which the user travels (the allowed values are shown in Table 2), the location of the accommodation, the type of accommodation (allowed values also shown in Table 2), an initial estimation of the budget, and the travel dates. Some of those variables are used to filter the results before they are shown to the user (budget and travel dates) or to locate the recommendations into a given geographical area (near the chosen destination).

The selection of the questions proposed in Fig. 4 is based on a study of previous data collected from tourists of the same region. We considered 30,000 questionnaires filled in by tourists in the last 10 years with the aim of discovering what kinds of activities tourists visit based on their characteristics. The study was conducted by the Costa Daurada Tourism Observatory. In order to find out the most relevant criteria, we computed a logistic regression to discover those variables that provide more information concerning the type of activities enjoyed by the users. The variables shown in Fig. 4 were selected, while other variables from the questionnaire were discarded due to their low discrimination value, such as the tourist age, sex, profession, social class or the number of previous visits to our region. A similar statistical analysis of the factors that have a stronger influence in the recommendation of touristic activities was also proposed in Heu et al. (2012).

Apart from the explicit information given at the beginning of the session by the user, the system is able to obtain explicit information from the evaluations that users can make on the activities they have already visited, expressing their degree of satisfaction. Users may rate activities with an integer value between 1 and 5, where 5 corresponds to the best. The form that allows the user to rate and write a review about activities is shown in Fig. 5. The system reminds users to access such form after his/her travel has finished.

3.3.2. Implicit information

The system also takes into account the actions performed by the user during his/her interaction with the system, in order to...
improve its recommendations. This information, implicit in the user behavior, is commented in this section.

Once the user obtains a list of recommendations (the manner the system produces the recommendations is explained later) he/she is able to make several actions on the proposed activities. The system is able to infer the user interests by capturing and analyzing these actions. This is very useful to adapt dynamically and automatically the user profile and make more precise the degree of interest of the user on each kind of activity during the recommendation session.

The user is able to select those activities he/she is interested in (see Fig. 6) and add them to a travel plan. On the left column of Fig. 6 there is a list of recommended activities and a list of activities visited by similar users. Each activity includes a percentage which is the degree of correspondence between the activity characteristics and the user preferences. On the right side, the location of the recommended activities is marked in a map. Other actions the user is able to make on activities are to request more detailed information on a specific event (see Fig. 7), to ask for activities close to the currently selected one (see also Fig. 7) or to obtain events that are similar to the current one (see Fig. 8). All these actions give evidences that the user is interested in the current activity in some way. On the other hand, it is also possible for the user to ask for a new list of recommendations; in this case, the activities over which the user has not made any action are considered as uninteresting for him/her. All these actions provide implicit information that is very useful in the recommendation process.

4. Recommendation of activities

This section explains how the system predicts the degree of interest of the user on each type of activity, that is, how an ontology-based profile is maintained and exploited. The aim is to suggest a ranked list of activities that are interesting to the user. The system combines content-based and collaborative recommendation techniques, as described in the following subsections.

4.1. Content-based recommendation

Content-based recommenders (Pazzani and Billsus, 2007) are based on a direct matching between the features of the activities to be recommended and the user interest in each of those features. The SigTur/E-Destination system contains a database with all the available touristic and leisure activities in the region (GIS database on Fig. 1). Each of the activities is labeled with a list of concepts belonging to the Tourism ontology introduced in Section 3.2. The basic aim of the recommender system is to
associate a degree of preference (between $-1$ and 1) to each concept of the ontology; from these preferences, it can then compute the interest that the user may have on each particular activity. The system does not only store the interest score (i.e., preference degree) for each concept, but also the level of confidence (between 0 and 1) on that value. This confidence level

---

**Fig. 5.** Explicit evaluation and comment of activities the user has visited.

**Fig. 6.** Screenshot of the recommendation pane. The user adds an activity to the travel plan by selecting the checkbox.
depends on the evidences that have led to the computation of the interest degree on that particular concept. In fact, as will be explained later, the system stores different interest scores and confidence values for each ontology concept, each one associated to a certain kind of recommendation technique. All these data are finally aggregated into a unique preference and confidence value, as will be described in Section 4.3.

4.1.1. Travel motivation

The initial information provided by the user is the motivation of the travel, given in the form shown previously on Fig. 3. Each of the levels of interest of the nine possible motivations is directly mapped into the corresponding concept of the ontology, with a full confidence. For instance, if the user specifies 85% for the motivation “beach”, the system applies a score value ($S$) 0.85 to
the ontology concept “Beaches” and the confidence level (CL) in that concept is set to 1.0 since that information has been provided directly and explicitly by the user.

4.1.2. Interaction of the user with the system

As commented in Section 3.3, the analysis of the actions of the user in the system also provides implicit information on his/her interests, which can be used to calculate the preference degree on each ontology concept. When the user is presented with a list of options, he/she can perform different actions on each activity. We have associated an interest score and a confidence level to each action, which are applied to the ontology concepts associated to the manipulated activity. The interest score is positive if the action shows that the user likes the activity (e.g., requesting more detailed information of an event), and negative if the action seems to indicate that the user is not interested in the activity after all (e.g., removing an event from the travel planner). The confidence level associated to each action reflects its subjective relevance.

The explicit ratings provided by the user also give a direct positive/negative feedback on a particular activity, which can be transferred to the ontology concepts it is related to. Ratings are given a full confidence level, since they are explicit information freely given by the user. The CL values for implicit actions are set lower than the ones for explicit actions, since they are considered less accurate (Kelly and Teevan, 2006). Table 3 shows the range of possible score (S) values and the default confidence level (CL) for each action. Finally, we also extract information from the absence of actions on a certain recommended activity. In that way, when the user asks the system to provide a new list of recommendations, we can know which activities have not been considered by the user in any way, and decrease the associated interest scores.

