UNIVERSITAT ROVIRA I VIRGILI

SEMANTIC RECOMMENDER SYSTEMS
PROVISION OF PERSONALISED INFORMATION ABOUT TOURIST ACTIVITIES.

Joan borràs Nogués

Dipòsit Legal: T 1339-2015

Joan Borràs Nogués

SEMANTIC RECOMMENDER SYSTEMS

PROVISION OF PERSONALISED INFORMATION ABOUT TOURIST ACTIVITIES

Ph.D. Thesis

Supervised by

Dr. Antonio Moreno and Dr. Aïda Valls

Department of Computer Science and Mathematics

UNIVERSITAT ROVIRA I VIRGILI

May, 2015

UNIVERSITAT ROVIRA I VIRGILI

WE STATE that the present study, entitled "**Semantic Recommender Systems - Provision of personalised information about tourist activities**", presented by Joan Borràs Nogués for the award of the degree of Doctor, has been carried out under our supervision at the Department Computer Engineering and Mathematics of this university.

Tarragona, 27th, April 2015

Doctoral Thesis Supervisors:

Dr. Antonio Moreno Ribas

Dr. Aida Valls Mateu

Acknowledgments

This work has been supported by the Spanish research projects DAMASK (Data Mining Algorithms for Semantic Knowledge, TIN2009-11005), SHADE (Semantic and Hierarchical Attributes in Decision Making, TIN2012-34369), and the collaboration agreement between the ITAKA research group and the Science & Technology Park for Tourism and Leisure in the area of technological applications for Tourism mobility management (SIGtur/E-Destination project funded by the FEDER European Regional Funds and the Government of the Province of Tarragona).

I would like to thank Dr. Salvador Anton and Dr. Antonio Paolo Russo for their expert advice in the Tourism domain and also for giving me the opportunity to work in such an innovative and motivating project. I also thank the directors of this thesis, Dr. Antonio Moreno and Dr. Aïda Valls for their guidance and advice on the development of the technological process.

Last but not least, I also want to thank my family, specially my wife for the stolen time and for encouraging me in the stressful moments.

Abstract

The technological revolution that enables the distribution and utilization of information by anyone has lead to an exponential increase on the information available on Internet. Such volume of information makes tedious the task of finding appropriate and relevant information throughout the huge available content.

Tourism is one of the sectors more affected by this fact, which has dramatically changed the way of travelling. Not long ago travellers visited places checking out information from book guides, paper maps or travel agencies. Currently, they can find any kind of information about touristic places with a simple click on the Web. However, as the available information about a touristic place is overwhelming, travellers planning trips have difficulties in order to seek and choose the most appropriate plan. Recommender systems can be used to overcome these problems by offering personalised information based on tourists' preferences.

This work studies how new improvements can be made on recommender systems using ontological information about a certain domain, in this case the Tourism domain. Ontologies define a set of concepts related to a certain domain as well as the relationships among them. These data may be used not only to represent in a more precise and refined way the domain objects and the user preferences, but also to apply better matching procedures with the help of semantic similarity measures. The improvements at the knowledge representation level and at the reasoning level lead to more accurate recommendations and to an improvement of the performance of recommender systems, paving the way towards a new generation of *smart semantic recommender systems*. Both content-based recommendation techniques and collaborative filtering ones may certainly benefit from the introduction of explicit domain knowledge.

In this thesis we have also designed and developed a recommender system that applies the methods we have proposed. This recommender is designed to provide personalised recommendations of touristic activities in the region of Tarragona. The activities are properly classified and labelled 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. A diversification process that computes similarities between objects is applied to produce diverse recommendations and hence increase user satisfaction. This system can lead to benefits in the impact of the region by improving the experience of its visitors.

Contents

List of figures

List of tables

Chapter 1 – Introduction

The last 10 years have witnessed an enormous increase in the amount of information available through Internet, caused mainly by the advent of the Social Web, which has transformed users from mere consumers of data into avid producers of up-to-date information, comments and opinions about every conceivable domain. Figure 1 is a snapshot of a website[1] that displays some interesting data (in real time) about the volume of information generated on the Web. In only 60 seconds it is possible to notice the huge volume of data generated through the most popular social platforms and applications (e.g. 3M Facebook posts, 342K tweets, 1K blog posts on Wordpress and 270K Google queries). This is just an example of how many data are being continuously transferred through Internet. A recent report written by CISCO (CISCO, 2014) predicts that the Internet traffic will nearly triple from 2013 to 2018, moving from 51 exabytes transferred per month to 132 (see Figure 2). All this amount of information available online clearly exceeds the cognitive capacity of users, which are constantly looking for the most appropriate and relevant information. Web search engines can return the most popular Web pages associated to a given query, but users are left with the daunting task of refining the search or manually exploring a large number of Web pages in order to come across the precise information they were looking for.

Figure 1. Amount of data generated in Internet in real time

[1] http://pennystocks.la/internet-in-real-time/ (last access February 2015)

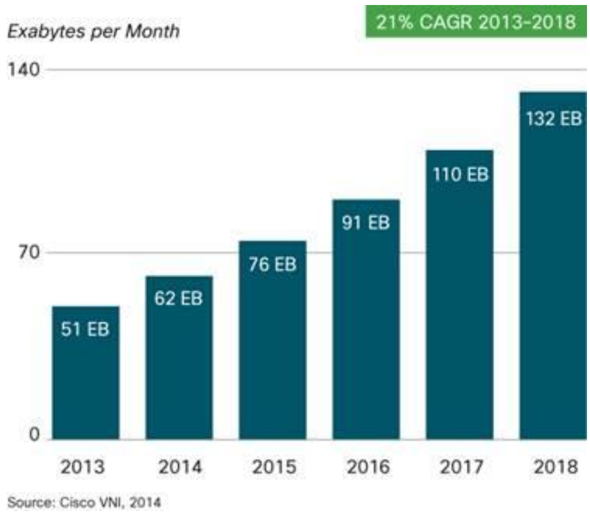


Figure 2. Cisco VNI Forecasts 132 Exabytes per Month of IP Traffic by 2018

In this scenario, intelligent decision support tools (and, in particular, *Recommender Systems*, (Adomavicius and Tuzhilin, 2005)) have appeared to help users to find the information they need to make their daily decisions efficiently. These systems, which provide customized information to users based on their preferences, restrictions, characteristics or tastes, have brought about a new era of personalized information on all domains (Gao et al., 2010). All retailers, regardless of whether they sell books, movies or computers, try to provide to each user the information about the products that match exactly his/her needs, filtering irrelevant options, saving the user from the task of analysing millions of alternatives and increasing not only the sales but also the satisfaction of the clients. *Amazon.com* was one of the first companies that offered customized information to its users, increasing their book sales (Schafer et al., 1999).

The Tourism sector has been largely affected by this growth of information, since it was one of the pioneers in the adoption and development of applications based on *Information and Communication Technologies* (ICT), driven by the explosive increase of the use of mobile and portable devices. In the last decade there has been a tremendous change in the way in which travellers prepare their trips. As shown in Figure 3, 74% of leisure travellers use Internet as a source of information for travel planning (Google, 2014). They spend a significant amount of time online exploring alternatives on what to do, where to sleep or where to eat in a given destination. A study from Expedia says travellers visit around 38 sites before booking a vacation (Expedia, 2013), going from 2.5 sites five weeks before the trip to 15.5 sites during the travel week (see Figure 4). Another study (TripAdvisor, 2014) says that 67% of travellers check TripAdvisor a few times a month, 53% of travellers will not book a hotel until they have read the reviews of the previous clients, and more than 80% of TripAdvisor users say that they use this platform since it helps them to have a better trip and choose the right hotel. From these figures it can be concluded that a large proportion of travellers are intensely technology-dependent in the preparation of their upcoming travel experiences.

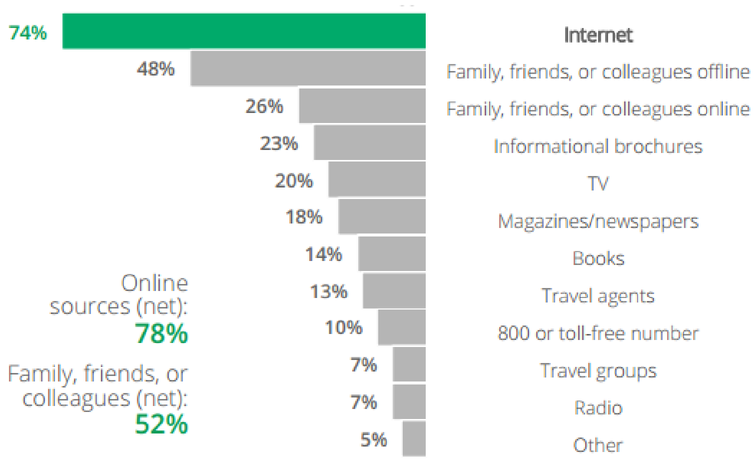


Figure 3. Travel planning sources for leisure travel (from (Google, 2014))

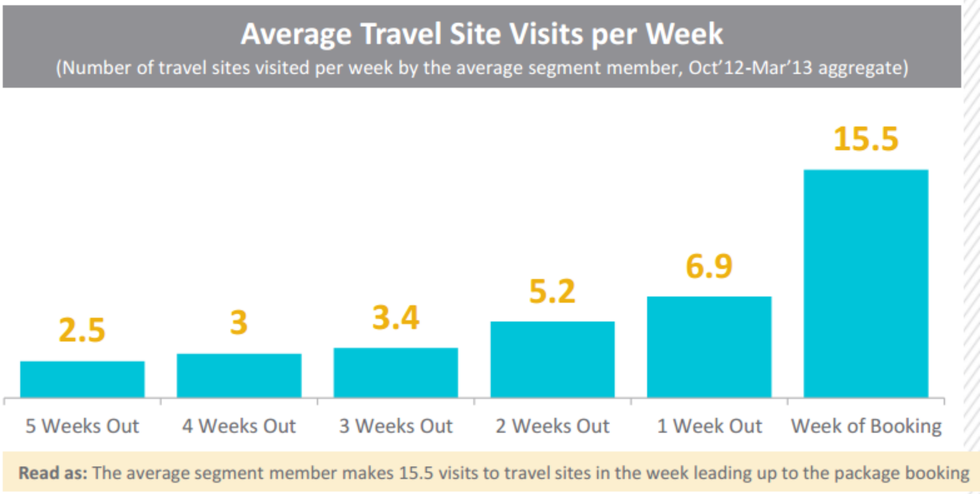


Figure 4. Average travel sites visit per week (from (Expedia, 2013))

Travel planning, which is a specific stage of the travel cycle, is an activity that directly impacts the quality of the final experience of the user. Thus, travellers devote a significant amount of time to gather the information they need to make the appropriate decisions in this task, such as choosing the destination, deciding the means of transport or selecting an accommodation (Gursoy and McCleary, 2004; Gretzel et al., 2006). This information-seeking behaviour prior to the user's decision making has brought up new opportunities for *Destination Management Organisations* (DMO) and travel companies, which try to engage the potential tourists into looking up their Web sites or using their mobile applications in the process of planning a trip to a region. These efforts by all the actors involved in destination management have led to a huge increase in the amount of touristic information available online. In a single specialised website such as TripAdvisor it is possible to find more than 4.5 million businesses and properties with more than 200 million reviews from travellers[2], so the whole

[2] http://www.tripadvisor.com/PressCenter-c4-Fact_Sheet.html (last accessed February 2015)

volume of Tourism information available on the Web is beyond imagination.

Ironically, the more information a traveller has, the more difficult and time-consuming it is to retrieve it, analyse it and use it to plan a specific trip. First, users must explore manually a large number of Web sites with accurate, up-to-date and trustable information, and then they must select from all that information which are the data that they really need to take their decisions. In the Tourism field *Travel Recommender Systems* (Ricci, 2002) try to match the characteristics of attractions with the user's needs. These systems are emerging as important tools in the development and management strategies of destinations, as local stakeholders are interested in promoting the global attractiveness of a particular region, especially of those activities that are less popular. An efficient design, organization and communication of opportunities in the region may lead to a more spatially, thematically and financially balanced tourism activity, with important returns in terms of sustainable development. 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 the users' profile. That relieves the users from having to manually evaluate all the possible choices and helps to avoid judgement mistakes when comparing the available alternatives. It is also important to note that the recommendation of touristic places is highly related to the spatial distribution of places and visitors. Therefore, it can be claimed that the combination of *Artificial Intelligence* (AI) techniques and *Geographic Information Systems* (GIS) within 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, 2002).

Recommender Systems use several methods to provide personalized information such as collaborative filters (Kruszyk *et al.*, 2007), content filters (Pazzani and Bilnius, 2007) and the use of stereotypes based on socio-demographic data (the use of these methods in Tourism recommenders will be commented in more detail in the next chapter). The idea of *collaborative filtering* is to make recommendations based on what similar users have visited and their level of satisfaction. *Content filters* generate recommendations based on the user's preferences. The objective is to find those places that fit better with such tastes. The *socio-demographic methods* analyse the user's basic demographic data to associate him/her to a predefined stereotype, for which preferred attractions are known in advance. In many cases recommender systems do not only take into account the preferences of the tourist but they also analyze the *context* in which the recommendation takes place (Dey and Abowd, 1999). This is especially useful when tourists are already at the destination and they are willing to use their mobile devices to customize their trips in real time. Each of these methods has its own pros and cons, so it is common to create *hybrid* systems that try to combine them in order to obtain a better overall performance (Burke, 2002).

An important current line of research is the enhancement of recommendations with *semantic* domain knowledge (Wang and Kong, 2007). A *semantic recommender system* bases its performance on a knowledge base, in which the domain knowledge is usually represented through conceptual maps (like taxonomies or thesaurus) or ontologies. *Ontologies* are formal, shared conceptualisations of a given domain in terms of classes, taxonomic and non-taxonomic relationships, attributes, instances and domain axioms. This knowledge can be used to represent both the features of the domain items and the user's interests. This fact allows an analysis of preferences at different abstraction levels and provides reasoning capabilities to the recommender system (Mobasher, 2007; Sieg et al., 2007). 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, or new information about the user can be obtained in an implicit way by observing his/her interaction with the system (e.g. analyzing the evaluations provided by the user and recommending items similar to the ones the user liked). Some approaches have proposed to build profiles using semantic knowledge, rather than mere numerical representations (Blanco-Fernández et al., 2011b). In those cases, the structure of the ontology may be used to spread the information about the user's preferences deduced by the system (this idea is explored in one of the contributions of this work, as detailed in chapter 3).

The accuracy of the predictions of a recommender system is usually evaluated with *precision* and *recall* metrics. The former indicates the percentage of recommended items which are relevant for the user, whereas the later is the proportion of user-relevant items that have actually been recommended. These measures are well suited to determine the level of accuracy in matching user's preferences; however, it may be argued (Mcnee et al., 2006) that accuracy is not the only aspect that produces an enjoyable user experience in the interaction with a recommendation system. *Diversity* has been suggested as one of the factors that can increase user satisfaction (Ziegler et al. 2005). Recommending a set of very similar items may be very accurate but it may be counterproductive and even unsatisfactory for the user. Increasing the diversity of the suggestions may produce *serendipity* to the user, which is the quality of surprisingly discovering new items that are somehow interesting. Moreover, such diversity may be also beneficial for retailers, which may increase the visibility and the sales of less popular items. Ontological domain structures can be exploited to compute similarities between items to produce diversified lists. This line of research has also been taken into account in this dissertation, as will be shown in chapter 4.

1.1. Objectives

In this thesis we have faced *two general goals*: (1) from the scientific point of view, to make relevant contributions in the emerging area of ontology-based recommender systems, and (2) from a more technical perspective, to actually design and develop a software system based on the methods proposed in the thesis that can have a real practical application with an impact in the territory within the scope of URV. In more detail, the final overall objective, which has been fulfilled, was to create a novel recommender system in the Tourism domain to improve the experience of tourists visiting the Tarragona province in the south of Catalonia (Spain), in close collaboration with the Science & Technology Park for Tourism and Leisure (PCTTO)[3]. It is worth noting that Tourism is one of the main research and development strategic areas of URV and of the Campus of International Excellence Southern Catalonia[4].

These two main goals can be divided in the following specific objectives:

1. Study the state of the art on recommender systems, specially focusing on those that have been applied in the Tourism domain. Analyse their main functionalities and the AI methodologies they apply, and identify points of improvement.
2. Study the mechanisms of preference modelling, focusing on the approaches that employ ontology-based user profiles, and especially analyze how they deal with the issues of initialization and dynamic update of the profile. Design a new semantic method to dynamically manage user profiles that allows improving the performance of traditional recommender systems.
3. Study the diversification methods that have been applied in recommender systems. Design a new semantic diversification method that improves the results of the existing ones.
4. In collaboration with the PCTTO, design and implement a software system that can be used by the tourists that plan to visit the province of Tarragona. The system should be able to provide personalised recommendations of touristic activities in the area, combining different kinds of recommendation techniques. It should be generic enough to be easily adaptable to specialised Tourism niches or to other geographical areas. Considering the importance of Enology in the area, an especialisation on eno-touristic activities should also be devised. The system should offer both a Web-based and a mobile-based interface, to facilitate the user interaction and to provide a better experience in the preparation stage and during the trip.

[3] http://www.pct-turisme.cat/
[4] http://www.ceics.eu/index.html

1.2. Contributions

The main specific contributions of this Ph.D. thesis towards the fulfilment of these objectives are the following:

1. **Study of tourism recommender systems.**

A comprehensive and thorough search of the smart e-Tourism recommenders reported in the Artificial Intelligence journals and conferences since 2008 has been made. We have performed a survey of the field, which provides some guidelines for the construction of Tourism recommenders and outlines the most promising areas of work in the field in the next years. The survey was published in the following paper:

Borràs, J., Moreno, A., Valls, A. (2014) *"Intelligent tourism recommender systems: a survey". Expert Systems with Applications 41.16 (2014): 7370-7389.*

It is worth mentioning that, according to the information provided by the editors, this survey has been downloaded more than 3,400 times since its publication on November 2014.

2. **Study and proposal of a new semantic preference management method.**

We analysed the state of the art on the management of semantic preferences, focusing on ontology-based models that maintain such information and how Machine Learning algorithms and decision aid methods take profit of these models. This study was reported in the following book chapter:

Valls, A., Moreno, A., Borràs, J. (2013). *"Preference Representation with Ontologies". Multicriteria Decision Aid and Artificial Intelligence: Links, Theory and Applications, pp. 77-99.* Eds: M.Doumpos, E.Grigoroudis. John Wiley and Sons.

We have proposed in this dissertation a new framework for managing personal preferences using ontologies. This approach applies a spreading algorithm to store and propagate preference values through the ontology structure. This framework also reasons about the uncertainty of these preferences. This approach has been presented, among others, in the following book chapter:

Borràs, J., Valls, A., Moreno, A., Isern, D. (2012). *"Ontology-based management of uncertain preferences in user profiles". Information Processing and Management of Uncertainty in Knowledge-Based Systems, Part II.* Eds: S.Greco, B.Bouchon-Menier, G.Colletti, M. Fedrizzi, B.Matarazzo, R.Yager. Communications in Computer and Information Science 298, pp. 127-136, Springer Berlin Heidelberg.

3. Study and proposal of new diversification methods in Recommender Systems.

We have analysed the main diversification mechanisms that are currently applied in Recommender Systems and we have proposed a new one based on semantic clustering. The main variations of the reviewed methods and the proposed one have been compared in the Tourism recommender developed in this work, concluding that the new semantic clustering diversification mechanism achieves very competitive results with an acceptable computational cost. This work is described in the following paper:

Borràs, J., Moreno, A., Valls, A. (2015). *"Diversification of recommendations through semantic clustering"*. Submitted to *IEEE Transactions on Knowledge and Data Engineering.*

4. Design and development of a Tourism Recommender System

Finally, we have applied the new techniques proposed in this dissertation in the design and development of a semantic Tourism recommender for the area of Tarragona (in close collaboration with the Scientific and Technological Park of Tourism and Leisure in Vila-Seca, Tarragona). This system combines several Artificial Intelligence techniques, including collaborative filtering, content-based recommendations, stereotypes, ontological representation and management of knowledge. An initial version of the system (SigTur/e-Destination) provided general recommendations on Tourism activities, whereas a posterior one (EnoSigTur) was more especialised in Enotourism. Detailed information on the architecture, functionalities and techniques applied in these systems has been reported, among others, in the following papers:

Moreno, A., Valls, A., Isern, D., Marin, L., & Borràs, J. (2013). *Sigtur/e-destination: ontology-based personalized recommendation of tourism and leisure activities. Engineering Applications of Artificial Intelligence, 26(1), 633-651.*

Del Vasto-Terrientes, L., Valls, A., Zielniewicz, P., Borràs, J. (2015) "*A hierarchical multi-criteria sorting approach for recommender systems*". Accepted for publication in the Journal of Intelligent Information Systems (in press).

Borràs, J., de la Flor, J. Pérez, Y., Moreno, A., Valls, A., Isern, D., Orellana, A., Russo, A., Anton-Clavé, S. *SigTur/E-destination: A system for the management of complex tourist regions. In: Information and Communication Technologies in Tourism Conference, ENTER 2011*, R. Law, M. Fuchs, F. Ricci, Eds, Springer Verlag, Innsbruck, Austria, 2011, 39-50.

De la Flor, J., Borràs, J., Isern, D., Valls, A., Moreno, A., Russo, A., Pérez, Y., Anton-Clavé, S. (2012). *Semantic Enrichment for Geospatial*

Information in a Tourism Recommender System. Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications, 134-156. Eds: L.Díaz, C.Granell, J.Huerta. IGI-Global.

Borràs, J., Moreno, A., Valls, A., Ferré, M., Ciurana, E., Salvat, J., Russo, A., Anton-Clavé, S. (2012). *Uso de técnicas de Inteligencia Artificial para hacer recomendaciones enoturísticas personalizadas en la Provincia de Tarragona. IX Congreso Turismo y Tecnologías de la Información y las Comunicaciones, TURITEC-2012.* Málaga, Spain, 2012, 217-230.

1.3. Document structure

The present document is divided into the following chapters:

- Chapter 2 details the main works on Tourism Recommender Systems. Different recommendation algorithms, Artificial Intelligence techniques and functionalities are analysed and described in order to define the most relevant methodologies and current lines of work in this area.

- Chapter 3 presents a study on the semantic management of user preferences, focusing on the use of ontologies as knowledge structures, and it proposes a novel approach that exploits ontology structures to manage user preferences.

- Chapter 4 analyses the diversification techniques applied in recommender systems in order to provide varied results and hence increase user satisfaction. A new diversity method based on semantic clustering is proposed and compared with the existing ones.

- Chapter 5 presents a Tourism Recommender System developed for the province of Tarragona which uses the novel techniques proposed in the previous chapters. It details the architecture of the whole system, the Tourism ontology that has been created, and the integration of several recommendation algorithms and Artificial Intelligence techniques.

- Chapter 6 makes a quantitative and qualitative evaluation of the recommender system and presents some adaptations to other domains and areas.