Each activity is mapped to one or more concepts in the lowest level of the domain ontology. For that reason, we have to update each concept separately. Let us name $K_{ucc}$ the set formed by all the actions made by the user u on activities labeled with the concept c. Then, for a given user u and a given concept c, the S and CL values are computed taking into account all elements in $K_{ucc}$ as follows:

$$ S^u(c) = \frac{\sum_{i \in K_{ucc}} S(i) \cdot CL(i)}{\sum_{i \in K_{ucc}} CL(i)} $$

(1)

$$ CL^u(c) = \frac{1}{|K_{ucc}|} \sum_{i \in K_{ucc}} CL(i) $$

(2)

Notice that the score indicating the interest of user u in a type of activities c ($S^u(c)$) is the mean value of all the scores of the actions performed by the user on the activities of that type ($K_{ucc}$), weighted by the confidence level of each action. The confidence level on the preference of user u for a type of activity c ($CL^u(c)$) is the average of the confidence levels of all the actions made by the user on activities of that type.

For instance, imagine the user requests detailed information (action3) and then adds to the travel plan (action1) the recommended activity “Museu Nacional Arqueológico de Tarragona”, whose associated ontology concepts are “ArqueologyMuseums” and “Roman”, but he/she does not make any action (action7) on the recommended activity “Col·lecció Cova museu de la Font Major” whose associated concept is “ArcheologyMuseums”. The score $S$ of the interest on the “ArcheologyMuseums” concept will be calculated as

$$ S = \frac{(S(action1)\cdot CL(action1))+(S(action3)\cdot CL(action3))+(S(action7)\cdot CL(action7))}{CL(action1)+CL(action3)+CL(action7)} $$

$$ = \frac{(1.0\cdot 0.5)+(0.6\cdot 0.5)+0.0}{0.5+0.5+0.0} = 0.823. $$

The confidence level value of the “ArcheologyMuseums” concept will be set to

$$ CL = \frac{CL(action1)+CL(action3)+CL(action7)}{\text{number-of-actions}} = \frac{0.5+0.2+0.15}{3} = 0.283. $$

Hence, the degree of interest predicted by the system for the concept “ArcheologyMuseums” is 0.823 and the confidence level on this value is 0.283. The system updates during the user session those two values for all the ontology concepts associated to the recommended activities with which the user interacts.

4.1.3. Ontology-based propagation of interest and confidence values

The Tourism ontology provides a hierarchical representation of the main kinds of activities in the domain. The information obtained from the user actions (described in the previous section) is mapped into preferences related to the concepts associated to the manipulated activities, which are nodes in the lowest levels of the ontology. It might be interesting to propagate that information up the hierarchy (Sieg et al., 2007), since the interest in one kind of activity also suggests some interest in the corresponding superclasses (someone interested in “ArcheologyMuseums” can be said to be interested in “Museums” and, in turn, that interest can also be moved to “Culture”).

Thus, a spreading algorithm has been used to propagate the preference values of the ontology nodes to their ancestors. Each node sends its interest score (related to the actions made by the user) to its superclasses. All the nodes that receive information compute the mean value of the preferences (and related confidence values) received from their children. A small reducing factor is applied at each level, so that the influence of the values decreases as we go up the hierarchy (for instance, in the previous example, the user might have a very strong interest in “ArcheologyMuseums”, which would be translated into a strong interest in “Museums” and a mild interest in “Culture”).

4.2. Collaborative recommendation techniques

Collaborative filtering techniques are recommendation methods based on the opinions of a set of users. They can focus on the items or on the users. The methods based on items (Linden et al., 2003) predict the interest of the user on an activity a considering the evaluation that this user has given to similar activities (defined as those that have been positively rated along with a by many users). On the other hand, user-based approaches (Jin et al., 2004) implement the “Word of Mouth” phenomenon, predicting the interest for an activity a through its evaluation by similar users.

In applications where the number of users exceeds the number of items, item-based recommendation methods present a better accuracy and efficiency (Desrosiers and Karypis, 2011). However, user-based approaches are more stable when items are dynamic, and they may also produce serendipitous recommendations. That
is a useful property to discover different types of items and produce more varied recommendations; thus, in this work we have considered user-based similarities.

Therefore, the main objective of our collaborative filtering techniques is to find users similar to the current one, so that the system can recommend him/her activities that were considered interesting by those similar users. The similarity between users can be computed in two ways: taking into account only the demographic information (two users are similar if they have close values in the demographic attributes), or considering the explicit ratings provided by the users (two users are similar if they give close ratings to the same activities). In the SigTur/E-Destination system we combine both strategies. At the beginning of the execution of the system, when the user has not yet provided any rating, the first kind of similarity is applied. When the user has already made a certain number of actions on the recommended activities, the second kind of similarity takes more relevance.

In order to perform a collaborative recommendation it is necessary to have a measure of comparison between two users, which gives us an estimation of their similarity. This measure can then be used to automatically build groups of similar users. In this work, we propose a similarity measure based on demographic and motiva tional attributes, which is explained in Section 4.2.1.

Due to its scalability in computation time, the \( K \)-means algorithm is applied to make the different clustering processes in the system, that will be commented in Sections 4.2.2, 4.2.3 and 4.2.4 (Ding and He, 2004). Thus, users are arranged in groups that have similar characteristics. The initial seeds (or prototypes) of the clusters are established using different techniques according to the type of recommendations that will be performed, as described in the following sections. On each step the distance between each user and the prototypes is computed, using the similarity function (5), and each user is assigned to the closest prototype. After that, the prototype of each class is recalculated, and the procedure is repeated again until it converges.

In particular, in SigTur/E-Destination clustering is applied with three different purposes: to obtain a basic initial set of tourist segments (4.2.2), to obtain classes of users with similar demographic characteristics (4.2.3) and to obtain classes of users with similar ratings (4.2.4).

4.2.1. Measure of similarity between two users

When a new user arrives, the system must compare it to other users of the system in order to find out who are the most similar ones and then start to making recommendations. The values that are considered in the comparison process are the travel motivations (Fig. 3) and the travel group composition, accommodation type and country of origin (Fig. 4).

As it has been said in Section 3.3.1, before defining the similarity measure, a logistic regression analysis was applied on the set of 30,000 hand-filled questionnaires to obtain the degree of relevance of each of attribute with respect to the discrimination of the travel activities performed by users. The result of this analysis is shown in Table 3, where we can see the weight associated to each attribute. It may be noticed that the composition of the travel group is the most relevant factor.

To calculate the similarity between two users \( u \) and \( v \), a novel method combining different aggregation operators is proposed. First, we measure the inverse of the distance on the values between \( u \) and \( v \) for each attribute separately. This gives us a vector of partial similarities \( x = (x_1, x_2, ..., x_{12}) \) where \( x_i = 1 \) if the two users have the same value on that attribute, and \( x_i = 0 \) if the values are completely different (see more details below). The vector \( x \) has initially 9 similarity values corresponding to the travel motivations, plus the similarity on the type of group, accommodation and country. To combine all this information into a unique value, we propose the use of two types of aggregation operators.

First, the partial similarities regarding the nine user travel motivations are aggregated using the OWA operator (Yager, 1988) in order to obtain a single similarity value with respect to the motivations. The OWA aggregation operator in a dimension \( n \) is a mapping \( R^n \rightarrow R \) that has an associated weighting vector \( W \) of dimension \( n \) with \( \sum_{j=1}^{n} w_j = 1 \) and \( w_j \in [0, 1] \), such that:

\[
\text{OWA}(a_1, ..., a_n) = \sum_{j=1}^{n} w_j b_j, 
\]

where \( b_j \) is the \( j \)-th largest of the \( a_i \).

The weighting vector to be applied in the aggregation of the travel motivations has been calculated using the classic Regular Increasing Monotone (RIM) linguistic quantifier defined by Yager in Yager (1996) as:

\[
Q_z(t) = r^t, 
\]

giving \( z \) the value 4 to allow a high degree of simultaneity. This means that we consider that the motivations of two users are similar only if most of their values are similar.

After that, this evaluation of the similarity with respect to the travel motivations is combined with the comparison of the demographic features using the LSP operator (Dujmović and Nagashima, 2006). This aggregation operator is particularly interesting because it permits 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 for making the global similarity evaluation. The final operator used to obtain the similarity between two users \( u \) and \( v \) is the following:

\[
\text{sim}(u, v) = (w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4)^{1/r}, 
\]

In this expression, \( r \) has been set to \(-3\) to specify a certain degree of conjunction. The values \( w_1, w_2, w_3 \) and \( w_4 \) are the relevance weights from Table 4 for travel group composition, accommodation type, country of origin and travel motivations, respectively. \( x_3 \) is set to 1 if \( u \) and \( v \) have the same travel group composition, and 0 otherwise. \( x_2 \) is set to 1 if the kind of accommodation of \( u \) and \( v \) is the same, 0.2 if they are similar (e.g., "Apartment rented through an agency" and "Rented apartment"), and 0 otherwise. \( x_3 \) is set to 1 if \( u \) and \( v \) have the same country of origin, and 0 otherwise. Finally, \( x_4 \) is the value obtained from the OWA operator explained previously given the motivations of \( u \) and \( v \). This similarity measure is used in different steps of the recommendation process as it is explained in the following sections.

4.2.2. Estimating the interests from similar segments of tourists

A common problem in collaborative recommender systems is the lack of users at the initial application stages. In order to solve this problem it was decided that, while the user database has a low number of users, general knowledge based on the characteristics of visitors (called tourist segments) to "Costa Daurada and Terres de l'Ebre" is used. Therefore, the system is initially enriched with the preferences associated to tourist segments obtained from a survey.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel group composition</td>
<td>0.37</td>
</tr>
<tr>
<td>Accommodation type</td>
<td>0.33</td>
</tr>
<tr>
<td>Country of origin</td>
<td>0.23</td>
</tr>
<tr>
<td>Travel motivation</td>
<td>0.07</td>
</tr>
</tbody>
</table>
of 30,000 questionnaires conducted in this area between 2001 and 2009.

An automatic clustering process, based on the well-known K-means algorithm, was applied to the set of questionnaires, using the measure of similarity described in the previous Section (5). Initial cluster seeds were selected by making a correspondence table between demographic data and travel motivations and finding out the most common relationships. The result of the clustering process, that was a set of 100 tourist types, was validated by calculating the optimal inertia (Gibert and Cortés, 1997) that quantifies both the separability between categories and the homogeneity within categories, considering different numbers of clusters and cut levels in the hierarchy. Afterwards, for each segment we build a prototype element, whose attribute values are obtained as an average of the segment members. So, the prototype contains values that are representative for its segment.

When a new user enters in the SigTur/E-Destination system, we search for the cluster (segment) that fits better with the characteristics of the new user, by comparing the information of the user with the prototypes of the clusters. SigTur/E-Destination also knows, for each cluster, which is the level of preference associated to the concepts of the ontology related to the travel motivations (for instance, clusters of senior travelers usually have a high preference value for the Relaxation concept, as they assign a high value to the Relaxation travel motivation). The degree of matching between the user and the segment is taken by the system as the confidence level that will be assigned to those concepts of the ontology.

4.2.3. Estimating the interests from users with similar demographic characteristics

Since the basic 100 segments provide only generic types of activities that tourists may be interested in visiting, it is necessary to have a way of obtaining more precise types of activities (lowest level concepts) and also particular activities to recommend. In order to do that, the system has to find out which are the users similar (from the demographic and motivational points of view) to the current one. Eq. (5), which takes into account the four aspects mentioned in Table 3, is used for this purpose. The idea is to evaluate periodical clusters of the full set of users stored in the database, in order to take into account the new users that are included in the system.

Thus, when a new user arrives, he/she first specifies his/her demographic data and travel motivations (Figs. 3 and 4). Then, given the current classification of users, the cluster that contains those users that have more similar characteristics is found. From that cluster, the system can compute, via Eq. (1), which is the score of each ontology concept, given the actions that have been performed on activities by the members of the cluster. The confidence level will be the product of the confidence obtained by Eq. (2) and the distance between the user and the cluster.