- Chapter 7 provides the final conclusions and comments some potential lines of future research.

Chapter 2 – Intelligent Tourism Recommenders

This chapter provides a comprehensive review of the Tourism recommender systems published in scientific journals and conferences since 2008, with an especial focus on the ones that employ AI techniques. Commercial products are not considered in this review, as it is usually not possible to know how they have been designed and implemented. Several aspects of these systems have been analysed, such as their interface, their functionalities, the recommendation mechanisms and the AI methods and techniques employed. This chapter provides an up-to-date state of the art of the field of intelligent Tourism recommenders, which may be useful not only to the scientists working in this field but to designers and developers of intelligent recommender systems in other domains.

In the next section we analyze which interfaces are commonly used by Tourism recommender systems to interact with users, discussing especially the differences between mobile and Web-based approaches. After that we survey the main functionalities offered by these systems, ranging from the recommendation of a tourist destination to the automatic construction of a detailed complex schedule of a visit of several days to a certain area. Section 2.3 comments the recommendation methods employed by e-Tourism recommenders, focusing on content-based and collaborative approaches. The next section exposes the use of AI techniques from different fields like multi-agent systems, approximate reasoning, knowledge representation, etc. A comparison with previous surveys on Tourism recommenders is given in section 2.5. The chapter concludes with a global analysis of the surveyed systems, conclusions that we have reached which have guided our work in this dissertation and some suggestions of lines of future work in the field.

2.1. Interface

This section analyses the user interfaces of recent Tourism recommender systems. Most of them offer a Web-based interface and/or an interface specifically designed to be used in mobile devices. Table 1 classifies the most relevant e-Tourism recommenders in these two broad categories, and Figure 5 shows the percentage of surveyed systems in each of them. A Web-based interface is the option chosen by most of the systems, since it permits an easy access from any computer connected to the Web without any kind of downloading, installation and configuration. However, due to the enormous increase in the use of smart phones connected to the Web in the last years, more than half of the reviewed systems have specific interfaces for mobile devices.

Table 1. Review of user interfaces

Interface	References
Web+mobile	(Venkataiaha et al., 2008), (Lamsfus et al., 2009), (Niaraki and Kim, 2009), (Vansteenwegen et al., 2010), (Gavalas and Kenteris, 2011), (Rey-López et al., 2011), (Ruotsalo et al., 2013), (Umanets et al. 2013)
Only web	(Coelho et al., 2009), (Huang and Bian, 2009), (Lucas et al., 2009), (Lee et al., 2009), (Ruiz-Montiel and Aldana-Montes, 2009), (Jannach et al., 2010), (Mínguez et al., 2010), (Sebastià et al., 2010), (Yang, 2010b), (García-Crespo et al., 2011), (Linaza et al., 2011), (Lorenzi et al., 2011), (Luberg et al., 2011), (Montejo-Ráez et al., 2011), (Sebastià et al., 2009) (Garcia et al., 2011), (Wang et al., 2011), (Koceski and Petrevska, 2012), (Gyorodi et al. 2013), (Kurata and Hara 2013), (Lucas et al., 2013), (Savir et al., 2013), (Cha, 2014)
Only mobile	(Castillo et al., 2008), (Ceccaroni et al., 2009), (García-Crespo et al., 2009), (Yu and Chang, 2009), (Ricci et al., 2010), (Martin et al., 2011), (Batet et al., 2012), (Martínez-Santiago et al., 2012), (Noguera et al., 2012), (Garcia et al. 2013a), (Meehan et al., 2013), (Rojas and Uribe 2013), (Yang and Hwang, 2013), (Wei et al., 2014) , (Braunhofer et al. 2015)

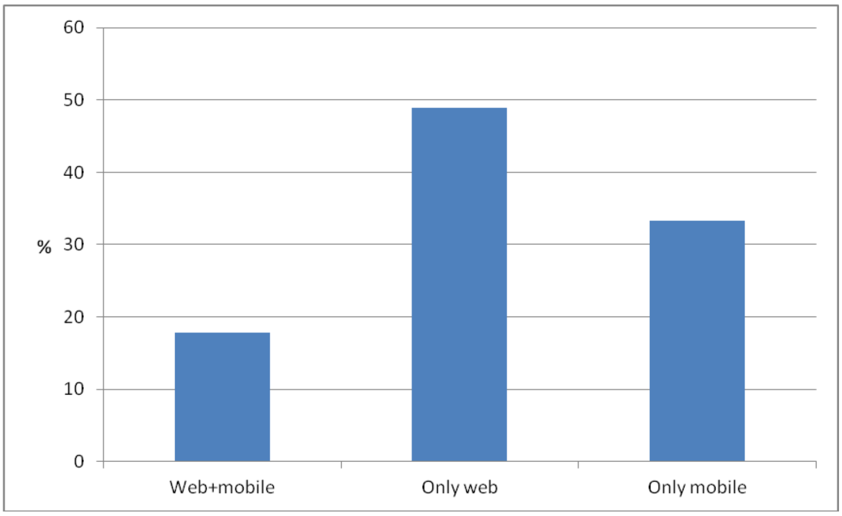


Figure 5. Interfaces used in the reviewed works (in %)

There are some recommender systems that have been designed as desktop applications and do not offer any of the two usual kinds of interfaces (e.g., (Kurata, 2011)). This kind of applications can usually be implemented more quickly than the mobile or Web-based ones; however, they require downloading and installing the program, which is not comfortable to most of the tourists that want to get recommendations as simply as possible without being bothered by technical details.

The following subsections review some approaches based on Web or mobile interfaces.

2.1.1. Web-based recommenders

The use of a Web-based interface is the most common option adopted by e-Tourism recommenders. This kind of interfaces allows tourists to look for information in a user-friendly manner. Users normally have a rich interaction with the system using a wide screen which allows displaying a large amount of data extended with maps, images or even high quality videos. Moreover, the mouse permits to interact easily with the computer and move through maps, perform zoom actions, select items or even drag and drop them. This is very useful for tourists when they are still in the planning stage of their trips. Nevertheless, Web-based applications are usually not designed to be used during the stay since most of the tourists will not have easy access to computers with Internet connection. Although an increasing number of tourists have mobile handsets or tablets with Internet connection, the information-ridden Web pages usually shown by recommenders cannot be easily read or manipulated on such small screens. In the remainder of this section we comment some interesting features exploited in Web-based interfaces to improve the interaction with the users.

(Venkataiaha et al., 2008) report the design of two visualisation systems (called *discrete* and *continuous*) for a tourism recommender and compare the interaction of the users in both cases. The former provides a high quantity of information in the screen at the same time, and it was determined that users needed too much time and effort to understand it. The latter aggregates all the information into a single video clip that combines the most relevant media content, including text, photographs and videos.

The approach shown in (Lee et al., 2009) is one of the firsts that embeds Google Maps Services[5] in their Web pages (Figure 6) in order to plot the travel route on a map, so that tourists can follow the personalized itinerary to enjoy cultural heritage and local gourmet food during their stay at Tainan City.

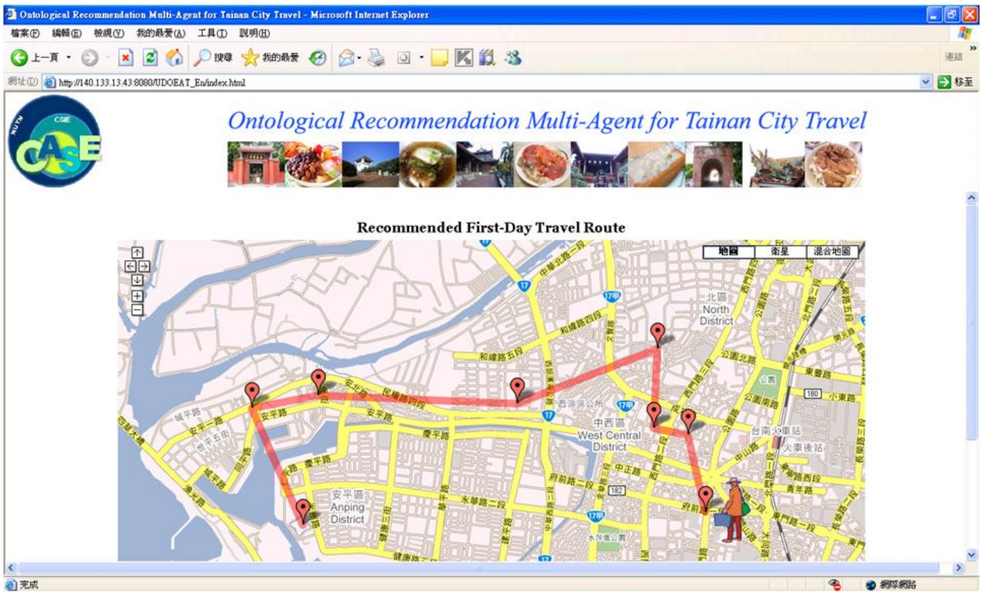


Figure 6. Personalized route through Tainan City (from (Lee et al., 2009))

Other Web-based recommender systems that display in a map the places scheduled to be visited in a single day are *e-Tourism* (Sebastiá et al., 2009),

[5] https://developers.google.com/maps/ (last access March 2014)

Otium (Montejo-Ráez et al., 2011) and *City Trip Planner* (Vansteenwegen et al., 2010). In this last system the user introduces his/her interests and he/she will receive a scheduled route of attractions for one day represented by a timetable (left image of Figure 7) and a map (right image of Figure 7).

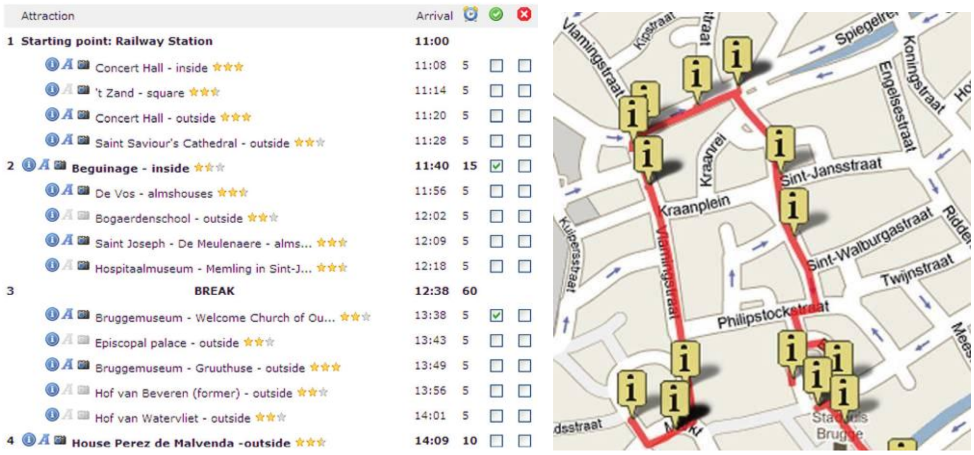


Figure 7. Web version of City Trip Planner: scheduled route plan and its representation on the map (from (Vansteenwegen et al., 2010))

The *VIBE virtual spa advisor* (Jannach et al., 2010) keeps an avatar-based conversation with the tourist in order to acquire the user's visit requirements through personalised forms. The main point of this approach is its dynamicity. If a new attribute has to be added to the product catalogue, it is automatically taken into account not only in the recommendation and preference elicitation processes, but also in the Web interface which is changed accordingly. The Web site has a section for domain experts, in which they can add or modify logical conditions that govern the conversational and recommendation procedures.

(Wang et al., 2011) show how Semantic Web technologies may be integrated with Web 2.0 services to leverage each other's strengths. To do so they propose an ontology-based tourism recommender that allows the automatic and dynamic integration of heterogeneous on-line travel information. The platform is built in Ruby on Rails with view extensions to create rich Ajax Web-based applications. They also use third party services to provide additional features, such as Google Map, Yahoo Weather, and WikiTravel.

2.1.2. Mobile recommendations

Systems that offer mobile interfaces have increased considerably in the last few years, due to the large number of users acquiring mobile devices with Internet connection or, more recently, the well-known smartphones. Mobile devices are small and their Internet connection is usually slow; thus, the quantity of information that can be shown in these devices cannot be compared with a standard Web page. Therefore, mobile Tourism recommender systems have to make an effort to provide only the information that is essential for the user, and it must be well structured in order to be displayed correctly in small screens. Moreover, the user's interaction with the system is limited, since even the basic actions made in Web-based interfaces (scrolling, introducing text) are not that easy. However, it is fair to say that the latest smartphones with bigger touchscreens provide a better user interaction. Furthermore, the main advantage of mobile devices is that they allow the use of the system in any place with an Internet connection, so that tourists may access information, discover places or modify their trips during the stay. Besides, most mobile systems are equipped with GPS and the recommender may know the present location of the user and it may offer geo-referenced information, advice or recommendations based on this knowledge.

One of the first approaches in the field that used mobile systems was reported in (Yu and Chang, 2009). This system, designed for PDAs, offers location-based recommendation services to support personalized tour planning. Recommendations are based on tourists' preferences, location and time. Figure 8 shows the mobile user interface in four separated screenshots. The first one shows the different mobile tourism services (restaurant, hotel, sightseeing spot, user profile, and tour plan recommendation). The second image illustrates the interface for setting user preferences. The third screenshot shows the recommended tour plan with information about the places to visit, such as names, descriptions, photos or visiting time frames. Finally, the last image displays the tour plan on Google Maps.

Figure 8. Prototype system for Windows mobile devices (from (Yu and Chang, 2009))

Another approach compatible with PDAs is *MTRS* (Gavalas and Kenteris, 2011). The authors argue that tourists may have problems to connect with the Internet, either because they are in a rural area or because they are foreigners and cannot afford the roaming costs abroad. They propose to solve this problem by installing an infrastructure to support proximity detection and a cost-effective means for remote content update. In fact, they propose to use small to medium-scale wireless sensor networks. Through this infrastructure, they introduce the concept of 'context-aware rating', in which user ratings uploaded through fixed Internet connection infrastructures (located at the rated places) are weighted higher to differentiate them from users that provide an evaluation using the Internet away from the visited place.

Another product using mobile devices is *MapMobyRek* (Ricci et al., 2010) that exploits quite well its interface by showing recommendations in lists and on maps. This system permits to compare two items with their characteristics displayed side-by-side in order to decide the one that is preferred.

GeOasis (Martínez-Santiago et al., 2012) acts as a tourist guide that describes the places to visit while the tourist approaches the recommended locations. The system uses the mobile GPS device to know the tourist location and speed in order to estimate the available time to give the explanations. Users can interact with the system in two ways: using a tactile interface or using a voice-based interface (voice recognition and text-to-speech software).

Despite the existence of several mobile tourism recommenders, not many of them use the newest technologies in mobile devices, such as the Android or iPhone platforms. Some examples that use these popular and rising platforms are reviewed below.

The *moreTourism* (Rey-López et al., 2011) Android-based platform provides information about tourist resources through the use of mashups, integrating images, videos, augmented reality services, geo-location, guide services, access to urban networks, etc. *LiveCities* (Martin et al., 2011) uses the notification service of Android systems to provide push information according to the user context. This information can be plain text, audio, video or HTML. The *STS* system (Braunhofer et al. 2015) is a powerful Android application with a good design interface that permits users to enter accurate information about their interests and opinions on the trip and the visited attractions (see Figure 9).

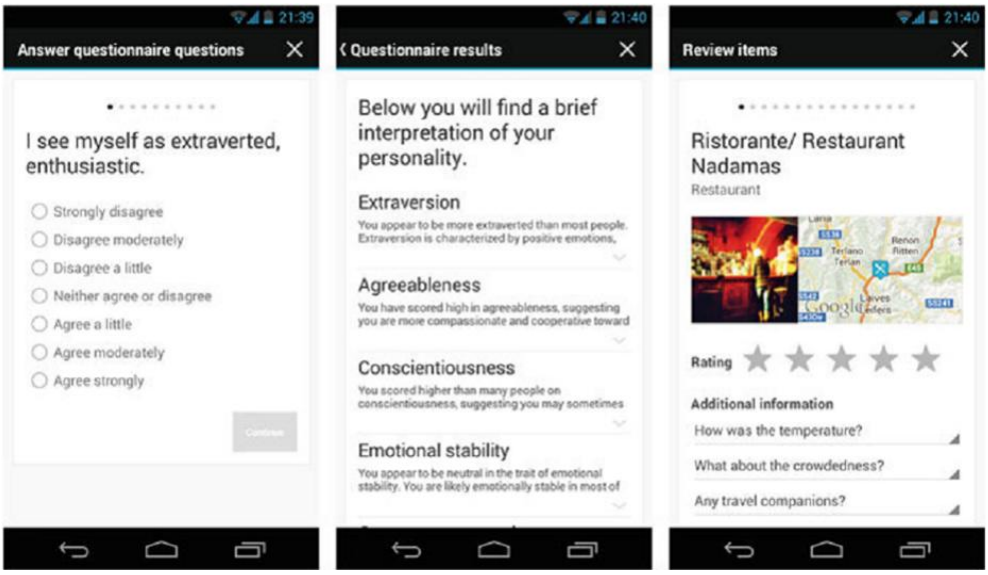


Figure 9. Preference elicitation from the mobile interface of the STS system (Braunhofer et al. 2013)

The recent *GUIDEME* system (Umanets et al. 2013) features a good implementation for mobile devices since its designers have not only developed an app for phones but also for tablet devices. In particular, the app is built for the iOS platform and it is adaptive to the screen sizes with specific adjustments for both iPhone and iPad devices. Figure 10 shows screenshots of its iPad version. *REJA* (Noguera et al., 2012) also works for iOS platforms.

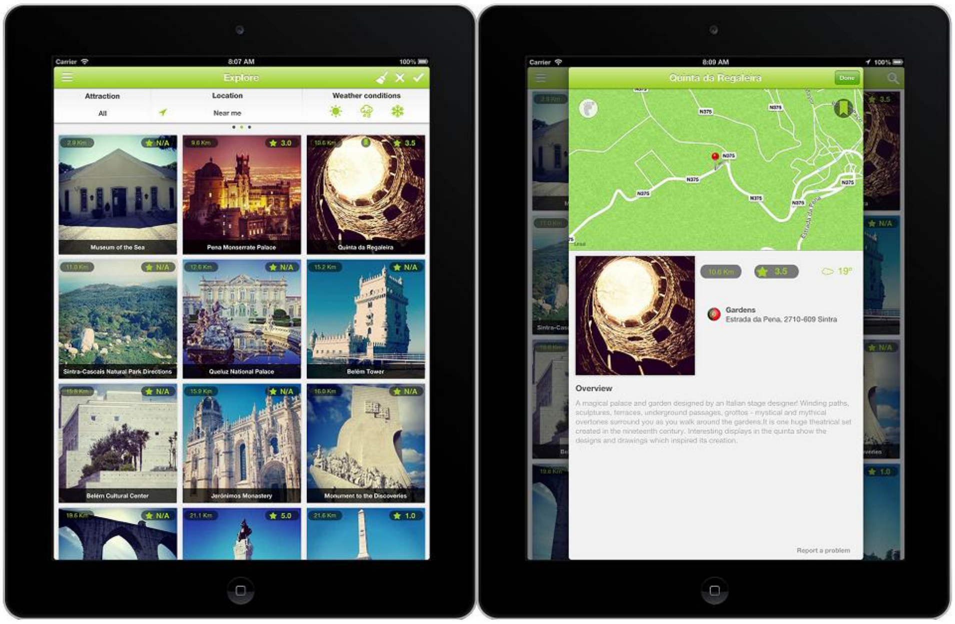


Figure 10. iPad version of the GUIDEME system (Umanets et al. 2013)

2.2. Functionalities

In this section we describe the general functionalities provided by the reviewed Tourism recommender systems. Table 2 catalogues the approaches in four broad groups, depending on the services they offer: suggestion of a destination and construction of a whole tourist pack, recommendation of suitable attractions in one specific destination, design of a detailed multi-day trip schedule, and social capabilities. Figure 11 gives a visual estimation of the percentage of systems that offer each of them. These aspects are commented in more detail in the following subsections, with examples of the most prominent proposals.

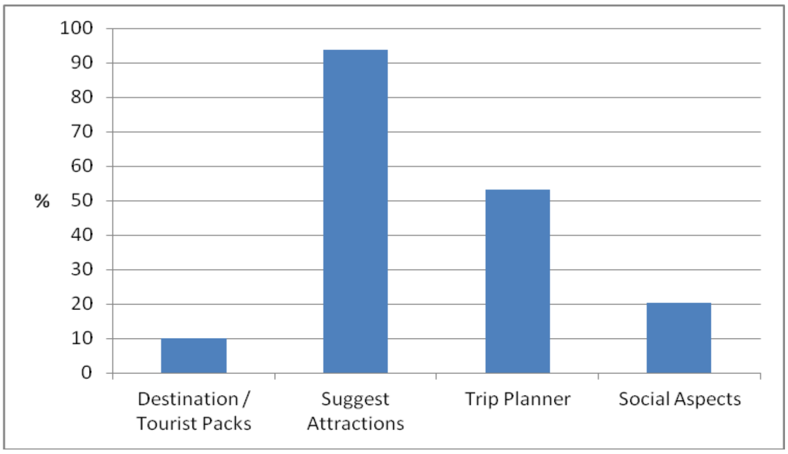


Figure 11. Functionalities offered by the reviewed approaches (in %)

Table 2. Review of user functionalities

Functionalities	References
Destination / Tourist Packs	(Seidel et al., 2009), (Yu and Chang, 2009), (Lorenzi et al., 2011), (Koceski and Petrevska, 2012), (Herzog and Wörndl, 2014)
Suggest Attractions	(Castillo et al., 2008), (Coelho et al., 2009), (Ceccaroni et al., 2009), (García-Crespo et al., 2009), (Huang and Bian, 2009), (Lamsfus et al., 2009), (Lucas et al., 2009), (Lee et al., 2009), (Niaraki and Kim, 2009), (Ruiz-Montiel and Aldana-Montes, 2009), (Yu and Chang, 2009), (Jannach et al., 2010), (Mínguez et al., 2010), (Ricci et al., 2010), (Sebastià et al., 2010), (Vansteenwegen et al., 2010), (Yang, 2010b), (Fenza et al., 2011), (Gavalas and Kenteris, 2011), (Kurata, 2011), (Linaza et al., 2011), (Lorenzi et al., 2011), (Luberg et al., 2011), (Martin et al., 2011), (Montejo-Ráez et al., 2011), (Rey-López et al., 2011), (Sebastià et al., 2009), (Garcia et al., 2011), (Wang et al., 2011), (Batet et al., 2012), (Koceski and Petrevska, 2012), (Martínez-Santiago et al., 2012), (Garcia et al. 2013a), (Gyorodi et al. 2013), (Kurata and Hara 2013), (Lucas et al., 2013), (Meehan et al., 2013), (Rojas and Uribe 2013), (Ruotsalo et al., 2013), (Savir et al., 2013), (Umanets et al. 2013), (Yang and Hwang, 2013), (Cha, 2014), (Han and Lee, 2014), (Wei et al., 2014) , (Braunhofer et al. 2015)
Trip Planner	(Castillo et al., 2008), (Coelho et al., 2009), (Ceccaroni et al., 2009), (García-Crespo et al., 2009), (Huang and Bian, 2009), (Lucas et al., 2009), (Lee et al., 2009), (Niaraki and Kim, 2009), (Yu and Chang, 2009),(Mínguez et al., 2010), (Sebastià et al., 2010), (Vansteenwegen et al., 2010), (Kurata, 2011), (Linaza et al., 2011), (Luberg et al., 2011), (Montejo-Ráez et al., 2011), (Rey-López et al., 2011), (Sebastià et al., 2009) (Garcia et al., 2011), (Wang et al., 2011), (Batet et al., 2012), (Kurata and Hara 2013), (Lucas et al., 2013), (Savir et al., 2013), (Cha, 2014), (Han and Lee, 2014) , (Herzog and Wörndl, 2014)
Social Aspects	(Coelho et al., 2009), (Ceccaroni et al., 2009), (García-Crespo et al., 2009), (Vansteenwegen et al., 2010), (Rey-López et al., 2011), (Sebastià et al., 2009) (Garcia et al., 2011), (Garcia et al. 2013a), (Meehan et al., 2013), (Umanets et al. 2013), (Yang and Hwang, 2013) , (Han and Lee, 2014)

2.2.1. *Travel destination and tourist packs*

Some of the reviewed systems focus on the recommendation of a destination that suits the user's preferences. This is the case of systems like *PersonalTour* (Lorenzi et al., 2011), *Itchy Feet* (Seidel et al., 2009) and *MyTravelPal* (Koceski and Petrevska, 2012). *PersonalTour* is used for travel agencies to help their costumers to find the best travel packages according to their preferences. Once the recommendation process is finished, a rated list of options is presented to the costumer. Table 3 shows an example of the hotel recommendation service. After that, the customer can rate each item of each travel service.