4.2.4. Estimating the interests from users with similar ratings

After the initial recommendation, the user may not be satisfied with the recommended activities because he/she does not fit exactly with the type of tourist that had a stronger matching with his/her demographic data and travel motivations. However, after the user interacts with the system and manipulates the current recommended activities, SigTur/E-Destination is able to perform more accurate recommendations, by finding other users that performed similar actions on the same activities (we call it ratings). In this step, the K-means clustering is also applied, although considering a different similarity measure and using as initial prototypes the users that have rated more items. Note that, in this case, the users with similar actions can have demographic values that are very different from the current one (different travel motivations, group composition, country of origin, etc.). The similarity between two users $u$ and $v$ (\(\log \text{Pearson} \)) to execute the clustering process is computed using the Frequency-weighted Pearson Correlation (Breese et al., 1998), where $\gamma_{u,v}$ are the activities that have been rated by both $u$ and $v$, $r_{ui}$ is the rating associated to the activity made by user $u$ on activity $i$, and $\text{mean}$ is the mean of the ratings of all actions made by user $u$:

$$\omega_{u,v} = \text{FPC}(u,v) = \frac{\sum_{i \in \gamma_{u,v}} r_{ui} - \text{mean}}{\sum_{i \in \gamma_{u,v}} (r_{ui} - \text{mean})^2}$$

In this expression there is a weight $\lambda_i$ for each activity, which has been included in order to increase the variety of the recommendations (Breese et al., 1998). As shown in Eq. (7), it takes into account the log-ratio of all users that have rated activity $i$ as well as the ratios of the actions of users $u$ and $v$ on activity $i$ with respect to all the actions they have performed. Thus, the activities that have been less frequently rated or have been less manipulated by all users have a higher relevance. On the other hand, activities rated more frequently by the two particular users $u$ and $v$ have a higher relevance. $U$ is the accumulated confidence level (CL) of all the actions of all the users in the system and $U_i$ is the accumulated CL of all users for a particular activity $i$. $R_u$ corresponds to all the CL values of user $u$ of actions done over activity $i$, and $R_u$ is the accumulated CL over all actions of such user. $R_u$ and $R_v$ are the same values for user $v$. The first factor in Eq. (7) was suggested in Breese et al. (1998), whereas the other two factors are novel.

$$\hat{\lambda}_i = \log \frac{|U|}{|U_i|} \hat{\lambda}_{1i} \exp \left( \frac{|R_u|}{|R_v|} \hat{\lambda}_{2i} \exp \left( \frac{|R_u|}{|R_v|} \hat{\lambda}_{3i} \right) \right)$$

Once the ratings-based clustering has been obtained, the process to follow is the same one that was explained in the previous section on the demographic-based clustering. The system finds the cluster that is more similar to the current user, and it can then calculate the preference level and confidence level on the concepts of the ontology, from the actions done by the users in the most similar cluster and the distance between the user and this cluster (Eqs. (1) and (2)).

4.3. Integration of all the recommendation techniques

In the previous sections we have presented several methodologies that can be used to discover which classes of activities are interesting for the user. For each methodology, the system has computed a preference value (with a certain confidence) for each concept of the ontology. More specifically, the following aspects have been taken into account:

1) The user's motivations (Section 4.1.1, Fig. 3).
2) The actions of the user over the activities shown by the system (Section 4.1.2, Figs. 6–8).
3) The explicit rating of the activities that have been visited (Section 4.1.2, Fig. 5).
4) The bottom-up propagation of the interest and confidence values over the ontology (Section 4.1.3).
5) The similarity between the user and the predefined tourist segments (Section 4.2.2).
6) The similarity between the user and clusters of users with similar demographic characteristics (Section 4.2.3).
7) The similarity between the user and clusters of users with similar ratings (Section 4.2.4).

With all these data, the system can compute a final preference value for the user $u$ for each concept $c$ of the ontology, with a certain confidence level. The preference value is the mean of the preference values obtained with the previous methods, and the confidence level
is the weighted mean of the respective confidence levels:

\[
CL_u(c) = \frac{\sum_{i=1}^{7} CL_u(c)}{7}
\]  

(8)

\[
S_u(c) = \frac{\sum_{i=1}^{7} S_u(c)CL_u(c)}{\sum_{i=1}^{7} CL_u(c)}
\]  

(9)

In order to get the final assessment of the interest in each ontology concept, the interest values are propagated down the ontology hierarchy. Thus, each node receives the contributions from its ancestors (e.g., if a user is interested in Museums, this interest will be transferred to the different kinds of museums available in the ontology). However, the interest in one class of activities cannot be translated directly into interest in all its subclasses (a user that is very interested in Sports will probably not be very interested in all kinds of sports); therefore, the downwards propagation is conditioned on the hierarchy depth (the longer we move from a node, the smaller is its influence). For each node concept \( c \) the final interest score \( S^0_u \) is given by

\[
S^0_u(c) = \frac{S^0(c)CL_u(c) + \sum_{p < c} S^0(p)CL_u(p)D(p,c)}{CL_u(c) + \sum_{p < c} CL_u(p)D(p,c)},
\]  

(10)

where \( S^0(c) \) are all the concepts that are ancestors of \( c \), and \( D(p,c) \) is a function that depends on the distance between nodes \( p \) and \( c \), which is computed as

\[
D(p,c) = Q(\text{number-of-levels-between}(p,c)),
\]  

where \( Q(0) = 1 \) and \( Q(l) = Q(l-1) - (z \times Q(l-1)) \)  

(11)

In this expression \( z \) is the weight factor, which has been empirically set to 0.2.

Once the system obtains the final score for each ontology concept, it rejects the ones that do not reach a minimum confidence level. Then, the system retrieves all the activities that are related to the remaining concepts and sets a preference score \( (S^0(a)) \) and a confidence value \( (CL_u(a)) \) for each activity \( a \), by averaging the values of the concepts associated to the activity.

A first filter on activities is applied, based on the price criterion. In the initial form the user has indicated the budget of the trip (Fig. 4), with a real value between 0 (low-cost trip) and 1 (luxury trip). The activities that are deemed to be too expensive for the intended budget of the user are removed at this stage.

After setting the preference and confidence values for each activity, the system must sort the recommendation list in order to provide the most interesting ones at the top. We have defined some heuristic rules in order to determine if one activity \( (a_1) \) must be above or below another activity \( (a_2) \) in the list. The rules are executed in the following order:

1- If \( CL_u(a_1) \) is high and \( CL_u(a_2) \) is low, then \( a_1 \) will be above \( a_2 \).
2- If \( S^0(a_1) \) is at least 0.01 points higher than \( S^0(a_2) \), then \( a_1 \) will be above \( a_2 \).
3- The activity that is closer to the user’s location will be above in the list if the difference between activities is higher than 3 km.

Fig. 9. Travel planner of activities.
4- If none of the previous rules have been applied, the activities are randomly ordered.

The order of the rules has been determined to try to avoid the fact that activities that have not been considered by the users (e.g., off-the-track or new activities) are hidden by the most popular activities. Therefore, these activities will appear randomly, until a certain amount of users has rated them. This final step is related to the fact that some authors claim that the recommendations ought to be diverse enough to cover different kinds of activities, so that the user is not offered a big set of very similar alternatives (Ziegler et al., 2005).

Once the list of N recommended activities has been obtained, the system represents them on a map with which the user can interact. It is possible to request N new activities at any moment. The user may select those activities that he/she wishes to visit and add them to a planning list. This list can be arranged by the user in different days, and the daily route between activities can be visualized (Fig. 9). This list can be printed or downloaded with the whole information of each activity and the directions to follow the route.

5. Validation

The validation of the system has been made from two different perspectives. The first one is to analyze the results obtained taking into account several stereotypes of tourists. The second one, more general, takes into account the whole system as a recommender product and analyzes the feedback received from several users during a public presentation at the meeting FITUR-2011. The region of “Costa Daurada and Terres de l’Ebre” is one of the main tourist destinations of Spanish coasts and the Mediterranean Sea. Its location provides a rich and varied region with an extensive coastline and an inner region with quiet villages and fields. Besides sun and sand, the region offers a wide variety of leisure, culture, nature and history. The region area is 6303 km² and the longest road distance from the bottom side to the top side is about 150 km. SigTur/E-Destination is enriched with 1300 diversified activities spread out over the territory.

5.1. Analysis with stereotypes

We have tested the recommender system with the simulation of four distinct tourist stereotypes, analyzing their profiles and the recommendations produced by the system. The stereotypes that have been considered are some of the most common tourist profiles that visit our region during the whole year. We have analyzed the degree of correspondence between the main motivations of each user and the recommendations provided by the system.

Table 4 represents the demographic and travel data related to the profile of each different user stereotype (country of origin, travel group composition, accommodation, destination city and dates of the travel).

Table 5 shows the degree of interest in each of the nine available motivations, with a percentage between 0 and 100. Tables 6–9 show the specific activities recommended to each tourist group. For each activity the table shows its name, direct concepts of the ontology to which it is associated, top-level ontology concepts that are superclasses of these direct concepts (and that correspond to the travel motivations selected in the stereotype) and the distance from the destination city to the location of the activity. The average distance of all recommended activities to the destination is given in Table 10.

To evaluate the quality of the recommendations, we have calculated the correspondence between two vectors: the first vector is given by the user motivations (Table 4) and the second is obtained through the analysis of the motivations represented by the recommended activities (column “Related motivation” on Tables 6–9). The process is to sort the first vector based on the user values and sort the second vector based on the number of recommended activities associated to each motivation. After that, a Spearman correlation is computed to evaluate if these two vectors are well correlated. The result values can be from –1 to 1, where –1 denotes a negative correlation, 0 a null correlation and 1 a positive correlation. The result values for each stereotype are shown in Table 11.

Let us now analyze the recommendation obtained for each stereotype. The first profile is a German senior group that is mainly interested in shopping and gastronomy events, and also has a small interest in relaxation and culture. They want to visit during one week in November the capital of the region (Tarragona), staying at a luxury hotel. The city offers a wide variety of cultural, gastronomical and commercial activities. The majority of the recommended activities are very close to the destination except a gastronomical event. The reason is that in the travel dates there are not any specific shows or events in Tarragona. The correlation between the recommended activities and the user motivations is 0.83, which shows that the recommended activities highly correspond to the user interests (for instance, four of the recommended activities are related to Shopping, which is the main motivation of these tourists) (Table 12).

Table 5
Demographic and travel data of stereotypes.

<table>
<thead>
<tr>
<th>User</th>
<th>Origin</th>
<th>Travel group</th>
<th>Accommodation</th>
<th>Destination</th>
<th>Trip period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Germany</td>
<td>Senior group</td>
<td>4–5 stars hotel</td>
<td>Tarragona</td>
<td>20/11/11–27/11/11</td>
</tr>
<tr>
<td>3</td>
<td>Spain</td>
<td>With children 6–12 yrs.</td>
<td>3 stars hotel</td>
<td>Cambrils</td>
<td>10/8/11–20/8/11</td>
</tr>
<tr>
<td>4</td>
<td>Spain</td>
<td>Couple &gt; 35 yrs.</td>
<td>4–5 stars hotel</td>
<td>Tarragona</td>
<td>5/3/11–10/3/11</td>
</tr>
</tbody>
</table>

Table 6
Interest value of the motivations selected by the stereotypes.

<table>
<thead>
<tr>
<th>User</th>
<th>Beach (%)</th>
<th>Shopping (%)</th>
<th>Relaxation (%)</th>
<th>Theme/Leisure parks (%)</th>
<th>Culture (%)</th>
<th>Nature (%)</th>
<th>Gastronomy (%)</th>
<th>Sports (%)</th>
<th>Shows/Events (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>60</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>69</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>58</td>
<td>31</td>
<td>7</td>
<td>93</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>34</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>16</td>
<td>53</td>
<td>0</td>
<td>75</td>
</tr>
</tbody>
</table>
### Table 7
Recommended activities for stereotype 1.