Table 3. Example of hotel recommendation in PersonalTour (adapted from (Lorenzi et al., 2011))

Id	Hotel name	City	Hotel category	Room category	Room type	Swimming Pool	WiFi
1	Libertel	Paris	Economic	Standard	Double	No	Yes
2	Palladium	Punta Cana	Resort	Luxe	Double	Yes	Yes
3	Amadeus	Milan	Economic	Standard	Single	No	Yes
4	Riu Palace	Cancun	First	Luxe	Double	Yes	Yes
5	WestIn	Aruba	Economic	Luxe	Double	Yes	Yes

Itchy Feet not only recommends tourism destinations but it also provides purchasing services for booking a trip and assistance from professional travel agents. Users make search requests, which are handled by autonomous agents that search for information in the internal database as well as in external data sources. The results are shown to the user through the interface, where recommended items (flights and hotels) can be selected and purchased.

MyTravelPal (Koceski and Petrevska, 2012) first recommends areas of interest over a region graphically (see Figure 12), where the size of the circle indicates the level of affinity with the user. Once the user focuses on a particular area, their tourist resources are also shown and sized depending on the affinity to the user profile.

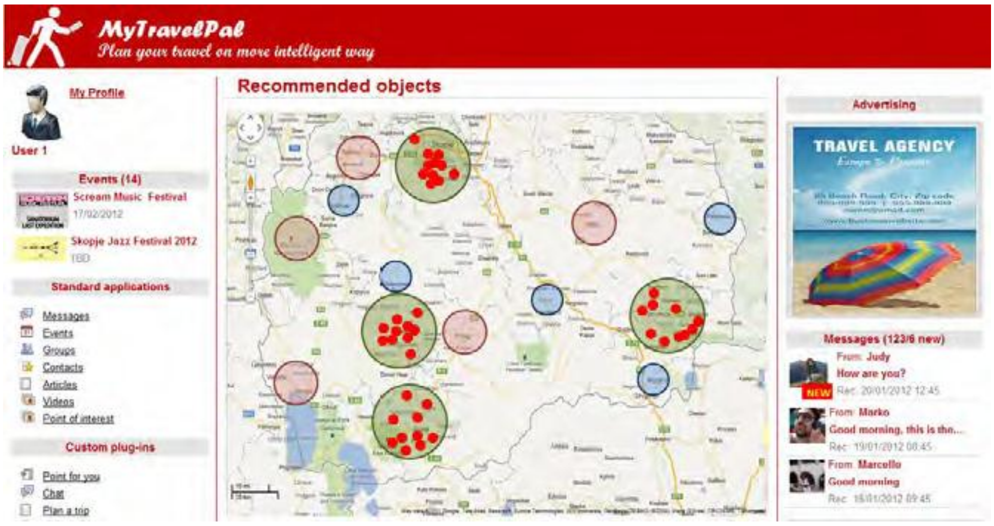


Figure 12. MyTravelPal – recommendation of regions of interest

2.2.2. Ranked list of suggested attractions

Most Tourist recommender systems tend to suggest places once the user has decided the destination of the trip or he/she is already there. These systems are more complex, since they can suggest a large number of attractions, accommodations, restaurants or even temporal events. In this context the capability of recommenders to classify and rank only those elements considered important for a particular user among the huge quantity of available information is very useful. With the support of these systems the users can find interesting places in an efficient way and even discover unexpected ones that may be of their interest. The activities to be recommended are normally stored in a static database, although some systems (e.g. *Otium* (Montejo-Ráez et al., 2011)) extract automatically information about events from the Web to ensure that they always provide updated information.

This kind of recommender systems (e.g. (Sebastiá et al., 2009; Ruiz-Montiel and Aldana-Montes, 2009; Fenza et al., 2011)) usually provide a list of activities that match the user profile, have been visited and/or positively evaluated by similar users in the past, or are similar to activities previously enjoyed by the user. Thus, they include mechanisms to compare the user preferences with the features of an object, or to compare the similarities between two users or two objects. The selection of the recommended items may also take into account contextual factors, like the present location of the user (Noguera et al., 2012). Some systems are also capable of justifying the provided recommendations (e.g. (Jannach et al., 2010)).

SMARTMUSEUM (Ruotsalo et al., 2013) is an example of a more complex recommendation system, which detects automatically if the user is outdoors or indoors, based on his/her location. For the first case, it can display the recommendations on a map. For indoor scenarios, it gives a list of the most relevant objects according to the user's preferences. This is useful, for instance, in museums, where the number of objects to see may be relatively high (Figure 13).

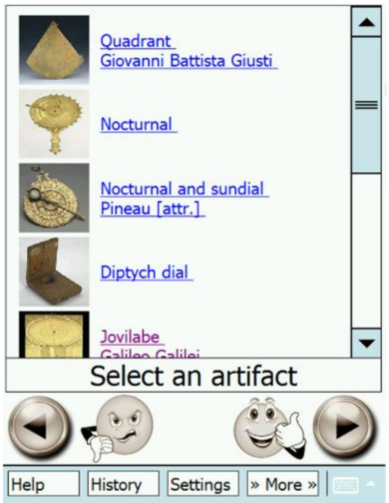


Figure 13. of recommendations of specific objects for indoor scenarios

2.2.3. *Planning a route*

There are several projects that not only provide a list of the places that fit better with the user's preferences but also help tourists to create a route through several attractions.

CT-Planner (Kurata, 2011; Kurata and Hara, 2013) offers tour plans, as shown in Figure 14, that are refined gradually as the user's expresses his/her preferences and requests (duration, walking speed, reluctance to walk, etc.).

It displays a radar chart that represents the user's preferences and a cartoon character as a navigator, in order to enrich the sense of user-friendliness and interactivity.

Figure 14. CT-Planner2 user interface (from (Kurata, 2011))

There are several systems that provide an initial set of recommended activities (or an initial plan), with which the user can directly interact to add more activities, remove activities, select an activity to be visited, change the order of visit, etc. The planning component of the recommender system takes into account important factors like the expected duration of the visit, the opening and closing times of the attractions and the distance between them. Some relevant examples include *City Trip Planner* (Vansteenwegen et al., 2010), *CRUZAR* (Mínguez et al., 2010), *Smart City* (Luberg et al., 2011), *Otium* (Montejo-Ráez et al., 2011) and *e-Tourism* (Sebastiá et al., 2009). A more detailed review of trip planning functionalities is available in (Vansteenwegen and Souffriau, 2011). Some advanced recommenders, like *SAMAP* (Castillo et al., 2008) and *PaTac* (Ceccaroni et al., 2009), are capable of analysing the connection possibilities between the activities using different means of transport (walking, by bike, by car, or by public transport). In the work reported in (Herzog and Wörndl, 2014) the organization of the trip may include multiple destinations and activities, and the system needs to find routes at different levels of a spatial hierarchy of regions. At the moment the system only recommends regions and each region has its own static routes. However, they plan to do recommendations of activities within each region in the near future.

Some of these systems incorporate more complex *Geographical Information Systems* to manage the geographical data associated to the touristic points and events. (Huang and Bian, 2009) argued that it is computationally unfeasible to maintain large amounts of spatial data and use them in planning procedures. Hence, they used existing geospatial Web service technologies, in concrete the ESRI ArcWeb Service[6], to obtain the location of the attractions, the distance between them given their street address, and driving directions between two attractions. *GeOasis* (Martínez-Santiago et al., 2012) continually calculates the position and the speed of the user. The estimated time to reach a place is considered in order to create the plan in real time. The key aspect is the prediction of where the user will be in the immediate future: in a city, near a city or on the road. If the user is already in a city, the planning algorithm checks the nearest places to the user without taking into account the route or the speed, since it is considered that the user is close to them. If the user is near a city, the planning algorithm checks the most relevant attractions in it. If the user is on the road, but not near a city, then the planning algorithm is more complex because it considers temporal constraints. The plan is not computed by the server but by the client application, since it is constantly checking the location by GPS. Routes are computed using Google Maps as an external resource.

Once the visit plan has been completely defined, the user may wish to retrieve the full schedule to follow the route. This retrieval can take different forms. Systems like *SAMAP* (Castillo et al., 2008) allow downloading a PDF file that contains a geo-referenced map with a detailed explanation of the plan. In others, like *City Trip Planner* (Vansteenwegen et al., 2010) and Otium (Montejo-Ráez et al., 2011), the user can download the route to a mobile phone.

2.2.4. Social aspects

Several projects (e.g. (Vansteenwegen et al., 2010; Ceccaroni et al., 2009; Unamets et al., 2013; Garcia et al., 2013a)) have paid special attention to the inclusion of *social* functionalities that allow users to share material (pictures, comments, evaluations) and interact with other tourists. These aspects may be very interesting to help to promote the use of a recommender among the visitors of a particular city. Recommenders like *moreTourism* (Rey-López et al., 2011) and *Itchy Feet* (Seidel et al., 2009) allow users not only to interact over popular social networks but also to create location-based activity groups that can be employed to post comments, join groups for doing common activities or interact with other users. The system *e-Tourism* (Garcia et al., 2011) allows creating plans that accommodate the preferences of a whole group of visitors. (Han and Lee, 2014) developed an approach that adaptively recommends clusters of landmarks using geo-tagged social media. The importance of landmarks is based on the trip's spatial and temporal properties. Figure 15 illustrates an example of recommendations of relevant landmarks.

In *iTravel* (Yang and Hwang, 2013) users communicate among them with mobile peer-to-peer communications to send ratings of attractions. Their navigation map not only displays the location of attractions but also the position of near-by users with which it is possible to communicate.

[6] http://www.esri.com/news/arcuser/0403/arcweb.html (last access March 2014)

Figure 16 shows a map with recommended attractions (green pins) and nearby users (blue pins).

The *VISIT* system (Meehan et al., 2013) applies sentiment analysis techniques (using the Alchemy API[7]) to analyse the updates about a given attraction in Twitter and Facebook and identify if users are expressing positive or negative comments about it. This information is shown with green and red colours in its interface, so that the user may easily identify the nearby places that visitors are liking (or disliking) in real time.

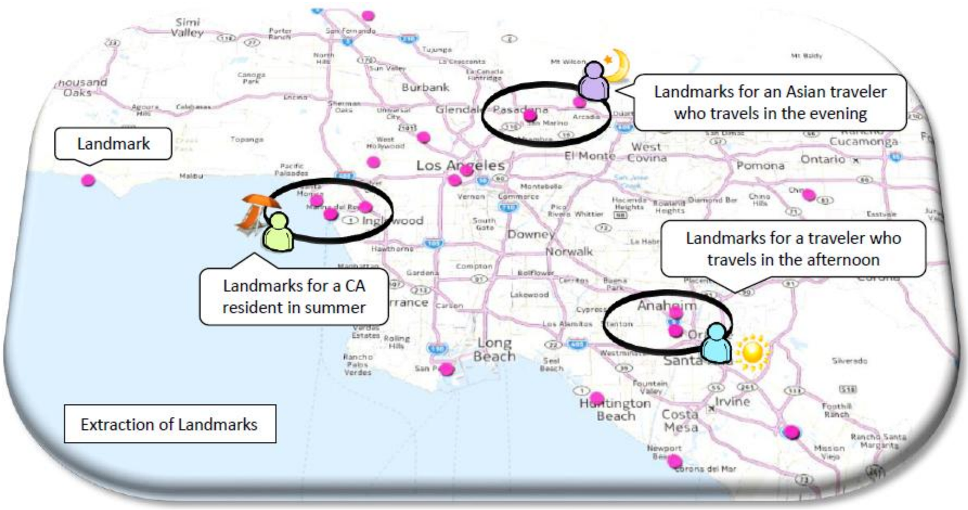


Figure 15. Landmark recommendations (from (Han and Lee, 2014))

Figure 16. Navigation map of iTravel (from (Yang and Hwang, 2013))

[7]AlchemyAPI. (2014) Alchemy API: Transforming text into knowledge. [Online]. http://www.alchemyapi.com/

2.3. Recommendation techniques in e-Tourism

Recommender systems have been usually classified, according to the way in which they analyze the information of the user and filter the list of items, into content-based, collaborative and demographic systems (Montaner et al., 2003; Burke, 2002; Manouselis and Costopoulou, 2007). In this section we introduce these three paradigms, analyzing its use in current Tourism recommender systems.

Content-based (CB) recommenders generate suggestions based on the preferences of the user, by calculating a degree of similarity between the user and the items to be recommended. The process is carried out by comparing the features of the item with the user's preferences. So, it is assumed that both users and alternatives share a common representation (e.g., they are composed of the same set of attributes or keywords). The output of the comparison process is usually an overall performance score, which indicates the degree of correspondence between the user's profile and each alternative. The higher the score is, the higher the performance of the alternative for a given user. Sometimes these methods also take into account the rating history of the user. In this approach, the recommendation system relies on having an accurate knowledge of the user's preferences to be able to select the appropriate items. The main disadvantage of content filters is known as over-specialization, which occurs when a recommendation system always tends to recommend the same items. For instance, if a user expresses a high interest in visiting museums, he/she may receive only recommendations associated to the visit of museums. Moreover, they can also suffer from the "*cold start*" problem when a new user enters in the system, because we can have poor knowledge about the user in an initial stage. The advantage of these methods is that they make recommendations according to the user's preferences and not subject to the opinions of other users.

In CB systems the recommendation process is mainly focused on defining an appropriate measure to compare a user and an item. The two most common approaches are the *aggregation* of ratings and the definition of a *distance* function.

- When the user profile is represented as a rating vector with the degree of interest of the user in each attribute, each rating can be interpreted as a performance score that can be used to evaluate an alternative. The goal is then to calculate an overall interest score for a certain alternative. The simplest approach consists of using an aggregation operator to combine the user ratings on the concepts that define a certain alternative (Batet et al., 2012). More sophisticated aggregation methods have also been applied, like AHP (*Analytic Hierarchy Process*) (Niaraki and Kim, 2009; Huang and Bian, 2009).

- When the items and the users are described by a list of keywords, some similarity measures can be applied. For example in (Lamsfus et al., 2009) items and users are described using concepts from an ontology, which defines archetypes of tourists (e.g. cultural, sportive or adventurous), and the cosine similarity between the two vectors (user and item) is calculated. A similar approach is proposed in (Gyorodi et al. 2013) with ad-hoc hierarchies of tags for locations that are rated by users. The locations ratings are then compared to the user's tags. In (García-Crespo et al., 2009) a feature-based similarity algorithm is applied, using several ontologies as reference. In (Fenza et al., 2011) classification rules are automatically generated and later used to define the degree of correspondence between the user and the item.

Collaborative (CL) systems make recommendations based on groups of users with similar preferences. The similarity between users is normally computed by comparing the ratings that they give to some of the items. When the system identifies who are the people that share similar interests with the current user, then the items that those people liked are recommended. In this approach, some feedback about the provided recommendations is necessary, in order to know which items the user has liked or disliked (e.g. which places he/she has enjoyed visiting). Two types of CL methods are distinguished: *user-based* and *item-based*. The former finds neighbours of a target user by matching his/her opinions with the ones of the other users in the system. The latter builds groups by finding similarities on the items that the users liked (or disliked) in the past.

Two weak points are recognized in CL systems: "*data sparsity*" and "*grey sheep*". The former occurs when the number of ratings from users is small in comparison with the total number of items, so that the probability of finding users that rate the same items is too low to make good estimations. The latter, "grey sheep", refers to a user with a profile different from the rest of users of the system. In this case, it is difficult to find appropriate items to recommend because we do not have information about similar users. Finally, this approach also suffers from the *scalability* problem if the community of users is large. On the other hand, the main advantage of collaborative filtering is to generate recommendations that may be more varied and surprising than the content-based ones.

Demographic-based (DM) systems rely on the personal data of the user (e.g. age, country of origin, level of studies, etc.). In this case, the recommendation is not based on the user's interests and preferences but on his/her personal characteristics. In this approach, users are usually assigned to a certain stereotypical class depending on their demographic data, so that the members of the same group share a common demographic profile. The system has internal knowledge about the standard preferences of each stereotype, which is used to provide the recommendations to the users. The definition of stereotypes of tourists is not new in this field. Many studies have defined segments of tourists according to their behaviour in different cities or territories (Brewer, 1984; Marques, 2009; Tsung-Chiung et al., 2012). A recent approach (Braunhofer et al., 2015) incorporates a questionnaire of 10 personality traits that allows classifying the user in a five-dimensional space (i.e., conscientiousness, agreeableness, extroversion, emotional stability and openness). Specific stereotypes provide precise descriptions of what tourists want and how they act in different situations. This information is normally used as a guide to conduct business with tourists, but it can also be exploited in recommender systems. These methods tend to provide generic recommendations, less accurate than those provided by other methods. However, they may be a good starting point to initialize the user's preferences when they are still unknown.

Since each of the approaches has some drawbacks, the combination of different techniques is also a widespread practice. Figure 17 shows the distribution of the different kinds of recommendation techniques used in the field of Tourism recommenders, in percentages. More than half of the works use a mixture of techniques (52%), combining mainly CB methods with CL filtering or with DM techniques. The rest of the systems apply a single approach, having a clear predominance for the techniques based exclusively on the description of the content of the alternatives (42% of the reviewed papers).

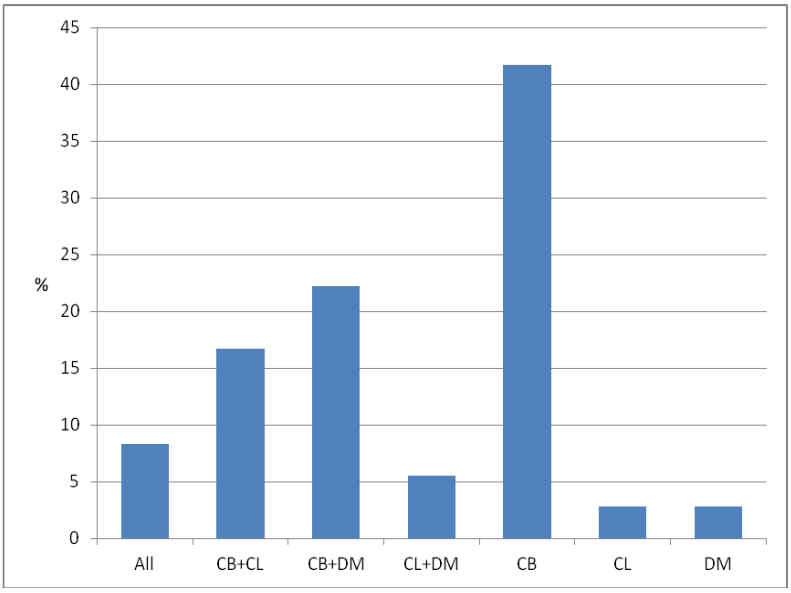


Figure 17. Use of recommendation techniques in Tourism recommenders

Hybrid systems can integrate these techniques in different ways. Three approaches can be distinguished:

1. Selection of the method: the system incorporates DM, CB and CL methods, but only one of them is applied depending on the particular situation of each user. For example, the first time the user arrives, a method based on demographic data is used. Later on, if similar users can be found, a CL recommendation is made; otherwise, a CB

procedure is applied. This is the case of (Martínez et al., 2009), (Huang and Bian, 2009) or (Noguera et al., 2012).

2. Sequential use: each recommendation technique is used in different stages of the process. For example, *SPETA* (García-Crespo et al., 2009) has four steps: first, contextual information (location, time) is used to make the first selection of appropriate options; second, a more fine grained set of results is obtained using knowledge-based filtering techniques, by calculating the semantic similarity between the user preferences and the touristic services; third, preferences and CL techniques are used to refine the set of options; finally, a vector of preferences is used to make the final selection. In (Braunhofer et al., 2014) CL filtering with DM and personal information is applied in a training phase to build a prediction model in different contexts. After that, CB techniques generate the list of recommendations by computing ratings for each item based on the current and predicted values.

3. Integrated use: both CB and CL techniques are combined during the execution. For example, in (Lucas et al., 2013) users are classified into groups using simultaneously personal data (DM), information about the content of the items previously selected by the user (CB) and the information of other users (CL). Then fuzzy rules are automatically generated so that new users can be classified into several groups, with different membership degrees. The list of recommended items is finally derived from a prediction based on the groups the user belongs to.

In the survey we have observed an increasing trend in the exploitation of CL filtering techniques since 2012, mainly in hybrid systems. More precisely, from 2008 to 2011 only 25% of systems used such method, whereas since 2012 the percentage has increased to 70% (see Table 4).