<table>
<thead>
<tr>
<th>Name</th>
<th>Concepts</th>
<th>Related motivation</th>
<th>Distance (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>El Corte Inglés</td>
<td>ShoppingCenter</td>
<td>Shopping</td>
<td>0.42</td>
</tr>
<tr>
<td>Mostra de l’Oli Siurana</td>
<td>FoodEvents</td>
<td>Gastronomy, Shows/Events</td>
<td>0.00</td>
</tr>
<tr>
<td>Venus Wellness Center</td>
<td>HealthResorts, HydroFacilities</td>
<td>Relaxation</td>
<td>0.66</td>
</tr>
<tr>
<td>Les Gavares</td>
<td>ShoppingArea</td>
<td>Shopping</td>
<td>4.71</td>
</tr>
<tr>
<td>Muralles passeig arqueològic</td>
<td>HistoryMuseums, HumanHeritage, Roman, Walls</td>
<td>Culture</td>
<td>0.59</td>
</tr>
<tr>
<td>Festa de l’oli nou DOP Siurana</td>
<td>FoodEvents</td>
<td>Gastronomy, Shows/Events</td>
<td>0.00</td>
</tr>
<tr>
<td>Golden Port Salou &amp; Spa</td>
<td>HealthResorts, HydroFacilities</td>
<td>Relaxation</td>
<td>10.35</td>
</tr>
<tr>
<td>Parc Central</td>
<td>ShoppingCenter</td>
<td>Shopping</td>
<td>0.77</td>
</tr>
<tr>
<td>Tasta Porrera</td>
<td>WineEvents</td>
<td>Gastronomy, Shows/Events</td>
<td>39.74</td>
</tr>
<tr>
<td>Mercat de Constanti</td>
<td>LocalMarket</td>
<td>Shopping</td>
<td>2.74</td>
</tr>
</tbody>
</table>

### Table 8
Recommended activities for stereotype 2.

<table>
<thead>
<tr>
<th>Name</th>
<th>Concepts</th>
<th>Related motivation</th>
<th>Distance (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cala de Llenguadets</td>
<td>Coves,UrbanBeaches</td>
<td>Beaches</td>
<td>1.21</td>
</tr>
<tr>
<td>Carrer Buigas – Saragossa</td>
<td>ShoppingArea</td>
<td>Shopping</td>
<td>0.44</td>
</tr>
<tr>
<td>Cala Granès</td>
<td>Coves</td>
<td>Beaches</td>
<td>3.14</td>
</tr>
<tr>
<td>Focs artificialis</td>
<td>TraditionalCelebrations</td>
<td>Shows/Events</td>
<td>0.97</td>
</tr>
<tr>
<td>Vela pet litoral Costa Daurada</td>
<td>Boating</td>
<td>Sports</td>
<td>1.34</td>
</tr>
<tr>
<td>Passeig marinít de Salou</td>
<td>ShoppingArea</td>
<td>Shopping</td>
<td>0.82</td>
</tr>
<tr>
<td>La Pineda, D’ocresilecs</td>
<td>Discos</td>
<td>–</td>
<td>3.20</td>
</tr>
<tr>
<td>Plata de Vilafortuny</td>
<td>NormalBeaches, UrbanBeaches</td>
<td>Beaches</td>
<td>4.63</td>
</tr>
<tr>
<td>El Vendrell,Center Comercial Les Mates Espai Lúdic</td>
<td>Discos, Bars, Pub</td>
<td>–</td>
<td>42.65</td>
</tr>
<tr>
<td>Cala de la Roca Plana</td>
<td>Coves, NaturalSpaces</td>
<td>Beaches, Nature</td>
<td>19.64</td>
</tr>
<tr>
<td>Port Harleym</td>
<td>ShoppingCenter</td>
<td>Shopping</td>
<td>2.03</td>
</tr>
<tr>
<td>Plata de l’Estany Salat</td>
<td>NormalBeaches, AquaticSports</td>
<td>Beaches, Sports</td>
<td>20.10</td>
</tr>
</tbody>
</table>

### Table 9
Recommended activities for stereotype 3.

<table>
<thead>
<tr>
<th>Name</th>
<th>Concepts</th>
<th>Related motivation</th>
<th>Distance (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Port Aventura Park</td>
<td>ThemePark</td>
<td>Theme/Leisure parks</td>
<td>10.43</td>
</tr>
<tr>
<td>Plata de la Riera</td>
<td>NormalBeaches, UrbanBeaches</td>
<td>Beaches</td>
<td>0.34</td>
</tr>
<tr>
<td>Port de Cambrils</td>
<td>ShoppingArea</td>
<td>Shopping</td>
<td>0.50</td>
</tr>
<tr>
<td>Costa Canibe Port Aventura</td>
<td>WaterPark</td>
<td>Theme/Leisure parks</td>
<td>9.92</td>
</tr>
<tr>
<td>Belles Aigues Spa</td>
<td>SpaResorts</td>
<td>Relaxation</td>
<td>0.21</td>
</tr>
<tr>
<td>Plata del Cap de Sant Pere</td>
<td>FamilyBeaches, UrbanBeaches</td>
<td>Beach</td>
<td>4.62</td>
</tr>
<tr>
<td>Verge del Carme</td>
<td>TraditionalCelebrations</td>
<td>Shows/Events</td>
<td>0.69</td>
</tr>
<tr>
<td>Carrer Barcelona</td>
<td>ShoppingArea</td>
<td>Shopping</td>
<td>7.59</td>
</tr>
<tr>
<td>Plata de la casa dels lladres</td>
<td>NormalBeaches, AquaticSports</td>
<td>Beaches</td>
<td>10.60</td>
</tr>
<tr>
<td>Plata de les Muntanyans</td>
<td>NormalBeaches, NaturalSpaces</td>
<td>Beaches</td>
<td>37.92</td>
</tr>
<tr>
<td>Aquatònic Espai Lúdic Termal</td>
<td>HealthResorts, HydroFacilities</td>
<td>Relaxation</td>
<td>7.25</td>
</tr>
<tr>
<td>Port Harley</td>
<td>ShoppingCenter</td>
<td>Shopping</td>
<td>2.03</td>
</tr>
</tbody>
</table>

### Table 10
Recommended activities for stereotype 4.

<table>
<thead>
<tr>
<th>Name</th>
<th>Concepts</th>
<th>Related motivation</th>
<th>Distance (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muralles passeig arqueològic</td>
<td>HistoryMuseums, HumanHeritage, Roman, Walls</td>
<td>Culture</td>
<td>0.59</td>
</tr>
<tr>
<td>Diada i IV Mostra del Bull</td>
<td>FoodEvents</td>
<td>Gastronomy, Shows/Events</td>
<td>15.40</td>
</tr>
<tr>
<td>Via T</td>
<td>ShoppingArea</td>
<td>Shopping</td>
<td>0.87</td>
</tr>
<tr>
<td>Fòrum de la colonia</td>
<td>HistoryMuseums, HumanHeritage, Roman</td>
<td>Culture</td>
<td>0.48</td>
</tr>
<tr>
<td>Muntanyans de Prades</td>
<td>NaturalSpaces</td>
<td>Nature</td>
<td>28.18</td>
</tr>
<tr>
<td>Tarragona, Pub i bars de la Part Alta</td>
<td>Bars, Pub</td>
<td>–</td>
<td>1.05</td>
</tr>
<tr>
<td>Casino Tarragona (Rambla Vella)</td>
<td>GameRoom</td>
<td>–</td>
<td>1.20</td>
</tr>
<tr>
<td>Aqueduecte de les Ferreres o Pont del Diable</td>
<td>Aqueduct, HumanHeritage, Roman</td>
<td>Culture</td>
<td>2.74</td>
</tr>
<tr>
<td>Mostra Gastronómica de l’Associació de Restauradors de Torredembarra</td>
<td>FoodEvents</td>
<td>Gastronomy, Shows/Events</td>
<td>15.40</td>
</tr>
<tr>
<td>Parc Central</td>
<td>ShoppingCenter</td>
<td>Shopping</td>
<td>0.77</td>
</tr>
<tr>
<td>De Cabra del Camp al coll de Romiguieres i les serres del Cogulló</td>
<td>MountainRoutes, RuralRoutes, Trekking</td>
<td>Nature</td>
<td>29.73</td>
</tr>
<tr>
<td>Mercat Central de Tarragona</td>
<td>LocalMarket</td>
<td>Shopping</td>
<td>0.45</td>
</tr>
</tbody>
</table>
The second tourist profile that has been considered is a group of young English friends that visit Salou in summer, staying in an apartment. Their main motivations are going to the beach, shopping and attending events. They also have a small interest in sports. The correlation between their motivations and the recommended activities is also very high, 0.8. It must be said that some of the recommended activities do not match directly with the user motivations, such as discos, but since these activities have been obtained through the collaborative mechanism (i.e., taking into account what other groups of young English friends made in Salou), which provides a better serendipity in the recommendation process. Nine of the twelve recommended activities are within walking distance from Salou.

The next stereotype represents a Spanish family that travels with children to a three stars hotel in Cambrils during the summer. Their main motivations are going to the beach and visiting theme and leisure parks. Moreover, they also express some interest in shopping, events and relaxation. Most of the recommended activities are quite close to the destination, except for the leisure parks (all of them are at a certain distance from Cambrils) and two beaches. The correlation value, 0.87, is the best one obtained among the different stereotypes.

The last simulated stereotype is a couple of Spanish adults that visits Tarragona in March staying in a four stars hotel. Their main motivations are culture and events, but they are also somewhat interested in gastronomy, shopping and nature. In this case the correlation value goes down to 0.7, because the collaborative methods influence the final recommendation. The recommended activities are close to Tarragona, except four that are not located in the city (two walks in natural spaces and two gastronomical events).

### 5.2. Feedback from real users

During the last FITUR’11 conference (International Conference of Tourism) held in Madrid, the SigTur/E-Destination project was presented to both professionals and non-professionals. Users interested in the product were briefed with the main features of the system as well as with basic notions about its use. Then they spent several minutes with the system adding their personal interests and then surfing through the obtained results. At the end of the act, a supervisor requested them to fill a questionnaire to evaluate their opinion about the system. As the questionnaire was freely filled by interested users in a very controlled setting, it was not necessary to include redundant or contradictory questions to assess the consistency of the answers.

Tables A.1 and A.2 in the annex show the whole questionnaire (questions and allowed values) used in the evaluation. At the end of 5 day, 78 forms were collected: 28 from Tourism professionals, and 50 from end users. Fig. 10 summarizes the main results obtained from the stakeholders. Two important conclusions from this evaluation were extracted: a recommender system able to acquire the preferences of the user is interesting and useful for tourists, and a Web-based approach is an appropriate option for this type of systems.

In more detail, most of the users reported a positive experience after its use. Concretely, more than 80% of those that were surveyed thought that the system is interesting and useful to know a particular region. Only 20% thought that the system is not useful to get information about destinations. Concerning the general perception of the system, more than 90% confirmed that the results of the recommender are accurate enough to be used to plan their holidays. More concretely, 24% of those that were asked would delegate planning the whole trip to the recommender, whereas 72% thought that this type of system is a good complement to the planning of a trip.

Internet was confirmed as the main source used to plan trips, delegating to a second term other sources such as travel agencies, specialized journals and books. Thus, a Web-based application seems a very good option for the implementation of a recommender system. Concerning the moment in which the recommender can be used, surveyed people thought that the best option is to utilize it before the trip. However, almost 40% said that it could also be employed during the trip. According to this second answer, it could be interesting to implement a mobile version able to run in smart phones in order to facilitate the use both before and during the trip.

Concerning the satisfaction with the usability and the obtained results, both items were well rated by respondents with 8/10 points in average.

Finally, concerning some general aspects of the application, the obtained results were also satisfactory (rates above 4/5).

Thanks to this feedback, the SigTur/E-Destination system has been improved during the last year. One of the main aspects that has been improved is the interface, giving always information about the whole process of recommendation, expanding the information about the recommended activities and including more (and more diverse) activities to the database.

### 6. Discussion and conclusions

Recommender systems can be an important tool in the provision of personalized advice to the visitors of a destination, making them aware of activities that are not the main focus of attraction and improving the chances of a better tourist flow and a more sustainable management. The Web-based interface of the presented system allows to plan the activities before or during the trip in a user-friendly graphical environment.