Table 4. Review of recommendation methods used

Recommendation method	**Reference**
ALL	(Lucas et al., 2009), (Ruiz-Montiel and Aldana-Montes, 2009), (Batet et al., 2012), (Koceski and Petrevska, 2012), (Garcia, Torre, and Linaza 2013), (Lucas et al., 2013), (Meehan et al., 2013) , (Braunhofer et al. 2015)
CB+CL	(Castillo et al., 2008), (García-Crespo et al., 2009), (Fenza et al., 2011), (Rey-López et al., 2011), (Noguera et al., 2012), (Rojas and Uribe 2013), (Cha, 2014)
CB+DM	(Coelho et al., 2009), (Ceccaroni et al., 2009), (Lamsfus et al., 2009), (Niaraki and Kim, 2009), (Mínguez et al., 2010), (Yang, 2010b), (Martin et al., 2011), (Sebastià et al., 2009) (García et al., 2011)
CL+DM	(Gavalas and Kenteris, 2011), (Wei et al., 2014)
CB	(Huang and Bian, 2009), (Lee et al., 2009), (Seidel et al., 2009), (Yu and Chang, 2009), (Jannach et al., 2010), (Ricci et al., 2010), (Sebastià et al., 2010), (Vansteenwegen et al., 2010), (García-Crespo et al., 2011), (Kurata, 2011), (Linaza et al., 2011), (Lorenzi et al., 2011), (Luberg et al., 2011), (Montejo-Ráez et al., 2011), (Martínez-Santiago et al., 2012), (Gyorodi, Gyorodi, and Dersidan 2013), (Kurata and Hara 2013), (Ruotsalo et al., 2013)
CL	(Savir et al., 2013), (Umanets et al. 2013), (Yang and Hwang, 2013)
DM	(Wang et al., 2011)

2.4. Use of AI techniques in Tourism recommender systems

This section makes a brief review of the main AI techniques and tools employed in Tourism recommender systems in the last years, which are summarised in the following table.

Table 5. AI techniques used in e-Tourism recommenders

AI techniques	References
Multi-agent systems	(Castillo et al., 2008), (Ceccaroni et al., 2009), (Lee et al., 2009), (Seidel et al., 2009), (Sebastià et al., 2010), (Lorenzi et al., 2011)
Optimization techniques	(Castillo et al., 2008), (Lee et al., 2009), (Garcia et al., 2010), (Vansteenwegen et al., 2010), (Garcia et al., 2013b), (Meehan et al., 2013)
Automatic clustering	(Castillo et al., 2008), (García-Crespo et al., 2009), (Martínez et al., 2009), (Fenza et al., 2011), (Gavalas and Kenteris, 2011), (Batet et al., 2012), (Noguera et al., 2012), (Lucas et al., 2013), (Kurata and Hara 2013), (Ruotsalo et al., 2013), (Han and Lee, 2014)
Management of uncertainty	(García-Crespo et al., 2009), (Huang and Bian, 2009), (Lamsfus et al., 2009), (Lamsfus et al., 2011), (Pinho et al., 2011), (Wang et al., 2011), (Hsu et al., 2012), (Lucas et al., 2013), (Meehan et al., 2013), (Ruotsalo et al., 2013), (Wei et al., 2014)
Knowledge representation	(Castillo et al., 2008) , (Ceccaroni et al., 2009), (Lamsfus et al., 2009), (Lee et al., 2009), (Sebastià et al., 2009), (Sebastià et al., 2010), (Garcia et al., 2011), (García-Crespo et al., 2011), (Wang et al., 2011), (Alonso et al., 2012), (Batet et al., 2012), (Martínez-Santiago et al., 2012), (Lucas et al., 2013), (Ruotsalo et al., 2013), (Cha, 2014)

2.4.1. Multi-agent systems

Agents are autonomous and proactive software entities capable of obtaining information from their environment and acting in an intelligent way upon it in order to try to accomplish a set of goals or objectives. *Multi-agent systems* are groups of agents that communicate between themselves to share information and resources, coordinate their activities and cooperate in the joint efficient solution of a distributed problem (Wooldridge, 2009).

Turist@ (Batet et al., 2012) is an agent-based system that provides personalised recommendations on cultural activities. The architecture of the system is shown in Figure 18. There is one agent for each kind of cultural activity, which maintains a small database with the events of that type available in the city (museums are the exception, as there is one specific agent for each museum in the city). The user interacts with the system through a graphical interface provided by a *User Agent*. A *Broker Agent* mediates the communication between the *User Agents* and the cultural activities agents. The user can make specific queries, can evaluate an activity that he/she has attended, or can ask for a personalised recommendation. The core of Turist@ is the *Recommender Agent*, which maintains a user profile for each tourist. This profile is initialised with some basic information on high-level cultural interests provided by the user when he/she uses the system for the first time. The *Recommender Agent* dynamically and automatically refines this initial knowledge about the user preferences by analysing the user’s queries and evaluations. The *User Agent* can also provide proactive recommendations, because it knows the position of the user in the city and can suggest cultural activities that fit the user’s preferences and are located in the vicinity. The system uses both CB and CL recommendation techniques.

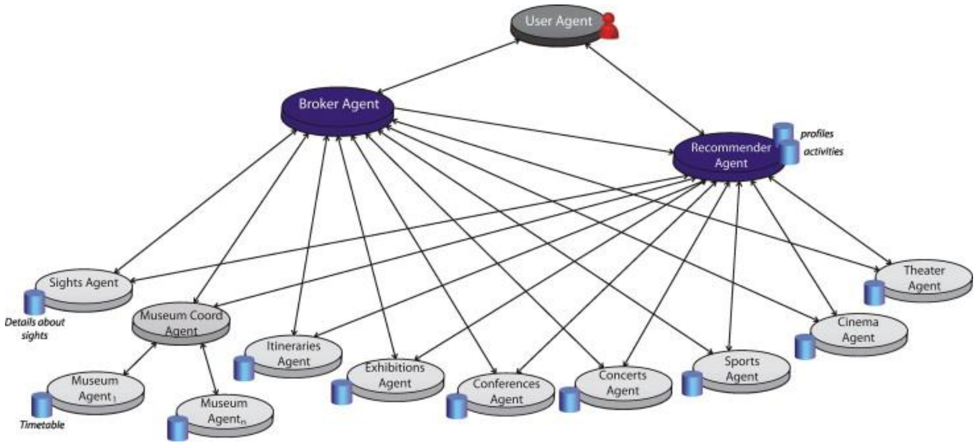


Figure 18. Multi-agent based architecture of Turist@ (from (Batet et al., 2012))

The idea of having an initial profile and refining it by analysing the explicit (evaluations) and implicit (actions) activities of the tourist is also given in (Ceccaroni et al., 2009). That work proposes to have a *Profile Management Agent*, which not only initializes the profile (by fitting the user into stereotyped classes) but also modifies it depending on the feedback provided by the tourist. In this agent-based proposal there are *Information Service Agents* that retrieve touristic information from databases and ontologies, and a *Personalization Agent* that, given the user profile and the available touristic data, applies CB recommendation techniques to select the items that should be suggested.

In *PersonalTour* (Lorenzi et al., 2011) there is a set of *Travel Agents*, and each of them is specialised in the recommendation of flights, hotels or attractions. When a new costumer arrives and expresses his/her preferences, these agents collaborate among themselves in order to propose a travel package to the tourist. The user can later evaluate each of the components of the package, providing a feedback to the system so that the degree of expertise of each *Travel Agent* can be conveniently updated.

Some recommenders (e.g. (Castillo et al., 2011; Sebastiá et al., 2010; Lee et al., 2009)) "agentify" the different components of the system (the interface with the user, the capture of his/her requirements and preferences, the analysis of the suitability of each attraction, the creation of a route among the selected points of interest), although there is not any kind of complex communication or coordination between them. In all these systems the agents seem to work in a sequential fashion, without any kind of coordinated effort. Therefore, the full potential of distributed, concurrent and coordinated behaviour of agents is not employed.

2.4.2. Optimization techniques

Many Tourism recommender systems have to solve complex planning and scheduling problems, which are well known to be NP complete and, therefore, cannot be optimally solved in an efficient way. In some cases, researchers have opted for the use of different kinds of optimization techniques which, although in many cases they do not guarantee the optimal solution, offer an affordable computational cost.

One example is the agent-based travel route recommender for Tainan (Lee et al., 2009), that uses *ant colony* optimization techniques. In these methods a set of autonomous entities (which represent the *ants*) cooperate through pheromone-mediated indirect and global communication to find a good solution to the travelling salesman problem (in this case, to plan a route that goes through different points of interest around the city). *CT-Planner4* (Kurata and Hara, 2013) uses a *genetic algorithm* to construct the plan to visit a city. In each iteration of a cyclic process it considers a population of different possible plans, which are evaluated according to their utility for the user; the best ones are mutated and recombined via crossover to generate another population for the next iteration. After a certain number of iterations, the best plan is finally selected. The authors of the *VISIT* system (Meehan et al., 2013) propose to make recommendations adapted to the context of the user, that is composed of different factors (location, time, weather, social media sentiment and user preferences). In that work, they suggest the idea of using an *artificial neural network* to assess the relevance of each context component for each user.

Some heuristic procedures to build travel itineraries were explored in the *City Trip Planner* system and related works (Vansteenwegen et al., 2010; Garcia et al., 2010; Garcia et al., 2013b). One possibility is the use of *Iterated Local Search*, a meta-heuristic iterative method that builds sequences of solutions generated by a local search. The heuristic perturbs the solution found by the local search (a route to visit some city attractions) to create a new solution. Then, it takes the best solution as the new starting point for the local search. The process is repeated until a termination criterion is met. Another option that was studied is the use of meta-heuristic iterative *Greedy Randomised Adaptive Search* methods (Souffriau et al., 2011). In each iteration a list of possible visits is generated from an initial solution which contains only the start and end of each tour. Those visits that have a heuristic value below a certain threshold are eliminated. A random visit from the remaining list is selected and applied to the current solution.

Most of the Tourism recommender systems that build personalised routes or itineraries implement an ad-hoc planning mechanism, but some of them apply more classical domain-independent AI planning techniques. For instance, in the *SAMAP* system (Castillo et al., 2008) the use of heuristic, A* and hierarchical temporal planners was explored.

2.4.3. Automatic clustering

Many Tourism recommenders employ techniques based on CL filtering, in which the users of the system are partitioned into groups that share some common characteristics. The basic idea of these methods is that it can be appropriate to recommend to the user those items that have been positively valued by similar tourists. The concept of similarity employed to group users may be based on demographic information, on the general preferences of the users over diverse types of touristic activities, or on the explicit ratings of individual activities. In any case, the automatic clustering tools developed in AI may be successfully used to classify the tourists. This section comments different alternatives that have been used in touristic recommender systems.

A very simple way of associating a new user with similar past users of the system is to employ the *k-nearest neighbours* approach (Dasarathy, 1991), calculating which are the *k* past users of the system who were more similar to the current one (e.g. (Martínez et al., 2009; Noguera et al., 2012)). Having done that, the information on those users may be employed to provide recommendations (e.g., the activities that were more highly valued for them). In *SAMAP* (Castillo et al., 2008) the similarity between users is based on the preferences expressed over the concepts of a domain ontology (a portion of it may be seen in Figure 19). For instance, the system could easily infer that a user that likes *Cinema* is more similar to a user that enjoys *Theatre* than to another that prefers *Sport* activities. Scalability is one of the main problems to be addressed when using this method.

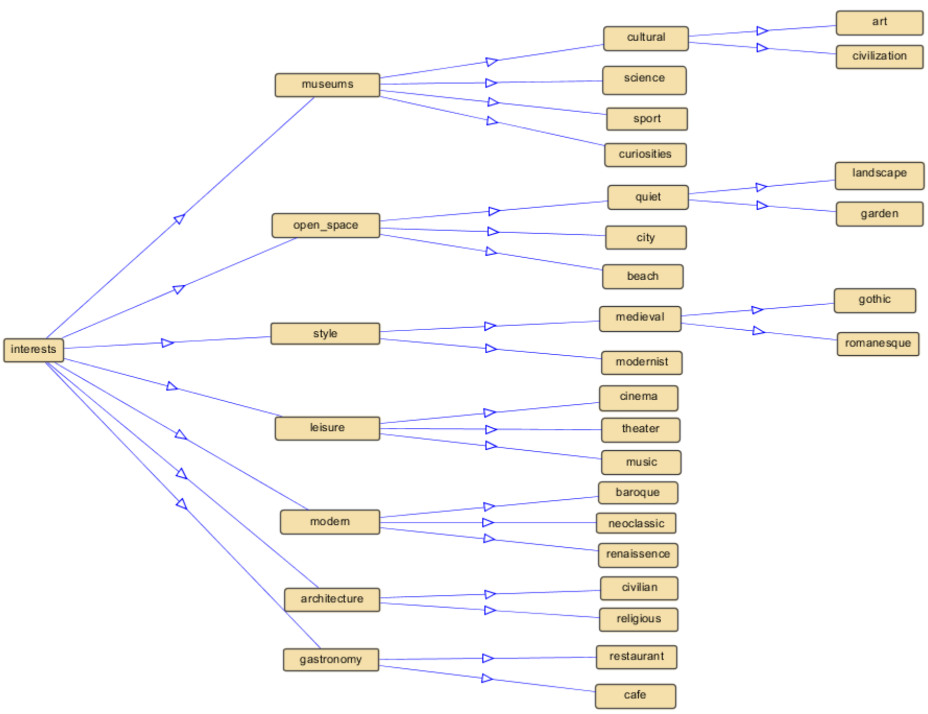


Figure 19. Portion of the SAMAP domain ontology (adapted from (Castillo et al., 2008))

A common option to group the users into different classes is to use the *k-means* algorithm (e.g. (Gavalas and Kenteris, 2011; Pinho et al., 2011)). The initial seeds of the *k* desired clusters are established in some application-dependent way. After that, there is an iterative process in which, in every step, the objects are sorted into the nearest cluster and the cluster prototypes are recalculated. The method converges to a solution when the objects belong to the same clusters in two consecutive iterations.

The recommender system described in (Fenza et al., 2011) proposes the use of the uncertain version of k-means, *fuzzy c-means*. The result of this algorithm is a fuzzy partition of a set of objects into clusters, so that each object has a degree of membership between 0 and 1 to each cluster, and the addition of the degrees of membership to all the clusters is 1. This algorithm is both applied to users and to touristic *points of interest* (POIs). After the definition of clusters of users and POIs, the system is able to derive rules that characterize them, that are used to integrate new users and new POIs to the clusters in which they fit better. This work also proposes to build *association rules*, which explain the relationship between clusters of users (plus contextual information) and clusters of POIs. These rules permit to determine the kind of touristic activities that should be recommended to a certain type of users. Very similar techniques are employed in the *PSIS (Personalised Sightseeing Information System)* recommender (Lucas et al., 2013).

Turist@ (Batet et al., 2012) also employs CL filtering recommendation techniques that require the definition of classes of similar users. The clustering is applied every time that 10 new users join the system, so classes are periodically recomputed. The employed clustering system is *ClusDM* (Valls, 2003), which builds a hierarchy of classes taking into account the interests of the users in general kinds of activities and their demographic data. The tree generated by the algorithm can be cut at different levels to generate partitions with the desired number of classes.

The use of *Support Vector Machines* (SVMs) as a classification technique in Tourism recommenders is suggested in the *SPETA* system (García-Crespo et al., 2009). Tourist preferences on several kinds of activities are stored in a vector, and the characteristics of each activity are also stored in the same way. Thus, SVMs may be used to compute the distance between the user's preferences and the recommendable items, so that the most appropriate ones can be efficiently found.

Another way of using clustering can be seen in (Han and Lee, 2014). POIs are classified in different clusters based on their geo-localization and attributes. Principally, POIs are recommended in groups that have similar characteristics and are not far away between them. Hence, they can recommend groups of places that can be visited in one day in terms of time and space.

2.4.4. Management of uncertainty

The task of recommending activities to a tourist is not simple, as there is not any clear and precise relationship between the characteristics and preferences of a visitor and the POIs available at a given destination. Some of the techniques developed in the AI field of *approximate reasoning* have been proposed to represent and reason about this uncertain relationship.

One possibility is to use *Bayesian networks* (Pearl, 1988). A Bayesian network is an acyclic graph in which edges represent relationships of causality or influence between nodes. Nodes that do not have any parent have an associated probability table, indicating how likely they are to occur. Nodes that have n parents have a conditional probability table of 2^n nodes, indicating how likely they are to occur depending on the presence (or absence) of their parents. A very simple use of Bayesian networks is presented in (Hsu et al., 2012), where a number of attributes (age, nationality, occupation, income, travel motivation, etc.) influence directly on the probability that a certain touristic point is interesting for the user. The initial Bayesian network was built after the analysis of more than 2400 questionnaires. A more complex application of this kind of networks is given in (Wang et al., 2011; Huang and Bian, 2009). They propose a network (see Figure 20) in which the age, occupation and personality influence the type of user which, along with the travel motivation, influences the probability of the user liking a certain kind of touristic destinations. Specific touristic events are not included in the network. (Wei et al., 2014) build a more complete Bayesian network that models not only demographic features but also the trip context (weather, time, etc.) and user behaviour (Figure 21).

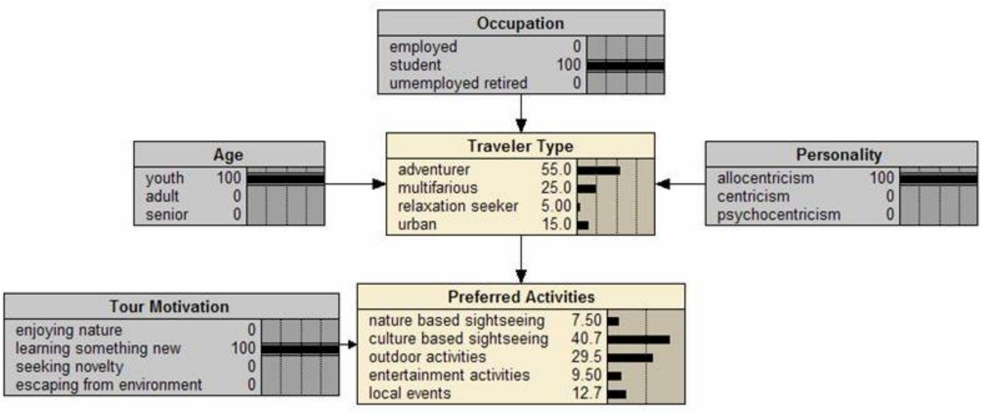


Figure 20. Use of a Bayesian network to detect the preferred kind of tourist activities (from (Huang and Bian, 2009))

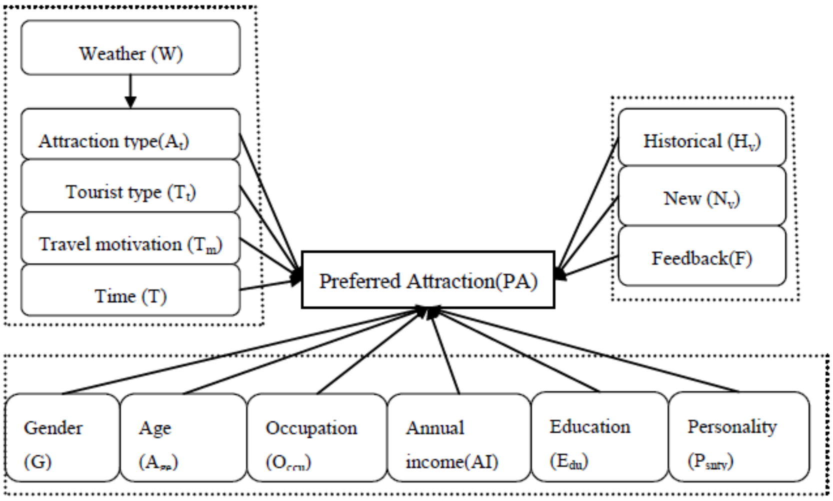


Figure 21. Bayesian network model of travel attractions (from (Wei et al., 2014))

Another common option to manage uncertainty is the use of *fuzzy logic*. A fuzzy variable make take as values a series of linguistic labels. Each linguistic label has an associated fuzzy set, in which every value in the domain of reference is assigned a membership value to the set between 0 and 1. In that way, fuzzy logic provides a generalisation of standard logic. Fuzzy sets and fuzzy reasoning may be used to represent the preferences of the user and to calculate how they fit with the characteristics of a tourist attraction (García-Crespo et al., 2009; Lee et al., 2009), to obtain the degree of membership of each user to different groups of users (Pinho et al., 2011) or to represent contextual aspects of the journey (Meehan et al., 2013). For instance, if the weather conditions are represented with a value between 0 and 1, instead of using a simple Boolean value for good/bad weather, it is possible to make a more fine grained analysis of the weather conditions and reason about its influence on the recommendation of each cultural activity.

Some touristic recommender systems also employ a *rule-based* approach, but without the addition of a fuzzy component. For instance, in the *CONCERT* system (Lamsfus et al., 2009; Lamsfus et al., 2011) there are rules that detect the events to be recommended depending on the user preferences and the context, such as this one:

hasFoodPreferencesRule: (v? red:type dcl:Visitor), (?v dcl:hasPreferences ?p) , (?p red:type dcl:FoodPreferencesDemographics), (?v dcl:usesDevice ?d), (?d dcl:isConnectedToNetwork ?n), (?n dcl:hasLocation ?l), (?l dcl:hasEnvironment ?e), (?e dcl:offersKindOfTourismConcepts ?s), (?s dcl:isRestaurantOfTypeVegetarian ?r) => print (?r dcl:isTourismServiceOfferedToVisitor ?v)

2.4.5. Knowledge representation

Recommender systems in e-Tourism need, as any knowledge-based intelligent system, a way to represent in an efficient way the domain knowledge, so that it can be used in their reasoning processes. The knowledge representation and reasoning techniques developed in AI are adequate tools for this purpose. In particular, nowadays the most common way of representing domain knowledge is the use of ontologies. An *ontology* describes a shared and explicit formal conceptualization of a given domain. Its main components are classes (representing concepts, usually organised in some kind of hierarchical structure), taxonomical and non-taxonomical relationships (Sánchez and Moreno, 2008a; Sánchez and Moreno, 2008b), axioms (Sánchez et al., 2012) and instances (representing specific objects).

There are several Tourism recommenders that employ ontologies to formalize the domain knowledge. Most systems have generic ontologies that store information about different aspects that have to be taken into account in the recommendation of cultural activities. Chapter 3, which is focused on the representation of the user profile using ontologies, treats this issue in more depth.

2.5. Related reviews

There have been some previous reviews explaining the application of recommender systems in the Tourism area, which are chronologically mentioned in this section. (Ricci, 2002) and (Staab and Werthner, 2002) explain in a very generic way the characteristics of travel recommender systems with some examples, without making an exhaustive review or comparing different approaches. (Werthner, 2003) gives also a very generic description of technological approaches applied to Tourism, where some examples related to Artificial Intelligence are mentioned. However, it does not provide any review or comparative analysis of different systems. (Berka and Plößnig, 2004) provides a brief guide on how to design recommmender systems for Tourism, but it does not attempt to make a survey of the area either. The survey that is more similar to this work is (Kabassi, 2010). It is mainly a classification of Tourism recommender systems (until early 2009) under different criteria: kind of objects they recommend (hotels, flights, restaurants, etc.), hardware support (computer or handheld device), individual/group recommendations, explicit/implicit acquisition of information from the user, recommendation technique (content-based, collaborative or hybrid) and personalization techniques (mainly decision-making tools and Bayesian networks). The authors of that paper basically group the systems in these categories, without making a deep analysis or explanation of all these possibilities. It does not provide any guideline on how to build this kind of systems and it does not consider the latest advances in the last five years, which are the basis of our study (advanced geolocalisation capabilities of mobile phones and tablets, context-aware recommendations, semantic management of preferences, use of social networks, etc.). (Gretzel, 2011) makes an analysis of Tourism recommenders from the point of view of social sciences, not from the technological perspective. The author of this paper argues that intelligent systems are necessary in the Tourism domain because there are many complex aspects to be managed: the mobility of tourists, the increased risk and uncertainty experienced in unfamiliar environments, the distributed nature of information sources, the idiosyncratic quality of tourism decision-making, the multi-faceted nature of tourism experiences, and the interdependency of subdecisions. A description of some systems that tackle those issues is done. The author also comments the main issues on the design and the evaluation of those systems, focusing on the user interaction, the context, the social perspective and the decision making process to maximize tourists utility; however, this work does not cover the use of intelligent techniques.