From the technical point of view, the development of SigTur/E-Destination has required a strong use of a wide set of Artificial Intelligence methodologies and tools. On the knowledge management side, a specific domain ontology provides a classification of the main types of activities and guides the knowledge-level inference process needed to assess the preferences of the user on each of them. Concerning the employees recommendation techniques, the system considers as much as information as possible to provide an accurate recommendation, including demographic and travel data, trip motivations, user stereotypes, the actions of the users on the platform, classes of users with similar tastes or demographic attributes, etc. The final recommendations have a good diversity and match quite nicely with the main motivations of the user. In terms of usability, this
The application has achieved the user-friendliness needed to reach non-experts users. The representation of the maps using Google Maps API makes it more familiar to the users, as the use of Street View and Directions Google services. Moreover, the ICEfaces framework feeds the website with rich components that allows user responses without full-page refreshes, thus giving to the user the experience of almost being use a desktop application.

Concerning the future work, the Tourism stakeholders in the territory feel that it could be interesting to make a new version of the system that could run on mobile devices, so that the utility of the system for visitors that are already in the Tarragona area is enhanced.

Acknowledgments

SigTur/E-Destination is a project funded by the FEDER European Regional Funds and the Government of the Province of Tarragona. Part of the work has been supported by the Universitat Rovira i Virgili (predoctoral grant of L.Marin), the Spanish Ministry of Science and Innovation (DAMASK project, Data mining algorithms with semantic knowledge, TIN2009-11005) and the Spanish Government (PlanE, Spanish Economy and Employment Stimulation Plan). The authors would like to acknowledge the extremely valuable contributions to the design and development of the system made by Salvador Anton, Paolo Russo, Alicia Orellana, Yolanda Pérez and Joan de la Flor.

Appendix A. Full questionnaire employed in FITUR-2011

See Tables A.1 and A.2.
Table A.1
Questions and allowed answers of the questionnaire.

<table>
<thead>
<tr>
<th>Id.</th>
<th>Question</th>
<th>Allowed answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What do you think about the application you just used?</td>
<td>Free answer</td>
</tr>
<tr>
<td>2</td>
<td>If this application was on the market, would you use for planning your holidays?</td>
<td>Yes/No do not know</td>
</tr>
<tr>
<td>2a</td>
<td>If the answer to question 2 was “Yes”, Which would be the use? What would be the advantages for you?</td>
<td>Free answer</td>
</tr>
<tr>
<td>2b</td>
<td>If the answer to question 2 was “No”, Why would not you use it? What disadvantages would it have?</td>
<td>Free answer</td>
</tr>
<tr>
<td>2c</td>
<td>If the answer to question 2 was “I do not know”, What aspects make you hesitate?</td>
<td>Free answer</td>
</tr>
<tr>
<td>3</td>
<td>Rate from 1 to 10 how useful is the application to organize activities during your holidays. (1 is the minimum rate, and 10 the maximum).</td>
<td>1–10</td>
</tr>
<tr>
<td>4</td>
<td>Rate from 1 to 10 how satisfied you are with the activities that have been proposed in this trial</td>
<td>1..10</td>
</tr>
<tr>
<td>5</td>
<td>How many of the activities that have been proposed are interesting for you?</td>
<td>One of these values: all/ almost all/ half of them/ almost none/ none</td>
</tr>
<tr>
<td>6</td>
<td>Concerning the obtained recommendations, have you missed any activity that could be interesting for you?</td>
<td>Yes/ No</td>
</tr>
<tr>
<td>6a</td>
<td>If the answer to question 6 was “Yes”, give examples.</td>
<td>Free answer</td>
</tr>
<tr>
<td>7</td>
<td>Referring to the information about the activities, have you missed any particular piece of data that could be interesting for you?</td>
<td>Yes/ No</td>
</tr>
<tr>
<td>8</td>
<td>How would you rate the following aspects of the application?</td>
<td></td>
</tr>
<tr>
<td>8a</td>
<td>Easy to use</td>
<td></td>
</tr>
<tr>
<td>8b</td>
<td>Time needed to get recommendations</td>
<td></td>
</tr>
<tr>
<td>8c</td>
<td>Look and interface</td>
<td></td>
</tr>
<tr>
<td>8d</td>
<td>Variety of the proposed activities</td>
<td></td>
</tr>
<tr>
<td>8e</td>
<td>General usability</td>
<td></td>
</tr>
</tbody>
</table>

Table A.2
Questions and allowed answers of the questionnaire (continued).

<table>
<thead>
<tr>
<th>Id.</th>
<th>Question</th>
<th>Allowed answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Usually, where do you find the information for planning the activities during your vacation? (Source)</td>
<td>One or more of these items: Internet, tourism offices, travel agency, books, journals, radio and TV, I like to improvise, recommendation from family/friends, and other sources</td>
</tr>
<tr>
<td>10</td>
<td>Regarding the organization of a journey, when could this type of system be useful for you?</td>
<td>One of the following items: before the trip, during the trip to get particular activities, both, neither, because I like to improvise</td>
</tr>
<tr>
<td>11</td>
<td>Would you use a recommender like this to plan your vacations?</td>
<td>One of these values: Sure/I would use it with some doubts/No</td>
</tr>
<tr>
<td>11a</td>
<td>If the answer to question 11 was “I would use it with some doubts”, Why?</td>
<td>Free answer</td>
</tr>
<tr>
<td>11b</td>
<td>If the answer to question 11 was “No”, Why not?</td>
<td>Free answer</td>
</tr>
<tr>
<td>12</td>
<td>If this tool was available on the market, would you recommend it to a friend?</td>
<td>One of these values: absolutely yes, maybe, No</td>
</tr>
<tr>
<td>12a</td>
<td>If the answer to question 12 was “Maybe”, explain the reason</td>
<td>Free answer</td>
</tr>
<tr>
<td>12b</td>
<td>If the answer to question 12 was “No”, explain the reason</td>
<td>Free answer</td>
</tr>
</tbody>
</table>

References