The main recommendation methods applied in Tourism are reviewed in (Felfernig et al., 2007). This paper presents some examples of the use of these techniques, but they are not deeply described nor compared. This paper emphasizes some interesting topics like group recommendation and context-aware recommendations in mobile devices. Finally, it is worth mentioning the paper (Vansteenwegen and Souffriau, 2011), that makes a deep overview of systems built between 2001 and 2011 that compose trip plans, although they only comment this single functionality. The authors compare each of the reviewed references in terms of these planning functionalities: personal interest estimation, selection and routing, mandatory points of interest, dynamic recalculation (update plan in real time when unexpected events occur), multiple day decision support (enable plans for multiple days), opening hours, budget limitations, max-n Type (limitation of activity types per day), mandatory types, weather dependency, scenic routes (build paths with beautiful views rather than the shortest ones), hotel selection, public transportation and group profiles. This paper describes how the orienteering problem and its extensions can be used to model trip planning functionalities.

In summary, as far as we know, there is not any recent survey of Tourism recommenders with the technological focus, novelty and breadth of coverage of the review presented in this chapter.

2.6. Conclusions

Tourism recommender systems give personalised and relevant suggestions to tourists whenever they intend to visit unknown places. They provide support tools to make the process of deciding what to do more manageable. In this chapter we have reviewed Tourism recommender systems published mainly in AI-related scientific journals and conferences since 2008.

We first analysed the interfaces used by these systems and we pointed out the predominance of Web-based approaches, which are especially useful for tourists when they are planning a visit before the stay. However, lately the usage of mobile platforms has widely increased, since they allow a direct access to the information about attractions during the stay. Moreover, they also permit to personalize and contextualize the gathered information, for instance taking the current location of the user into account. However, we have noticed that new mobile platforms such as Android or iPhone have been weakly exploited. Since these platforms are currently being widely used for tourists, it is necessary to address the development of applications for those systems and to create responsive Web designs that permit to adapt the content to any viewing device. Tourism recommender systems, as we have seen, not only manage textual information, but most of them use images, pictures and interactive maps. **Therefore we consider crucial the design of both a Web and a mobile platform for a recommender system, as we have done in this thesis.** In the Web version we take profit of large screens to visualize a complete plan of items with their multimedia content and geographic localization. In the case of the mobile version we take advantage of the contextual information to improve the accuracy of the recommendations and adapt it to the dynamically changing circumstances of the trip.

Recent recommender systems, known as *social recommender systems* (Noel et al., 2012) exploit the power of social networks. In addition to offering social functionalities, these tools facilitate the use of collaborative filtering techniques, since this kind of technologies permit new forms of rating items or collecting user information at an individual level or at a social level. These tools can be used both to identify groups of similar items and to build groups of like-minded users. For example, in *moreTourism* (Rey-López et al., 2011) the users have an associated tag cloud with terms relevant to their profile, and a new tag is created for each attraction based on the tags of the users who liked it. This information is used to compare the tag clouds of users and items and find coincidences. *TasTicWiki* obtains information about the user interactions with the items by analyzing the searches, readings and editions in a wiki (Ruiz-Montiel et al., 2010). This information is used to calculate the satisfaction degree that an article in the wiki has for a certain user. Another example is found in *SPETA* (García-Crespo et al., 2009), which maintains a social network profile of the users, so that their contact data are taken into account in order to analyze the interactions between them. *Trust* is another component that appears when dealing with social recommenders. It has been argued that ratings of credible users should be treated with higher weights than others (Gavalas and Kenteris, 2011). **In our thesis work, we make use of social networks to allow users to share their trips with their contacts. Moreover, one of the developed systems contains a section where users can check trips from other users.** However, we have left for the future work the study of new ways of exploiting all the data provided by social networks and other Web 2.0 applications, including the relationships between users and the different kinds of content they provide (comments, pictures, ratings), to improve the information that the recommender has on their interests. This aspect is certainly very relevant in the Tourism field, due to its highly social nature. Thus, it is important to include in Tourism recommenders as many possibilities of sharing information (pictures, videos, comments, ratings, localisation, etc.) as possible. The analysis of the social relationships of the users is a recent area of work that can surely lead towards the discovery of more accurate recommendations that fit better with the user's tastes, by taking into account the opinions of his/her closest friends, weighting the opinions depending on the strength of the relationship with the acquaintance, etc.

The recommendation process is a crucial aspect in Tourism advisory systems, hence we have analysed the main mechanisms used in the reviewed articles. The most popular approaches use content-based, collaborative and demographic-based techniques. These techniques suffer from several problems when applied individually. Hence, **a good practice is the combination of several techniques together to overcome their drawbacks, as has been done in this dissertation**. A special characteristic in Tourism, which distinguishes it from other domains in which recommenders have been applied, is the mobility of the users, which may need recommendations in different moments and in different places. For this reason, this particular type of recommender systems has started to incorporate *context-aware* techniques. The success of this approach is due to the widespread use of mobile devices. Many Tourism recommenders run on phones, so the user's location can be used to guide the filtering of the items to be shown (Lamsfus et al., 2009; Kurata, 2011; Yang, 2010b). Not only the current location of the user is important, but also the places that have already been visited (Gavalas and Kenteris, 2011; Umanets et al. 2013). Other features that are considered as contextual information in Tourism recommender systems are, for instance, the current weather to decide if it is more appropriate to recommend indoor or outdoor activities (García-Crespo et al., 2009; Gavalas and Kenteris, 2011; Braunhofer et al. 2013) or the motion speed and time to generate plans (Noguera et al., 2012). In the system described in (Niaraki and Kim, 2009) a complex model of the context is considered for constructing personalised route plans. The context information is organized on a hierarchy, including aspects related to the traffic, weather, safety (like telephone booth, side road parking, medical centre, etc.), facilities (gas station, etc.) and tourist attractions (fishing zone, recreation place, seaside, etc.). In (Amato et al., 2013b) four main parameters for the context are set: (i) time (time needed by the user to reach the place, the opening/closing times, etc.); (ii) location of the user and the place; (iii) weather and environmental conditions (e.g. temperature, humidity, rainfall degree, wind, season, moment of the day, etc.); (iv) social factors (number of users close to the place and number of positive/negative feedbacks). Moreover, the same authors extended their work (Amato et al., 2013a) to indoor scenarios to analyze room crowd, room fitness, network performances, location and time interval. They use a pre-filtering strategy to select those alternatives that satisfy the user's needs and a post-filtering strategy to arrange the recommended items based on their contextual values. This dimension is devised as a crucial point in the success of recommender systems in Tourism, due to the inherent mobile behaviour of the users in this specific application domain. **In our work, we take advantage of contextual data to improve the quality of recommendations**. We have modelled information of location, travel dates, opening-closing days of items, budget and size of the travel group to pre-filter and post-filter those items that fit with the users' conditions. We consider a good improvement the use of weather information or other related contextual data for the future.

It is also becoming increasingly clear that, in order to provide precise recommendations, it is necessary to move away from purely textual information and represent in a semantic way (e.g. through the use of ontologies) both the preferences of the user and the features of the different kinds of cultural and leisure activities. Having this structured information, it is possible to define and use complex semantic similarity techniques to compare users, compare objects or compare the preferences of the user with the characteristics of the objects. We have designed a **Tourism domain ontology** for the system that allows to classify objects and manage user preferences. Moreover, **a new framework that exploits the ontology hierarchy has been designed to represent and reason about user preferences**. These preferences can be acquired explicitly or implicitly. The most common method is the acquisition of explicit information. However, we consider applying a combination of both methods. Even though implicit information is inherently more uncertain, it is also less intrusive for users and it is easy to collect it directly by monitoring their interaction with the system.

Content-based systems focus on recommending items similar to the user's profile, which may cause overspecialized results, leaving aside other items that might be interesting for the user. This is an important issue in some applications in the field of Tourism. Some recommender systems aim at making publicity of "different" or new sorts of activities which may be ignored by most visitors (e.g. a new restaurant or a new guided tour). It has also been argued that a smart recommender should provide a diversified list of recommendations (e.g., even if the system knows that the user is interested in going to the beach, it is not very exciting to show a list of ten different beaches and not to suggest other kinds of related activities). In (Savir et al., 2013) a measure of balance between the number of attractions of a certain type and the minimum rating threshold is proposed in order to keep a fixed diversity level in the activities proposed in a trip. In (Ruotsalo et al., 2013) the objects of a museum are gathered in clusters sharing the same features so that the recommendation procedure picks a representative number of objects from each cluster to increase the diversity of the proposal made to the visitor. In this line, we have studied the diversification mechanism applied in recommender systems. Some of them have been tested in our approach and compared against a **new diversification method based on semantic clustering**.

In the remainder of this thesis we explain how the main issues that have appeared in this review have been handled. In particular, the next chapter describes a novel way to manage uncertain preferences using semantic domain knowledge. After that, in chapter 4 a clustering-based diversification method, also based in the use of a domain ontology, is proposed and compared with existing techniques. Finally, in chapter 5 we show how we have developed a Tourism Recommender System that takes into account the conclusions taken in this review: the use of Web and mobile platforms, the use of semantic knowledge to manage preferences, the combination of different recommendation mechanisms, the use of contextual information and the application of an explicit diversification mechanism.

Chapter 3 – Ontology-based management of uncertain preferences

In the current context of information overload, people are daily confronted with many situations in which a decision must be taken in the presence of a wide set of alternatives defined on a large number of criteria or attributes. *Recommender systems* (RS) can be very helpful in these situations, because they can analyse automatically all the information available on the possible alternatives, compare it with the user preferences or interests, rate the alternatives and present to the user the most appropriate ones. The representation and management of the user preferences is a key component in RS because the solution must be based on the user interests and needs. Thus, a basic component of RS is the *user profile*, which stores the information about the user's preferences on the domain.

A current research trend is the design of *semantic recommender systems* (SRS), in which the semantic information about the domain, usually represented in the form of an ontology, is used to represent both the user profile and the recommendable items. As pointed out in (Cantador and Castells, 2011), SRS provide the benefits of *semantic richness* (preferences are richer and more detailed than the standard ones based solely on keywords), *hierarchical structure* (allowing an analysis of preferences at different abstraction levels) and *inference capabilities* (the structure of the ontology may be used to reason about the preferences of the users on all the domain concepts). The comparison between two values using keywords is simply based on their equality/inequality (and sometimes is related with some kind of ordering of the categories), due to the lack of proper methods for representing the meaning of the terms. Using semantic variables it is possible to establish different degrees of similarity between values (*e.g.,* "trekking" is more similar to "jogging" than to "cooking"). Semantic similarity functions between semantic values usually depend on the ontological knowledge available for the domain of discourse (Jiang and Conrath, 1997; Resnik, 1995; Sánchez et al., 2010).

In this chapter we present a *semantic-based approach to store and exploit the personal preferences of a user* with respect to a complex domain. Recent Artificial Intelligence knowledge models, such as ontologies, provide tools for representing the elements of a certain domain (i.e. concepts), as well as their interrelations, in a machine understandable language. They allow mapping words to concepts, so that terms can be interpreted according to their taxonomical and semantic relations with other terms (Studer, et al., 1998). These models facilitate the design and implementation of reasoning tools that exploit the knowledge they store. A great effort has been done in some communities to develop shared domain ontologies. A paradigmatic example is the definition of shared vocabularies and thesaurus in Medicine, like SNOMED CT[8] (*Systematized Nomenclature of Medicine, Clinical Terms*), which is an ontological/terminological resource distributed as part of UMLS (*Unified Medical Language System*). It is used for indexing electronic medical records, ICU monitoring, clinical decision support, medical research studies, clinical trials, computerized physician order entry, disease surveillance, image indexing, consumer health information services, etc. Another example is the *Thesaurus on Tourism and Leisure Activities* developed by the World Tourism Organisation[9].

Thanks to the availability of these large, detailed and generally accepted ontologies, a new generation of ontology-based techniques is appearing (clinical support systems (Pisanelli, 2004), semantic clustering (Batet et al., 2010; Aseervatham and Bennani, 2009), semantic anonymization (Martínez et al., 2012), semantic browsing of digital document resources (Collins et al., 2005), etc.). In particular, SRS use the semantic knowledge stored in the ontology to provide personalized and accurate recommendations to the user. In this case, the ontology is usually tailored to store the degree of interest of the user with respect to each of the concepts of the domain.

Some authors have already proposed works with ontology-based user profiles, in which the ontology components (especially the concepts and the taxonomic relationships between them) are used to spread preference information through the ontology, to compare users to form clusters of people with similar tastes (in collaborative filtering systems) or to match the user preferences with the representation of each item (in content-based RS). In those systems the user profile is usually built and maintained through explicit information provided by the users or by analyzing their interaction with the system. The work presented in this chapter considers the *uncertainty* associated to these kinds of information and proposes a general framework that allows representing and reasoning about the uncertainty associated to preferences in ontology-based SRS.

This chapter starts with an introduction to ontologies and its main features. Then, in section 3.2 it is explained how ontologies have been used in recommender systems, both to represent the domain items and the preferences of the users, followed by a review of approaches that apply ontologies in these systems. Afterwards, we propose a new framework for managing uncertain preferences exploiting the hierarchy of an ontology domain.

---

[8] http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html (last access on March 2015)
[9] http://www.wtoelibrary.org/content/m7434p/ (last access on March 2015)

3.1. Ontologies

Ontology-based intelligent systems have powerful modelling and reasoning capabilities. The use of explicit domain knowledge, represented in the form of an *ontology*, permits a high degree of knowledge sharing, logic inference and knowledge reuse (Wang et al., 2004). These knowledge structures basically describe the main concepts (and the relationships between them) in a particular domain, along with their properties and restrictions on their use, giving a precise meaning to each concept. Ontologies have several components on which intelligent systems may apply reasoning procedures. The main ones are classes, instances, properties and rules. A brief explanation of these features and how they are used in some ontology-based systems is given in the following list:

- *Classes* are the abstract representation of the different concepts of a domain. They usually correspond with the nouns found in the domain. For instance, a class could be 'city', 'accommodation' or 'singer'. Each class has a certain number of features, represented with *slots*. For instance, the 'singer' class could have slots identifying aspects like the birth place of the singer, his birth date, his number of Grammy awards, etc.
- *Instances* of a class represent specific individuals that belong to that class of objects. For example, in the Music domain we may have instances of the class 'singer' like 'Elton John' or 'Madonna'. In the Tourism domain, 'Berlin' and 'The Plaza Hotel' are instances of the classes 'city' and 'accommodation' respectively. Instances have a particular value associated to each of their slots, including those slots inherited from all their superclasses.
- *Properties* permit to establish binary semantic relationships between classes. The most common is the '*is-a*' property which indicates that a class is subclass of another class. For example, 'football' is-a 'sport' means that the 'football' class is subclass of 'sport' (and, therefore, it inherits all its characteristics). This property defines a taxonomical structure of classes, which is normally a tree or an acyclic graph. Any other property between classes is considered non-taxonomical. For example, we could define the property 'locatedIn' between 'accommodation' and 'city', and use it to indicate that 'The Plaza Hotel' is located in 'Berlin'.
- Ontology *rules* are the translation of mathematical axioms that impose some constraints on the objects that can be related via a certain property or on the values that a certain slot may take. These rules may be used by ontology-based systems to implement complex reasoning mechanisms. For instance, an axiom could specify that a certain binary relationship P between classes has the transitive property; then, if the system knows that aPb and bPc, it can infer that aPc.

Table 6. List of ontologies and their main features.

Ontology	Properties (examples)	Inst.	Rules
IPTC ontology (Cantador, 2008)	is-a, SubjectQualifier, MediaType, Gender	Yes	No
(Ceccaroni et al., 2009)	is-a, isPermormedAt, uses, isAbout	Yes	No
SPETA (García-Crespo et al., 2009)	is-a, locatedIn, interestedIn, hasCurrency	Yes	Yes
OntoMOVE (Bhatt et al., 2009)	SubClass, EquivalentClass, DisjointWith, SameIndividual, differentFrom	Yes	Yes
(Middleton et al., 2009)	is-a (3 levels)	Yes	No
ContOlogy (Lamsfus et al., 2010)	is-a, type	No	No
CRUZAR (Mínguez et al., 2010)	subClassOf, partOf, hasQuality, location, date	Yes	Yes
OntoCrawler, OntoClassifier (Yang, 2010a)	is-a	No	No
(Dongxing et al., 2011)	is_a, hasPart, hasFunction, useMaterial, hasProperty, hasFeature, has_Standard	Yes	Yes
e-Tourism (Garcia et al., 2011)	is-a	Yes	No
(Luberg et al., 2011)	is-a	No	Yes
(Ruíz-Martínez et al., 2011)	is-a	Yes	Yes
(Wang et al., 2011)	is-a	No	Yes
(Alonso et al., 2012)	is-a	Yes	Yes
(Debattista et al., 2012)	isComposedOf, hasConstraint, hasNegation, hasObject, hasSubject, etc.	Yes	Yes
(Di Noia et al., 2012)	genre, director, subject, broader	Yes	No
(Lemos et al., 2012)	participatesInEvent, isA	Yes	Yes
GeOasis (Martínez-Santiago et al., 2012)	has-visited, is-located-in, is-selected, is-in-area, is-part-of, is-point-of	Yes	Yes
(Parundekar and Oguchi, 2012)	hasName, hasLocation, hasAverageRating, hasCarWash	Yes	No
(Rospocher and Serafini, 2012)	hasData, hasConclusion, produceConclusion, etc.	Yes	No
(Bouneffouf, 2013)	is-a	No	No
(Cena et al., 2013)	is-a	Yes	No
(Moscato et al., 2013)	is-a	Yes	No
SMARTMUSEUM (Ruotsalo et al., 2013)	is-part-of	Yes	No
(Cha, 2014)	is-a	No	Yes
(Al-Hassan et al., 2015)	any object property	Yes	No

As an example, Table 6 shows the main features of an illustrative set of ontologies from different domains used by ontology-based recommender systems. Most of them have been designed and built ad-hoc for a particular system. All the approaches use the '*is-a*' relationship (or its equivalent form 'subClass' or 'is-part-of') in order to categorize the main domain concepts in a taxonomical hierarchy. Most of them also use more complex non-taxonomical relationships. For instance, (Cantador, 2008) defines an ontology about news that includes metadata elements like 'Subject Qualifier', 'Media Type', and 'Gender'. The Tourism ontology defined in (Mínguez et al., 2010) has properties like 'partOf', 'hasQuality', 'location' or 'date'. Another example is (Bhatt et al., 2009), that uses mathematical properties such as 'equivalent', 'inverse', 'transitive' or 'functional', among others. This permits description logic reasoners to exploit the ontology, deductively inferring new facts from the available knowledge. (Rospocher and Serafini, 2012) go one step further and use object properties not also to relate instances but also for explanation purposes: for instance, the 'ProduceConclusion' property allows keeping track of what data triggered a certain conclusion. (Dongxing et al., 2011) uses properties of documents in order to define different relationships like 'hasPart', 'hasFunction', 'useMaterial', 'hasProperty', 'hasFeature' or 'hasStandard'. In this approach, two types of semantic rules are employed to describe the low-level features of customer preferences and to build an ontological knowledge base. One is used to combine preference terms and concepts. For example, the term 'RED' and the concept 'SH-FLIP-PHONE' can form a new preference concept 'SH-RED-FLIP-PHONE'. The other rule type is applied to combine two concepts such as 'F-WORD' and 'F-TEXT' that produces the new concept 'F-WORD-TEXT'. This new concept generation is based on specific relationships, such as 'is_a', 'hasFunction' or 'hasMaterial', which give a meaning to the new concept. In the SPETA ontology (García-Crespo et al., 2009) there are properties like 'locatedIn (indoor or outdoor)', 'interestedIn', 'hasCurrency', etc. This ontological knowledge permits the system to answer questions like what activities can be visited by a certain type of tourists, which is the location of interesting places and when they can be visited. This information is inferred by using ontology rules, such as '*closeOnDate(?attraction, ?date)*' that specifies that the attraction is closed on a particular date. Rules in (Luberg et al., 2011) could be like '*fact(? X type architecture 0.9*?N) :- fact (? X type church ?N)*', which indicates that if an item belongs to the type 'church' with score *N*, it can also be considered of the type 'architecture' with a score 0.9**N*. Another approach using rules is (Alonso et al., 2012), that provides rules with meanings like '*if possible, P_1 is preferred with weight W_1, or if P_2 is not possible, then P_2 is preferred with weight W_2*'. (Lemos et al., 2012) infers new knowledge from previous facts (ubiquitous geolocation snapshot of user activity, call history, custom habits) or future plans (planned events) by applying rules based on Drools[10]. (Debattista et al., 2012) use case-based reasoning (CBR) techniques to automatically learn context-aware rules through the di.me Rule Management Ontology (DRMO[11]) that permits to make recommendations based on the user's context-aware history. SPARQL queries are used to trigger certain rules that are modelled on the Event-Condition-Action (ECA) pattern concepts. The ECA pattern is a structure used in event-driven architectures, where the event part specifies on what event this rule might be triggered, the condition specifies under which conditions the actions should be triggered and the action part contains what is executed to lead the system to a new state, causing data to be changed. Figure 22 depicts how a rule is represented with a *drmo:Event*, which is composed (*drmo:isComposedOf*) of a number of *drmo:Condition* '*blocks*' and triggers (*drmo:triggers*) and one or more *drmo:Action* instances.

[10] http://www.drools.org/ (last access March 2015)
[11] http://www.semanticdesktop.org/ontologies/2012/03/06/drmo/ (last access March 2015)

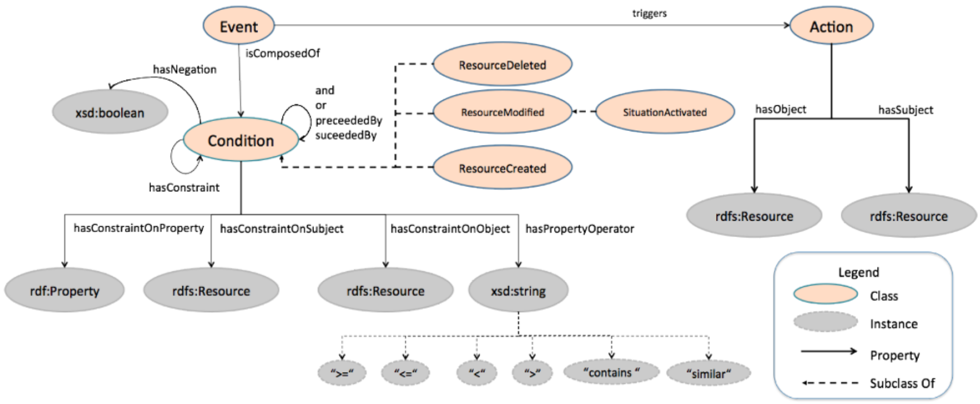


Figure 22. A Rule Management Ontology (from (Debattista et al., 2012))

3.2. Use of ontologies in Semantic Recommender Systems

The main objective of recommender systems is to predict the degree of interest of a user for an object given the user preferences and the features of the object (Montaner et al., 2003). The system can then provide to the user a ranked list with the alternatives that fit better with his/her preferences. Ontologies can be applied to extend the traditional text-based recommender systems with semantic domain knowledge, with the aim of improving the accuracy of the recommendations. The hierarchical organization of the concepts in ontologies permits to make a representation of both the characteristics of the alternatives and the users' preferences at different levels. Then, reasoning mechanisms can be applied to propagate the information through the ontology in order to make a suitable comparison of the properties of an object with the interests of a user, to compare the properties of different objects, or to compare the interests of different users. Most semantic recommender systems use ontologies to represent both the information about the alternatives and the knowledge about the user preferences. These two possibilities are commented in more detail in the following subsections.

3.2.1. Representation of alternatives

3.2.1.1. Semantic representation of domain items

In the context of a recommender system, the information contained in the ontology is normally used to represent the main features of the different alternatives that the user is considering. For instance the authors of (Garcia et al., 2011) designed an overall taxonomy in the Tourism domain to describe attractions in general categories such as 'Gothic Art', 'Museums', 'Religious Buildings', etc. The particular attractions were represented as the instances of this ontology. Another example in the field of Tourism (Huang and Bian, 2009) defines a different set of classes to organize the items, such as 'Attraction', 'Location', 'OpenTimes', 'AdmissionFees' and 'Activity'. In the music recommender system reported in (Celma and Serra, 2008), classes are used to describe the relevant features of a song, such as 'genre', 'singer', 'title', 'duration' or 'tempo'. The route planning system defined in (Niaraki and Kim, 2009) uses an ontology that represents road variables, like the traffic, safety, road facilities, weather conditions and attractions, to find the optimum path in the road network. Other systems like *GeOasis* (Martínez-Santiago et al., 2012), SMARTMUSEUM (Ruotsalo et al., 2013) and the one proposed in (Alonso et al., 2012) also include ontologies to model the different kinds of touristic activities and to be able to reason on them in a semantic fashion. These systems use ontology-based similarity measures to deduce if two kinds of activities are similar, and this knowledge may also be used to compute the similarity between users and provide recommendations based on collaborative filtering techniques.

There are systems that use several ontologies, which focus on different dimensions of the domain. For instance, in the Tourism recommender system shown in (Ruiz-Montiel and Aldana, 2009), there is a domain ontology formed by classes that describe implicitly the properties of a service (with classes such as 'Inexpensive Service', 'Accommodation Service' or 'Charming Accommodation Service') and a separate user ontology whose classes describe personal information, such as gender, age or touristic interests. *PaTac* (Ceccaroni et al., 2009) includes separate ontologies with knowledge about cultural activities, restaurants, entertainment, hotels, etc. (see Figure 23). They are linked with standard temporal and geo-location ontologies provided by the W3C[12] consortium and with a user model ontology that contains different kinds of touristic stereotypes. (Lamsfus et al., 2010) presents a semantic-based digital broadcasting contextual tourism information system. They have created a network of ontologies, called ContOlogy, which integrates 11 ontologies, 86 classes, 63 properties and 43 restrictions. These ontologies represent the information about visitors, preferences, roles, activities, environment, devices, network, motivations, location, time and tourism objects.

Most of the examples shown in Table 6 are based on ontologies that have been built ad-hoc to be used in the recommender system. However, different organizations and committees are defining public ontologies, which usually cover a larger set of concepts including many more different types of taxonomical and semantic relations. From our analysis of the recent literature, we have found few semantic recommender systems that make use of existing ontologies or vocabularies. These are the cases of (García-Crespo et al., 2009) using the YAGO ontology[13] and (Celma and Serra, 2008) with the RDF Site Summary[14] and FOAF (Friend of a Friend)[15] ontologies. The upper level ontologies in *PaTac* (Ceccaroni et al., 2009) are

[12] http://www.w3.org/ (last access March 2015)
[13] http://www.mpi-inf.mpg.de/yago-naga/yago/ (last access March 2015)
[14] http://web.resource.org/ (last access March 2015)
[15] http://www.foaf-project.org/ (last access March 2015)

based on various standards, such as W3C's Time[16], Geoposition[17], General User Model Ontology (GUMO) (Heckmann et al., 2007), FOAF and UMBEL[18]. SMARTMUSEUM (Ruotsalo et al., 2013) also employs a limited subset of the GUMO approach. (Debattista et al., 2012) represents contextual information with instances of the Context Ontology (DCON[19]). In this work past context snapshots can also be timestamped and made persistent as instances of the User History Ontology (DUHO[20]).

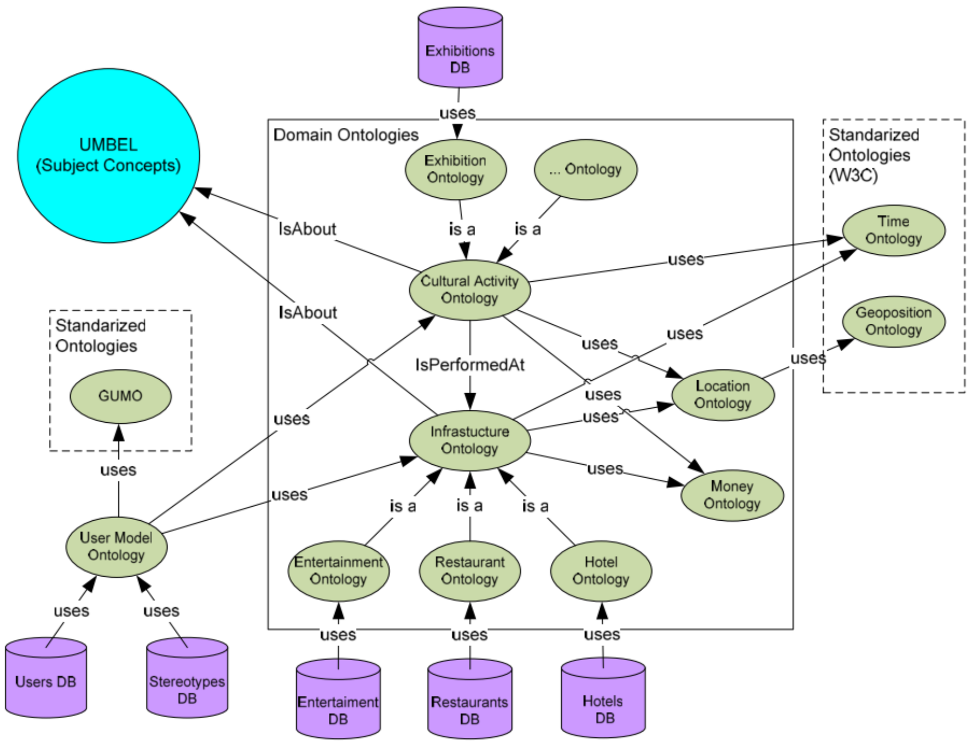


Figure 23. Multiple ontologies with relationships (from (Ceccaroni et al., 2009))

3.2.1.2. Ontology population

In general, alternatives are represented as instances of the ontology. In some cases, each alternative is restricted to be an instance of a unique class in the ontology. In this model, each alternative is associated to a single concept, for example 'The Plaza Hotel' is an instance of the class 'Accommodation' and of no other class. However, it is common that an alternative can be an instance of several disjoint classes. Sometimes the classes which are allowed to be instantiated are only the ones in the leaves of the taxonomy (i.e. the most specific concepts) (Garcia et al., 2011). When an alternative is associated to multiple concepts, they can also be referred as different

16 http://www.w3.org/TR/owl-time/ (last access March 2015)
17 http://www.w3.org/2003/01/geo/ (last access March 2015)
18 http://www.umbel.org/ (last access March 2015)
19 http://www.semanticdesktop.org/ontologies/2011/10/05/dcon/ (last access March 2015)
20 http://www.semanticdesktop.org/ontologies/2011/10/05/duho/ (last access March 2015)

annotations or keywords describing the alternative. The analysis of these multiple concepts requires some kind of multi-criteria approach.

The process of associating an alternative to the classes is called *initialization* or *ontology population*. If the set of alternatives is not fixed, some process for including new instances dynamically must be defined. In some cases, it may also be interesting to define a way to reduce the number of alternatives in the system, if they can be obsolete after a certain time (e.g., via "forgetting" rules). The initialization process can be done manually by a domain expert, who enters the information of each new alternative and instantiates it in the corresponding ontology classes. An expert criterion is used in (Albadvi and Shahbazi, 2009) to design and populate the ontology. Marketing managers defined the most important nodes, such as book, CD/DVD, story or comedy. Managers also defined grain nodes as a flexible way to apply multiple rules at a time by grouping similar rules together. Moreover, they defined category attributes such as price, brand, or size that are inherited from the product category. Different products were then associated to those categories. *e-Tourism* (Garcia et al., 2011) uses the edges linking an item with an associated value to indicate the degree to which an item belongs to a certain category (see Figure 24).

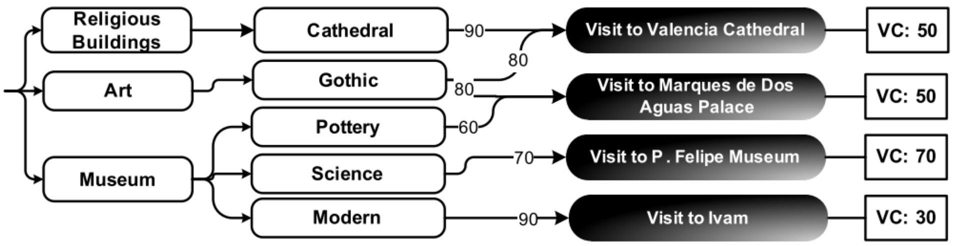


Figure 24. Ontology designed in e-Tourism (from (Garcia et al., 2011))

However, this manual process may be long, tedious and error-prone. A way to reduce the cost of the construction of the ontology (Ruíz-Martínez et al., 2011) is to populate it in automatic fashion, by analysing electronic resources (e.g. Web pages), extracting the appropriate information about tourist activities and creating the associated instances. A similar proposal was made in (Vicient et al., 2013). (Celma and Serra, 2008) developed a Web crawler that extracts metadata to fill up the ontology with instances of songs, artists or concerts. It also discovers automatically relationships between artists like 'isRelatedWith', 'isInfluencedBy' or 'isFollowerOf'. In (García-Crespo et al., 2009) the ontology is populated with a large number of instances extracted from DBpedia[21], which contains more than 2.49 million of structured items from Wikipedia[22]. (Cantador, 2008) analysed 137,254 Wikipedia entries to populate 744 classes with 121,135 instances. Other approaches consider the textual information provided by documents. In this case, some natural language processing tools are needed. For example, (Yang, 2010a) extracts information from documents using

[21] http://dbpedia.org/
[22] http://www.wikipedia.org/

computational linguistic techniques like normalization, segmentation, stop word filtering, word stemming and TF/IDF calculation (term frequency/inverse document frequency). Different weights are assigned to the keywords according to their level in the hierarchy. (Dongxing et al., 2011) analyses the frequency of the terms in documents (alternatives) to represent weighted features for each document. A similar procedure is done in (Middleton et al., 2009) which automatically constructs clusters of papers according to their similarity, to assign them to the same concepts in the ontology. (Debattista et al., 2012) uses several open data sources, such as Sindice[23] (to crawl linked-data resources from different sites) and LinkedGeoData[24] (that serves the geo-referenced information collected by OpenStreetMap[25] and makes it available in RDF). (Di Noia et al., 2012) reuses datasets publicly available in the Linked Open Data cloud like DBpedia and Linked Movie Database (LinkedMDB[26]). Figure 25 shows an excerpt of the graph containing objects and properties from these sources.

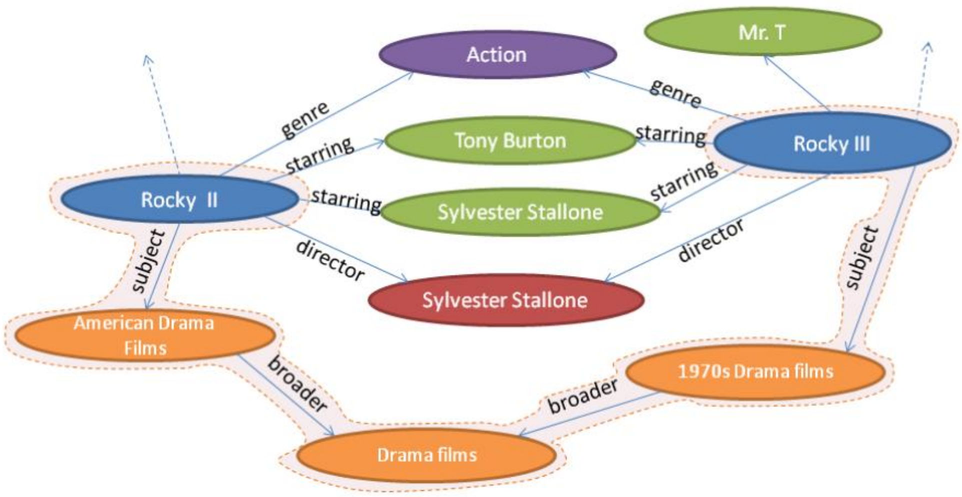


Figure 25. Sample RDF graph extracted from DBpedia and LinkedMDB (from (Di Noia et al., 2012))

3.2.2. Ontology-based user profiles

3.2.2.1. Semantic representation of preferences

The second use of ontologies on semantic recommender systems is in the definition of the user profile. Recommender systems need to know the preferences of the user. Different ways of making use of ontologies in the user profile can be found in the works we have studied. The simplest model associates to each user a list of keywords corresponding to the names of the classes in the ontology in which the user is interested (Shoval et al., 2008; Bhatt et al., 2009; Ruiz-Montiel and Aldana, 2009; Lamsfus et al., 2010). However, this kind of representation does not provide much information to the system. A more widespread approach consists on associating a vector of

[23] http://sindice.com/ (last access March, 2015)
[24] http://linkedgeodata.org/ (last access March, 2015)
[25] http://www.openstreetmap.org/ (last access March, 2015)
[26] http://www.linkedmdb.org/ (last access March, 2015)

features with the user. Each feature corresponds to a different concept in the ontology (i.e. a semantic category). Then, in each user's vector a rating of each feature is stored. This numerical value indicates the degree of interest of the user with respect to the concept (Hagen et al., 2005; Sieg et al., 2007; Cantador, 2008; Sendhilkumar and Geetha, 2008; Jiang and Tan, 2009; Middleton et al., 2009; Zheng, 2011). This vector approach facilitates the inclusion of other types of features in the profile, such as demographic information, as in (Mínguez et al., 2010; Codina and Ceccaroni, 2010; Wang et al., 2011; Garcia et al., 2011). Some works have also taken into account some measure of the credibility associated to the information stored in the profile. The rating values may be uncertain because the user gives an approximate score or due to the inference mechanisms used to obtain those values (as will be explained in section 3.2.2.2). A confidence degree can be associated to each rating in the profile and can be used as a weighting factor in the exploitation stage (Codina and Ceccaroni, 2010).

Finally, we can also find some works that build a specific tailored ontology for each user. In (Albadvi and Shahbazi, 2009) a subset of concepts from the ontology is selected by the user. Those concepts are considered as the ones relevant for the recommendation. In (Blanco-Fernández et al., 2011a) the user may select a subset of the concepts and attributes of the general ontology to generate its own "ontology of interest". Then a semantic network is created, whose nodes are the class instances selected in a pre-filtering phase. The ontology of interest is used to identify links that relate the nodes to each other. A degree of interest is associated to each node to reflect the significance of the relationship between the alternative and the user preferences.

In some decision aiding tools the user profile is not updated because the system is designed to solve a single problem once. In recommender systems the framework is completely different, since usually the goal is that the user becomes a usual client of the product. Therefore, it is crucial to maintain the user profile up to date in order to provide appropriate recommendations to the same person along time. The usual procedures for initializing and updating the ontology-based user profiles are presented in the next section.

3.2.2.2. Dynamic preference adaptation

In order to produce personalized recommendations to the same user along time, the system has to model his/her interests in the user profile and maintain them up-to-date. *Feedback* information is used to modify the profile when some change on the user's preferences is detected. Different type of data can be studied to model the user profile, as not only the user interests on the specific domain, but also the user context (such as the user location) is relevant. This information can be collected explicitly or implicitly (Marin et al., 2013).

Explicit feedback is obtained by means of the direct interaction with the user. The decision maker is requested to fill in some form (giving his/her opinion on different values of the criteria or indicating his/her location) or to rate a set of alternatives. This approach gives quite precise knowledge because the data is given directly by the user. However, it is usually considered quite an intrusive way of elicitation, and many users are not keen on spending time in answering this kind of questions.

Techniques based on *implicit feedback* aim at collecting the user information analysing his/her behaviour in the system, such as the alternatives that are selected, purchased or viewed. More sophisticated tools study the sequence of actions done by the user on a certain alternative, or even the amount of time spent with each alternative. The main advantage of these methods is that an additional effort from the user is not required. However, implicit information is more uncertain than explicit information, so less confidence must be given to it when the profile is modified.

3.3. Review of semantic preference management in recommender systems

When the user profile is based on ontologies, new techniques for the initialization and the adaptation of the knowledge about the user's preferences must be designed. This section reviews the main approaches to these questions.

Table 7 shows some details about the semantic recommender systems that define some kind of ontology-based user profile updating mechanism. The first four columns distinguish different techniques for the initialization of the profile. In early approaches like (Sieg et al., 2007), the concepts of the ontology that are associated to the user profile are obtained from the analysis of the queries that the user makes to the recommender system. Similarly, in SMARTMUSEUM (Ruotsalo et al., 2013) the user can search for concepts in an auto-completion field to indicate his/her interests. It is also quite common to obtain the initial description of the user by means of forms, which may contain questions about preferences and/or demographic data (Ruiz-Montiel and Aldana, 2009). Demographic information may be used to infer new preferences by analysing the relations in the ontology. (Niaraki and Kim, 2009) consider both preferences and demographic information (including age, gender, nationality, marital status, language, religion, socioeconomic conditions, residence location and ethnicity). Sometimes it is claimed that requiring so much information by means of forms is not appropriate because many users will abandon the system even before starting to use it. Hybrid approaches are used to alleviate this effect. For example one may use information about the user context to infer some of this data, given that the personal characteristics determine the human behaviour and the behaviour determines the context, and viceversa. In

(Lamsfus et al., 2010), the system stores the user's context, such as the weather, the location and the time of the day, which are gathered from Internet or from the mobile device of the user. In (Niaraki and Kim, 2009) the authors propose a model that relates the user profile with the contextual information. Another example that models the context is SMARTMUSEUM (Ruotsalo et al., 2013), which stores the user's GPS location, duration of the visit or companion. The context of the profile managed in (Bouneffouf, 2013) is based on location, time and social parameters (e.g people that are near the user). They use the context to determine the user's interests in the current situation. For instance, a tourist may be interested in food when he/she travels whereas he may be more interested in sports when being at home. Figure 26 shows their ontologies for each type of context.

Table 7. Ontology-based profile management

Reference	Initialization				Update		Domain inference
	Queries	Form about Preferences	Demo-graphic Form	User Context	Explicit	Implicit	
(Sieg et al., 2007)	●					●	●
(Wang and Kong, 2007)		●			●		
(Cantador, 2008)	●	●	●	●	●	●	●
(Sendhilkumar and Geetha, 2008)	●					●	
(Shoval et al., 2008)		●				●	
(Albadvi and Shahbazi, 2009)					●	●	
(Ceccaroni et al., 2009)		●	●		●	●	
(Jiang and Tan, 2009)	●					●	●
(Bhatt et al., 2009)	●				●		●
(Middleton et al., 2009)						●	●
(Niaraki and Kim, 2009)		●	●	●		●	
(Partarakis et al., 2009)						●	
(Ruiz-Montiel and Aldana, 2009)		●	●		●		
(Codina and Ceccaroni, 2010)		●			●	●	●
(Lamsfus et al., 2010)				●			●
(Blanco-Fernández et al., 2011a)		●			●		●
(Alonso et al., 2012)				●	●	●	●
(Debattista et al., 2012)				●		●	●
(Di Noia et al., 2012)		●			●		
(Lemos et al., 2012)		●		●		●	
GeOasis (Martínez-Santiago et al., 2012)	●			●		●	●
(Parundekar and Oguchi, 2012)	●			●		●	
(Rospocher and Serafini, 2012)	●	●	●	●	●		
(Bouneffouf, 2013)				●		●	
(Cena et al., 2013)	●					●	●
(Moscato et al., 2013)		●	●		●	●	●
SMARTMUSEUM (Ruotsalo et al., 2013)	●	●	●	●	●	●	
(Cha, 2014)		●		●	●	●	
(Al-Hassan et al., 2015)					●		●

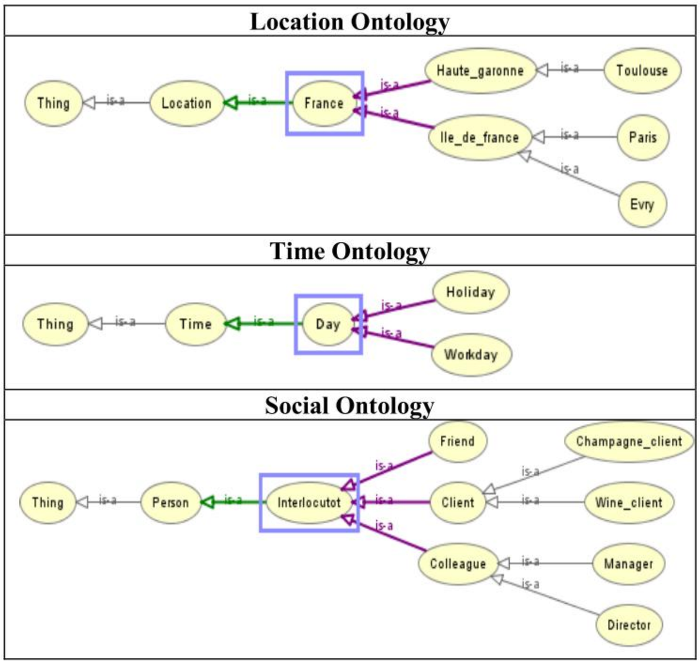


Figure 26. Location, time and social ontology (from (Bouneffouf, 2013))

Since recommendation is not a one-time task, in addition to the initial construction of the user profile, a RS must also assure that accurate recommendations will be made in the future. Therefore, we can find different techniques for *updating* the user profile during a session. The use of implicit methods (72%) is more widespread than the one based on the explicit requirement of feedback (51%). In Table 7 we can also observe that around 27% of the papers use a combination of both approaches. Explicit knowledge elicitation has been used both in profiles based on annotations and those that consider feature vectors. For the former case, (Bhatt et al., 2009) propose an incremental procedure to allow the experts to refine the semantic categorization stored in the system. For the latter one, (Wang and Kong, 2007) use explicit information of the user to update his/her degree of interest on the concepts. The user has to rate the recommended alternatives and then the degree of interest on the related concepts is modified according to the given ratings.

Several papers exploit the implicit information provided by the user by tracking his/her behaviour. For instance, (Sieg et al., 2007) increment or decrement the preference weights based on bookmarking, frequency of visits and time spent on each alternative (a Web page, in this case). (Shoval et al., 2008) updates the importance score of each concept based on the number of its 'clicks' divided by the total number of 'clicks' of the user. (Jiang and Tan, 2009) present a method based on probabilities (Bayesian networks) for learning relations of interest. In (Sendhilkumar and Geetha, 2008) a weight degree is specified for each user action: save (1), print (1), copy (0.25-0.75) and bookmarking (1). Those weights are applied to modify the current user profile according to the actions done on each of the proposed alternatives. Another example that exploits implicit information is (Moscato et al., 2013), where the user’s interests are gathered from social networks (e.g. Facebook) when he/she signs in. Although they also propose a small questionnaire at the registration, the main maintenance of the user's preferences is done by tracking the user's behaviour.

About 48% of the reviewed papers include some domain inference mechanism. Some of them extend the user profile by exploiting the ontology hierarchy (Codina and Ceccaroni, 2010) to discover new knowledge about the user's preferences. For example, if a user expresses an interest in *Culture* (parent class of *Museums*) it can be deduced that he may also be interested on *Museums*. On the other way round, if a user is interested in *Museums*, *HumanHeritage* and *Monuments* it could be inferred that he is interested in *Culture* in general. A derivation method for building a sub-ontology for a certain user is given in (Bhatt et al., 2009). From the partial specification of the user's interests on a base ontology, a complete and independent sub-ontology is generated. The derivation itself is achieved by the application of different processes, like optimisation schemes and consistency checking. More complex approaches like (Lamsfus et al., 2010) or (Cantador, 2008) extend the user interests with spreading activation algorithms that iteratively propagate the weights of user preferences through the ontology relations. This kind of algorithms explore networks by considering the relationships between nodes. They start associating to a set of nodes a weight value or "activation level" and then these weights are iteratively propagated or "spread" to the linked nodes. The strength of the propagation normally decreases as the distance with the initial nodes increases. The process is repeated until there are no more nodes related to the initial ones. (Blanco-Fernández et al., 2011a) present an approach in order to overcome two severe problems suffered by the traditional spreading activation algorithms. The first one is related to the kind of links that are used: some approaches only have simple relationships, which only allow making few inferences and hamper the discovery of new knowledge about complex relationships. The second problem is the propagation of static weights through the network. In order to overcome these drawbacks, they use more complex associations between nodes based on properties, such as 'hasActor', 'hasIntendedAudience', or 'isAbout'. This variety of associations permits to establish different ways to propagate the preference weights, leading to enhanced recommendations. Moreover, each semantic relation considers a different strength degree, which enables to update the weight properly. For instance, they consider the length of the property and the existence of a common ancestor between two nodes, among other data. Similarly (Jiang and Tan, 2009) do not only consider the distance between nodes, but they also use taxonomical and joint relationships. They provide a decay factor over time in the spreading process in order to represent short term preferences rather than long term ones. This allows modeling the confidence on the inferred values. (Al-Hassan et al., 2015) proposes a new function named *Inferential Ontology-based Semantic Similarity* (IOBSS) that measures the semantic similarity between items in a specific domain of interest by taking into account not only the hierarchical relationships but also the shared attributes and implicit relationships through the network of concepts, giving rise to richer similarities.

3.4. A framework for managing uncertain preferences with ontologies

Ontologies define a set of concepts related to a certain domain as well as the relationships among them. This structure may be exploited to represent and reason about the preferences of a user. In recommender systems the user profile is usually built and maintained through explicit information provided by the users (filling forms, rating items) or implicit information related to the interaction of the user with the system (saving items, deleting items). The work presented in this section proposes a general framework that allows representing and reasoning about the uncertainty associated to preferences in ontology-based semantic recommender systems. To do so, the concepts of the ontology represent the uncertainty of the degree of interest of the user. Both the degree and the uncertainty are propagated through the related ontology concepts to manage user preferences. These preferences are represented in each concept with a fuzzy set indicating their degree of interest.

3.4.1. A fuzzy approach to store the user profile in an ontology

In a recommender system the domain ontology permits to classify the objects to be recommended. We consider that each object is an instance of one (or several) of the lowest level classes of the ontology (*i.e*. the leaves). Thanks to the taxonomical structure of the concepts in the ontology, we can reason about the objects at different levels of generality. We propose to use the domain ontology to represent the preferences of the users of the recommender system. The users can be interested on some of the concepts of the ontology with different levels of engagement. We propose to use the *Fuzzy Set* theory to represent the relation between the user and the different concepts of the ontology. A fuzzy set X is defined by a membership function of the objects of the domain Y. The membership degree to the set X (denoted as μ_X) of a certain object is a number that indicates to what extent the object belongs to the concept X. While in Boolean logic the membership is limited to 0 and 1, in Fuzzy logic we can have values in the continuum [0..1], which permits to have a richer gradation of values. For example, each person belongs to the set of “Tall” with a different degree that depends on his/her height (Figure 27).

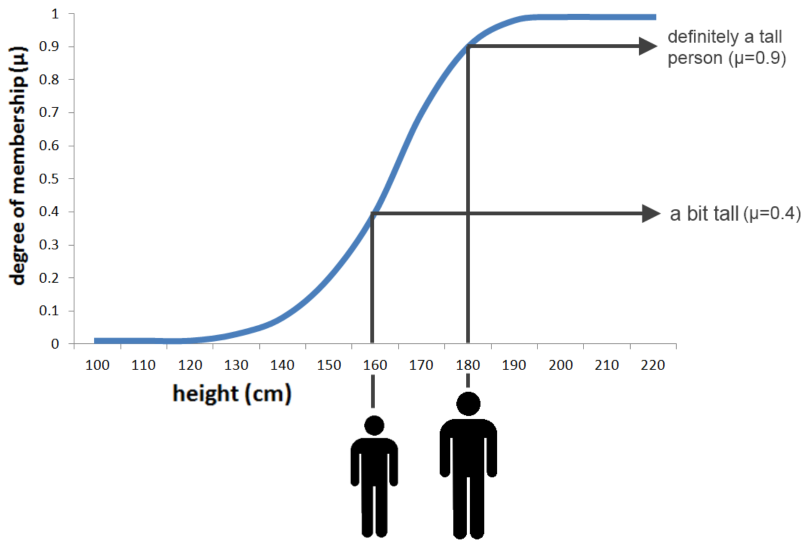


Figure 27. Fuzzy sets associated to the linguistic labels of the variable Height

Fuzzy sets have been used in this thesis because they allow the representation of incomplete or imprecise information. We propose to have a fuzzy set for each concept of the ontology. The elements of the fuzzy sets are the users of the system.

Proposition 1. Let us consider a fuzzy set for each concept c of the ontology, so that, for each user u, $\mu_c(u)$ gives the membership degree of u to the concept c.

This membership degree is personal for each user and represents his/her degree of interest in a certain concept c. If the user is completely interested in c, then $\mu_c(u)=1$. Oppositely, when $\mu_c(u)=0$, we assume that user u is not interested at all in concept c.

When a certain user u needs a recommendation, we propose to find the values of $\mu_c(u)$ for all the concepts in the ontology. Once the ontology has been completely labelled with $\mu_c(u)$, the recommender system will be able to find the most appropriate items for this user, taking into account that each object is an instance of some of the concepts. The values of $\mu_c(u)$ will be calculated using explicit and implicit information elicited from the interaction of the user with the system. Due to this process of estimation, there is a strong uncertainty in the preference values. To manage this uncertainty, we will consider the following confidence degree:

Proposition 2. Let us consider a confidence level $CL_c(u)$ between 0 and 1 that quantifies the confidence associated to the estimation of the membership degree of u to the concept c, denoted as $\mu_c(u)$.

A large value of $CL_c(u)$ indicates that we can trust the value of $\mu_c(u)$ as the true degree of interest of the user u for the concept c, whereas a low value indicates that the estimation is not so reliable. In this way, not only the degree of membership to the concepts in the ontology is considered to select the best alternatives, but also the confidence on the estimation of those values is taken into account. For instance, the recommender system may decide to ignore the values with a low confidence level, because they have not achieved enough support.

In summary, the user's personal profile consists on a copy of the ontology that stores the degree of interest of this user on each concept, as well as the related confidence levels. As an example, let us consider a recommender system for the members of a Hiking association. Figure 28 shows a small portion of the domain ontology, which can be used to recommend events, news or conferences of interest to the association members. As said before, it is assumed that all the recommendable items are instances of the lowest level concepts (*OilRoutes*, *WineRoutes*, *DrivingRoutes*, *Trekking*, etc.). The instances do not belong to the profile; they are stored in a separate database.

3.4.1.1. Initialization of the profile

Each concept within the ontology maintains an interest degree $\mu_c(u)$ estimated by the system, which is calculated from the collection of user information through the session. The collected data can be extracted explicitly or implicitly from the user. For the initialization of the user interests the application asks him/her to fill in a form where the user can express the interest on a certain number of general domain aspects, represented by first-level ontology concepts (in the example shown in Figure 28, those general concepts are *Routes* and *Sports*). Rating values range from 0.0 (no interest) to 1.0 (highest interest). The confidence level associated to these ratings is 1.0 because the value is fully reliable since it is given directly by the user.

3.4.1.2. Propagation of the initial preference and certainty values

The hierarchical structure of the ontology may be exploited to transfer the preference information through the nodes. In particular, a *downwards propagation* of the initial preference and confidence values obtained for the first-level ontology concepts is performed.

Figure 28 shows an example of initialization of values given that the user explicitly expresses a high interest in the first-level concept *Routes* ($\mu_{Routes}(u)=0.8$, $CL_{Routes}(u)=1.0$) and a low interest in *Sports* ($\mu_{Sports}(u)=0.3$, $CL_{Sports}(u)=1.0$). This suggests that the user is interested in general in different kinds of routes, which are represented by its descendants, except for those routes that are related to sports. Therefore, the system has to transfer the interest shown in the most general concept to its subclasses until the concepts in the lowest level (that are used to instantiate the items to be recommended) are reached.

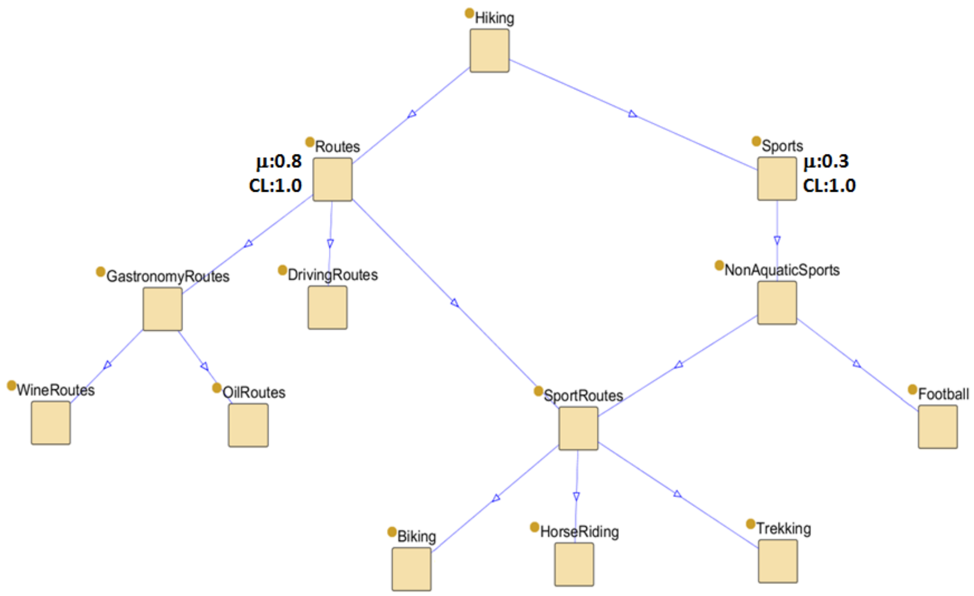


Figure 28. Initialization of the Hiking ontology with µ and CL values

We know that the user is highly interested in *Routes* in general but, in fact, there is some level of uncertainty that the interest is equal in all its children. For example, a young user can be more interested in *DrivingRoutes* than in *GastronomyRoutes*. Therefore, the level of uncertainty on the membership value of the children must be increased as we move to deeper levels of the ontology. The further we are from the first-level concepts, the more uncertain the preferences are. We propose to copy the membership degree of the user to the parent class to all its descendants, but decreasing the degree of confidence at each level by a factor α, which can be customized to the needs of the application. For instance, taking α=0.33, a value of confidence of 0.34 would be given to the preference in the *WineRoutes* concept, because it is two levels away from the *Routes* concept. *CL* values of the other concepts are shown in Figure 29.

Definition 1 (*Downwards propagation of the initial preferences*)

The preference associated to a concept c is calculated as an average of the preferences of his parents (χ^c), weighted by their confidence values. The confidence value associated to c is the average of the confidences in his parents, decremented by α:

$$\mu_c(u) = \frac{\sum_{i \in \chi c} \mu_i(u) CL_i(u)}{\sum_{i \in \chi c} CL_i(u)} \qquad CL_c(u) = \frac{\sum_{i \in \chi c} CL_i(u)}{|\chi c|} - \alpha \tag{1}$$

At the same example, the preference μ of the descendant nodes of the *Routes* concept maintain the same score except for those concepts that are also descendants of *Sports*, as it is the case of *SportRoutes* and its descendants. In those concepts, the average of both ascendants is computed giving a value of μ=0.599. The score has been decreased due to the low preference value given to the *Sports* concept.

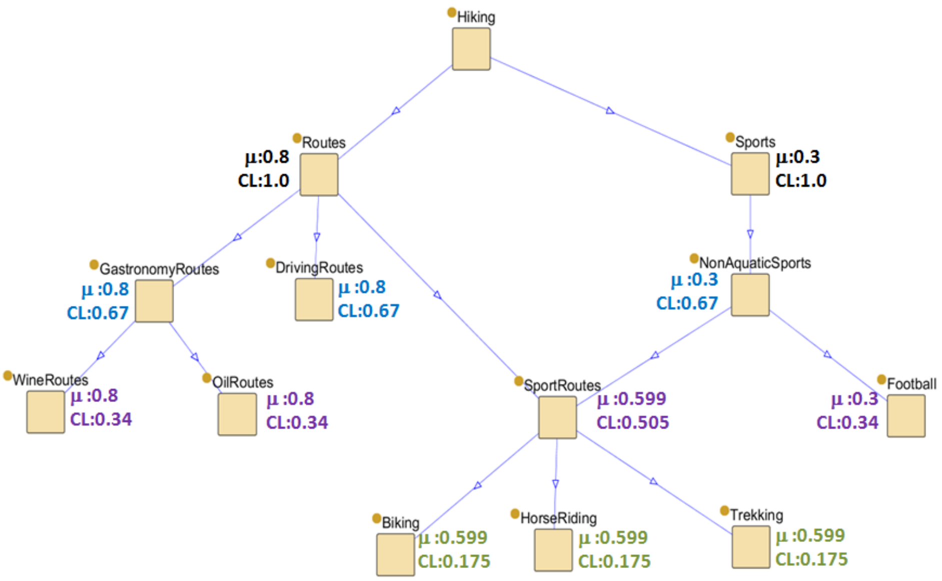


Figure 29. Downwards propagations of the ontology preferences with α=0.33

3.4.2. Dynamic refinement of the user profile

During the execution of the recommender system we can gather additional knowledge about the user's interests. The evidences provided by the different types of actions on the objects are used to modify both the membership degrees of the user to the related concepts and their confidence level. The information obtained about an object *i* affects directly the concepts which *i* is instantiating (which are leaves in the ontology).

We distinguish two main types of information that can be obtained from the interaction of the user with the recommender system:

A) Since each object is labelled with concepts at the lowest level of the ontology, we can learn about the interest of the user on these concepts by studying the actions he/she does on them, which can be either *positive* (*e.g.* saving a recommended item) or *negative* (*e.g.* removing a saved item). For this type of indirect feedback, the confidence level should be low.

B) Recommender systems may ask the user to rate some items shown to him/her. In this case, the rating values on the items can also be used to estimate the membership degree of the user to the lowest level concepts. The confidence level can be high because this is explicit information provided by the user.

Table 8 summarizes the scores *s* (between -1 and 1) and the weights *w* (between 0 and 1) associated to each user action. This feedback is useful to refine the estimation of the membership degree of the user by inferring his/her interests based on the behaviour of the user in front of the previously recommended objects.

Table 8. User actions allowed by the system.

User actions	Explicit	Implicit	s	w
Save recommended item		●	0.5	0.5
Remove recommended item		●	-0.5	0.5
Request detailed information about an item		●	0.1	0.2
Request item similar to the current one		●	0.15	0.3
Rate an item	●		[-1.0, 1.0]	1.0

Assume that we have observed a set of actions A_c on a group of objects that are instances of the concept *c*. The scores and weights associated to these actions are aggregated together as follows:

$$\Delta_c = \frac{\sum_{a \in A_c} s_a w_a}{\sum_{a \in A_c} w_a} \qquad CA_c = \frac{MIN(\lambda, \sum_{a \in A_c} w_a)}{\lambda} \tag{2}$$

As can be seen in equation (2), the aggregated confidence of the actions is normalized using a parameter λ, which can be set to a level above which a higher amount of evidence is not required. If the aggregated confidence in the actions is higher than the current confidence level of the concept ($CA_c \geq CL_c$), then its preference and confidence values are updated as follows:

$$\mu_c = \begin{cases} if\ (\Delta_c > 0) & MIN(1, \mu_c + \beta \times \Delta_c) \\ else & MAX(0, \mu_c + \beta \times \Delta_c) \end{cases} \qquad CL_c = \beta \times CA_c + (1-\beta) \times CL_c \tag{3}$$

β is a parameter between 0 and 1 that graduates the level of change between the current values and the scores and weights given by the user actions. The higher its value, the bigger is the impact of the actions on the change on the preference associated to the concept.

Figure 30 illustrates several actions a user has done at the previous example: a) requests more information (*s*=0.1 and *w*=0.2) and saves (*s*=0.5 and *w*=0.5) a *WineRoutes* item; b) requests more information (*s*=0.1 and *w*=0.2), saves (*s*=0.5 and *w*=0.5) and rates highly (*s*=1.0 and *w*=1.0) a *HorseRiding* item; and c) rates with low values two *Football* items (*s*=-0.8 and *w*=1.0 for the first and *s*=-0.6 and *w*=1.0 for the second). These action values are applied to the related concepts with the equation (3). For instance both the preference and the confidence on the *WineRoutes* and *HorseRiding* concepts have increased due to the positive actions done on the related items. On the other hand, the negative actions applied to the instances of *Football* have led to a decrease of the preference to the minimum (0), with a high confidence.

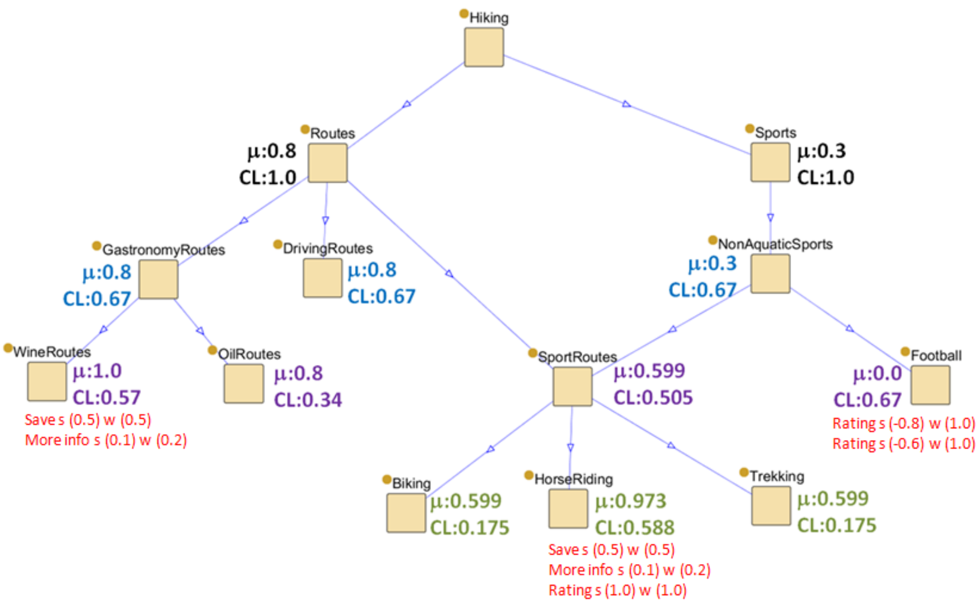


Figure 30. Transmitted user actions to leaf nodes of the ontology

3.4.2.1. Upwards propagation

At this point, the feedback of the user has been used to modify the information stored at the lowest-level concepts of the ontology. After the system has collected a sufficiently large set of user actions, the values can be propagated through the ontology to update the values of other related concepts. In a first step, we make an *upwards propagation* to the ancestor concepts of the modified leaves. For instance, if the user has rated positively some specific instances of *WineRoutes*, the system can transmit a positive impact towards its ancestors *GastronomyRoutes* and *Routes*. Again, the more distant an ancestor is, the more uncertainty we have.

Note that several children of the same concept may have been modified (e.g., the user may have interacted with instances of *WineRoutes* and *OilRoutes*, both children of *GastronomyRoutes*). Let us assume that φ^c is the set of concepts that are children of c and have confidence values higher than a certain threshold (concepts that don't have enough confidence should not influence on their parents). The aggregated preference and confidence values of the children of c may be computed as follows:

$$\Delta_c = \frac{\sum_{i \in \varphi^c} \mu_i CL_i}{\sum_{i \in \varphi^c} CL_i} \qquad\qquad CA_c = \frac{\sum_{i \in \varphi^c} CL_i}{|\varphi^c|} \tag{4}$$

If the aggregated confidence of the children of c, CA_c is higher than a threshold, then its preference and confidence values are updated as shown in equation (5). β is the parameter used in equation (3), which regulates the degree of change.

$$\mu_c = \frac{(1-\beta)\times\mu_c CL_c + \beta\times\Delta_c CA_c}{(1-\beta)\times CL_c + \beta\times CA_c} \qquad CL_c = \beta\times CA_c + (1-\beta)\times CL_c \tag{5}$$

The upwards propagation is illustrated in Figure 31 for the previous example indicating with red arrows the values that are propagated upwards. In this case, only those concepts with a *CL* higher than 0.35 are updated (which is the case of the three leaves that have been modified). Positive actions on the instances of a class lead to an increase of the preferences of its ascendant concepts, as can be seen for *GastronomyRoutes* and *Routes*. On the other hand, *NonAquaticSports* increases slightly its preference value, due to the mixed influence of two concepts that have both positive (*SportRoutes*) and negative (*Football*) scores.

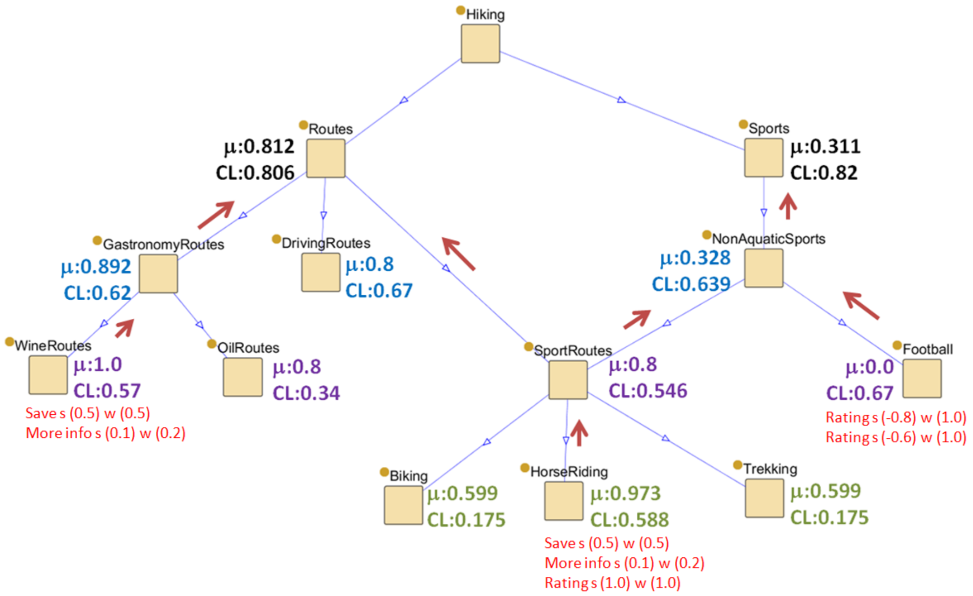


Figure 31. Upwards propagation of the ontology preferences with β=0.5, λ=1.5 and minimum upwards confidence 0.35

3.4.2.2. Downwards propagation

Once the upwards propagation has been completed, a second step propagates the preference and confidence values to the descendants of the updated nodes. For instance, if the preference of the user in *SportRoutes* has been modified due to the rating of some *HorseRiding* activities, a modification of the values for *Biking* and *Trekking* seems reasonable, due to their high semantic similarity with *HorseRiding*.

In this downwards propagation, the information of a concept *c* is modified according to the preference and confidence values of its parents, χ^c, as long as these confidence values exceed a given threshold. The aggregation of the information of the parents is done equivalently to the upwards case, as follows:

$$\Delta_c = \frac{\sum_{i \in \chi c} \mu_i CL_i}{\sum_{i \in \chi c} CL_i} \qquad\qquad CA_c = \frac{\sum_{i \in \chi c} CL_i}{|\chi c|} \tag{6}$$

If CA_c is higher than a given threshold and c has not been updated during the upwards propagation, its information is changed according to equation (5).

Figure 32 shows in red arrows those values that are propagated downwards. The concepts with a crossed square are the ones that were already modified in the upwards propagation (and hence they are not modified again) as well as those concepts in which CA_c is equal or lower than 0.35. As we can see, the updated concepts in the downwards propagation are *OilRoutes*, *DrivingRoutes*, *Biking* and *Trekking*. As expected, all these concepts increase their preference and confidence values due to the positive actions performed by the user on closely related concepts (*WineRoutes* and *HorseRiding*).

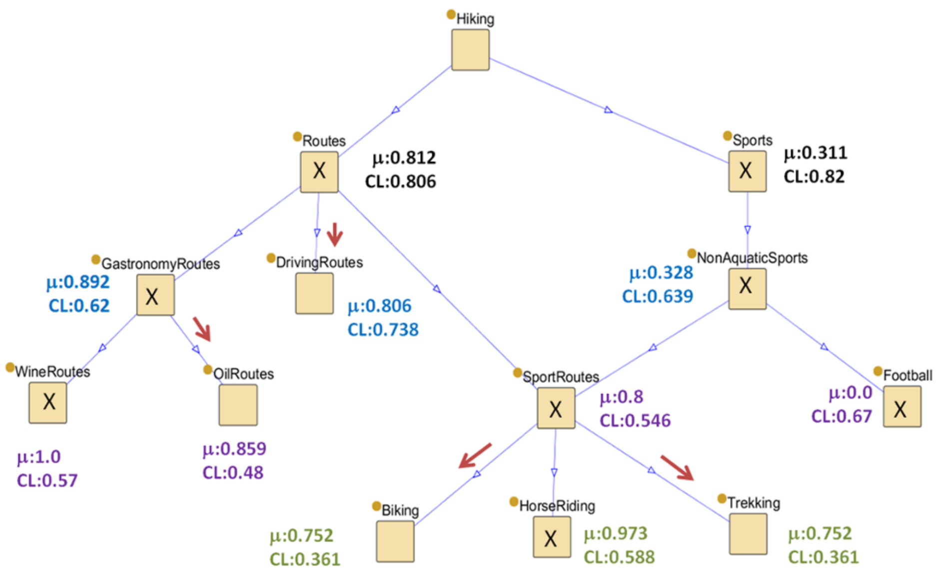


Figure 32. Downwards propagation of the preference and confidence values with β=0.5, λ=1.5 and minimum downwards confidence 0.35

3.5. Conclusions

The management of semantic domain knowledge by recommender systems is an exciting current line of research. The general idea is that the use of a domain ontology (both to represent user profiles and domain items) may lead to a complex semantic analysis of how similar are two users or how close the preferences of a user are to the characteristics of an item.

In this chapter we have provided a brief survey of some of the more recent works on ontology-based (or semantic) recommender systems. After that, we have proposed a novel framework, founded on fuzzy systems, that suggests to store, in the user profile, the preference value of the user for each concept and a confidence degree on this value. We have also described how these values may be easily initialised (e.g. from a simple description of the high-level interests of the user on the most general classes of the ontology) and how they can be dynamically modified by analysing different kinds of interaction of the user with the recommended items. As preferences change dynamically, this framework could be used to model situations in which the actual preferences of the user change over time.

This framework for managing uncertain preferences has been successfully applied in a tourism recommender system, to be described in chapter 5 of this dissertation. However, it is general enough to be usable in different applications, because the system actions (and their scores and weights) and the parameters for preference adaptation can be customized. Our future work on this topic includes a thorough analysis of the influence of the different updating parameters in the dynamic change of the user preferences, the study of different ways in which the information about preferences and certainties may be used by the recommender systems, and the test of this general framework in other domains.

Chapter 4 – Diversification of recommendations through semantic clustering

Precision and *recall* are the metrics more commonly used to measure the accuracy of the suggestions made by recommender systems. The former indicates the percentage of recommended items which are relevant for the user, whereas the later is the proportion of user-relevant items that have actually been recommended. These measures are indeed important to quantify the degree to which the recommended items match the user's interests. However, it may be argued (Mcnee et al., 2006) that other factors also have a strong influence on the overall satisfaction of a user with a RS, being the *diversity* of the recommended items one of them (Ziegler et al., 2005). The intuitive idea is that the recommendation of a set of very similar items may technically be very accurate, since all the items may match quite precisely the user's preferences, but at the same time it may also be counterproductive and unsatisfactory for the user. The recommendation of almost identical items (e.g. books of the same genre by a single author) is boring, unengaging and devoid of *serendipity* (the quality of presenting options that surprise the user and permit him/her to discover new items that may also be interesting, like books of the same genre by other authors, or books by a known author that explore other genres).

The main idea of *topic diversification* is to study how a RS can *balance* the provision of accurate recommendations with the suggestion of items that are different enough to attract the attention of the user and improve his/her experience with the system. The equilibrium between accuracy and diversity is not easy to achieve, as the increase in one of them often leads to the decrease of the other one. If the system does not use diversification mechanisms, the recommended items may be too similar and the system may not be very helpful neither for the user nor for the retailer (that aims to sell all the variety of products, not only those that are most popular and well-known by the majority of users). However, suggesting many items that do not match precisely the user's preferences may also decrease the confidence on the RS and lead to its rejection. Some works actually suggest using two different lists, one with the standard recommendations and another one with related but unexpected items (Ge et al., 2010).

This chapter focuses on the study *of diversification mechanisms*, understood as algorithms that select a small set of items to recommend to the user from a possibly large set of items that have been previously filtered and ordered by the RS according to the user profile. In this chapter the main techniques that have been suggested to diversify a set of recommendations are shown, and some variations and a new method based on clustering are proposed. These novel methods have a low time complexity and provide a good level of diversity with an insignificant loss of accuracy. All the diversification techniques commented in this work have been experimentally tested using SigTur/E-Destination, the recommender system for tourism activities detailed in the next chapter.

The rest of the chapter is organized as follows. In the next section we briefly review previous works on the diversification of recommendations. Section 4.2 explains a new semantic measure of similarity between objects, which is later used to measure the diversity of a set of recommended items. Section 4.3 presents a list of diversification methods that includes some variations of previous techniques and a new one based on clustering. The balance between accuracy and diversity offered by all these methods has been experimentally tested. The results of these tests, including the computational costs of the different algorithms, are detailed on Section 4.4. The final section makes some final conclusions and presents potential lines of future work.

4.1. Related works

The techniques that have been proposed in the literature to present a varied list of recommendations may be divided into three main categories. The first group, which is the main focus of this approach, consists on the application of a diversification algorithm on the list of results calculated by a standard RS (which have already been selected according to their similarity with the user's preferences). These algorithms basically change the order of the items in the set of recommendations, ensuring that the first items on the list (the ones that will be finally shown to the user) are both diverse and accurate. The second group integrates the analysis of diversity within the actual ranking procedure of the RS, so that both accuracy and diversity are taken into account at the same time. Finally, the last group includes those techniques that do not focus on individual diversity but on *aggregate* diversity (the level of diversification of suggestions of the RS throughout all users). These methods try to make sure that all the items (even those that are new or unpopular) are actually recommended to some users. Some examples of these three categories are commented in the following paragraphs.

One of the first approaches that studied the diversification of a list of recommended items was (Smyth and McClave, 2001). In this work the RS starts by building a ranked list L of recommendable items, taking into account the user's preferences. The first item of this list is added to the final list T of items to be recommended. Then, the system analyzes all the items in L and looks for the item that has more *quality*, which is measured by multiplying the similarity of the item to the user's preferences by the diversity of the item with respect to all the items already stored in T. The item with more quality is added to T. This process is repeated until T contains the number of items that the system intends to recommend to the user (typically the size of T is small –8 or 10 elements- whereas L may have hundreds of items). This algorithm is computationally expensive, since the diversity of each element of L with respect to the set of items already added to T must be checked in each iteration; that's why the authors also propose a *bounded* version of the algorithm, in which only the first B items of L are analysed in each iteration. Another work (Ziegler et al., 2005) added a parameter to this algorithm that permits to adjust the desired level of diversity. In this way the designer of the RS may decide to have more accuracy or more diversity in the offered recommendations, depending on the specific domain of application. In this work each item is represented with a set of attributes, and the values that these attributes can take are structured in a taxonomy. This fact allows the computation of the semantic similarity between pair of items. Another approach in which the level of diversity may be adjusted is reported in (Aytekin and Karakaya, 2014). In this work the domain items in L are clustered, taking into account the ratings given by the users. They only consider one element of each cluster in each iteration of the selection procedure; therefore, the computational cost is much lower than the one of the previous methods. Their results show good levels of diversification with a small decrease in accuracy. Another approach of the same family is presented in (Zhang and Hurley, 2008), in which an optimisation method that maximizes the diversity of the recommendation set while keeping an adequate level of accuracy is proposed. The optimisation problem is solved by reducing it to a trust-region problem.

All the works mentioned on the previous paragraph focus on increasing diversity by selecting carefully a set of items from a ranked list of options, previously computed by the RS in some way (usually with a content-based or a collaborative filtering procedure). Other approaches integrate the diversification mechanisms within the actual ranking procedure of the RS. For instance, Vargas (Vargas et al., 2011) is inspired by diversification techniques used in Information Retrieval, in which results associated to different meanings of the query are shown to the user. His idea is that a set of diverse recommendations may be obtained by showing to the user the results suggested by different recommendation mechanisms. In (Candillier et al., 2011) it is stressed that the selection of an appropriate recommendation technique for a particular user in a specific context is crucial to provide satisfactory results, as the same user may be interested in precise or diverse recommendations in different settings. The same authors propose in another work two similarity measures, *topicality* and *topical diversity*, that may be used to assess the degree of variety of a set of results (Candillier et al., 2012). They conclude that the aggregation of these similarities offers results with a good trade-off between accuracy and diversity. Zhou introduces a recommendation algorithm called *heat-spreading*, inspired on the physical process of heat diffusion (Zhou et al., 2010). The idea is to propagate the values of the history of objects evaluated by a user to its neighbourhood. A combination of this method with a classical one focused on accuracy gives results that, in some cases, produce gains both in accuracy and in diversity. Another proposal (Akiyama et al., 2010) considered the degrees of serendipity and unexpectedness of each item within the recommendation process. The former represents the dissimilarity of the item with respect to the user profile, whereas the later measures the uncommonness of the attribute values of the item within the whole item set. Some authors (Iaquinta et al., 2010) have pointed out that it is more probable to offer serendipitous results when the RS does not have a large confidence on the information about the user preferences.

The last type of techniques tries to offer *aggregate* diversity, not *individual* diversity. Thus, the aim is to provide a diverse set of recommendations globally, taking into account all the users of the system. These systems are mainly based on collaborative filtering. For instance, Niemann and Wolpers (Niemann and Wolpers, 2013) define a notion of similarity between items that takes into account not only their direct co-ocurrence in the purchasing list of users, but also their *second-order* co-ocurrence (two items are similar if each of them appears frequently with a third common item). Therefore, this method finds new links between items that were never bought together. The rating predictions for unfrequent items are increased, hence improving the aggregate diversity. The work reported in (Adomavicius and Kwon, 2010) proposes different ways of increasing the weight of the items that have been less frequently rated, in order to try to improve their chance of being recommended and increase the aggregate diversity of the RS. One of them is to rank in an ascending order the items based on their number of ratings, from the *lowest* to the highest, so that the most unusual items appear on the top positions. A minimum rating value is set to avoid recommending bad items. Their best results range from a diversity gain of up to 20-25% with only a 0.1% accuracy loss, up to a 60-80% diversity gain with a 1% accuracy loss. Another example of aggregate diversity is proposed in (Gan and Jiang, 2013). The main idea is to adjust the similarities between users with a power function to reduce the adverse effects of popular items in user-based collaborative filters. With this method the influence of the most similar users is enhanced, and an increase in both accuracy and diversity is reported.

In this thesis we want to study the influence of several diversification mechanisms on the results of the personalised recommender of tourism activities detailed in chapter 5. Thus, the rest of the chapter will focus only on the analysis of methods of the first family, which select the items to be shown to the user from the ranked set of options calculated by the RS. Aggregate diversity will not be considered, since the aim is to show to each individual user of the recommender system a varied set of alternatives (keeping a good level of accuracy). However, in the Tourism domain it is also important to make sure that all the activities available on a given area are recommended to some customers (even those that are not very popular), so we intend to include a more detailed study of aggregate diversity and serendipity in our future work.

In section 4.3 we describe the basic diversification mechanisms proposed in the literature, some variations and a new one based on semantic clustering. Before that, in the next section we describe the semantic similarity measure that will be used to assess the degree of diversity of the items in a list.

4.2. Diversity measure

In order to implement a diversification algorithm it is necessary to know how similar (or, actually, dissimilar) two objects are. The use of domain knowledge, in the form of an *ontology*, permits to define *semantic* similarity measures. An ontology, as explained in the previous chapter, is a knowledge structure that represents, in an explicit and formal way, the manner in which a certain domain of interest may be conceptualised. Its main components are *concepts* (classes of objects that share a common property), taxonomic and non-taxonomic *relationships* between them, and *instances* (specific objects of the domain). For instance, Figure 33 shows a small portion of an ontology of Tourism activities. The concepts shown in the figure are taxonomically related (e.g. *WineRoutes* is a subclass of *GastronomyRoutes*, which is in turn a subset of *Routes*). Each instance (in this case, each particular touristic activity) will be associated to a set of classes; for example, a concrete enological route on a horse could be related to the classes *WineRoutes* and *HorseRiding*. Intuitively, the shorter is the taxonomical distance between two concepts in the ontology, the more similar they are. Following the same example, a touristic route themed on oil (tagged as an *OilRoute*) should be more similar to an enological route (classified as *WineRoute*) than to a tour taken on bycicle (labelled with the *Biking* tag).

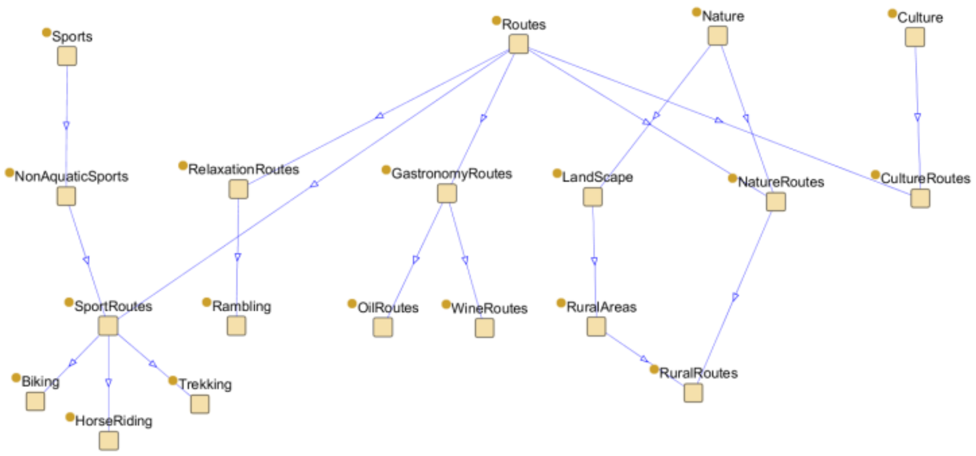


Figure 33. Portion of a Tourism ontology

This intuitive notion of semantic similarity may be implemented in different ways. One possibility is to count the number of links between two items (e.g. 2 from *OilRoutes* to *WineRoutes*, but 4 from *OilRoutes* to

Biking). Another possibility, which is the one that will be used in this work, is to consider the number of shared ancestors between two items (e.g. *OilRoutes* and *WineRoutes* have 2 common ancestors, whereas *OilRoutes* and *Biking* only have 1 common ancestor). The *ontology-based semantic distance* (OSD) between two concepts t_i and t_j (equation (7)) is measured as the square root of the ratio between the number of different ancestors and the total number of ancestors of both concepts (Moreno et al., 2013b). This distance ranges from 0 (the distance between a concept and itself) to 1 (the distance between two concepts that do not have any common ancestor). In this equation *A(t)* is the set that contains the concept *t* plus all its ancestors (super-classes).

$$OSD(t_i, t_j) = \sqrt{\frac{|A(t_i) \cup A(t_j)| - |A(t_i) \cap A(t_j)|}{|A(t_i) \cup A(t_j)|}} \quad (7)$$

The *ontology-based semantic similarity* (*OSS*) between two concepts is defined as the inverse of the *OSD* (*1-OSD*). Table 9 shows this similarity between the concepts *OilRoutes*, *WineRoutes* and *Biking*. The less common ancestors between two concepts, the larger is the distance between them (and the lower is their similarity).

Table 9. *Ontology-based semantic similarity* between concepts

	OilRoutes	*WineRoutes*	*Biking*
OilRoutes	1	0.7	0.45
WineRoutes	0.7	1	0.45
Biking	0.45	0.45	1

We want to consider the case in which each recommendable item may be associated not only to a single class of the ontology but to a *list* of classes. Thus, we need to define a similarity measure between lists of concepts. Given two lists, the idea will be to measure their resemblance by somehow aggregating the pairwise similarity between the items in both lists. For instance, a simple option could be to take the average similarity between the pairs of concepts. However, this option may return the same aggregated result on very different lists (e.g. the result 0.5 would be obtained with the lists of similarities (0,0,0,1,1,1) and (0.5,0.5,0.5,0.5,0.5,0.5)). In this dissertation we propose to use the *Ordered Weighted Aggregation* (OWA) family of operators (Yager, 1988) to aggregate the pairwise similarities between the members of two lists. An OWA aggregator is defined with a mapping $R^n \rightarrow R$ that has an associated weighting vector *W* (see section 5.3.2.1 for more details on how to construct this vector) of dimension *n* with $\sum_{j=1}^{n} w_j = 1$ and $w_j \in [0,1]$, so that

$$OWA(t_1, ..., t_n) = \sum_{i=1}^{n} w_i t_i \quad (8)$$

Thus, the similarity between item *a* (associated to a set of concepts a_i) and item *b* (associated to a set of concepts b_j) may be calculated as follows:

$$sim(a,b) = OWA(\{\forall a_i : \max_{\forall b_j} OSS(a_i, b_j)\} \cup \{\forall b_j : \max_{\forall a_i} OSS(b_j, a_i)\}) \quad (9)$$

Thus, first we calculate, for each concept associated to item *a*, which is the most similar concept in *b*, and this maximum similarity is stored in a list. After that, we repeat the process for all the concepts related to *b*, and the maximum similarities to concepts in *a* are added to the same list. Finally, all these values are aggregated, using the OWA operator, into a single final similarity value. The weighting vector regulates the desired degree of andness/orness to be used in the aggregation.

In the diversification algorithms used in the next section it will also be necessary to compute the similarity of an item *a* with respect to a list *l* of items (to decide whether the new item is different enough from all the items in the list to be added to it). In this case, we will also apply an OWA operator to aggregate the similarities between the item and each of the members of the list:

$$sim\_list(a,l) = OWA(\forall l_n : sim(a, l_n)) \quad (10)$$

In this expression l_n are the items of the list *l* and *sim* is the formula used in equation (9).

4.3. Diversity methods

As shown in section 4.1, there are several methods that try to improve the diversity of the results offered by a RS. We will focus on those methods that, given a long ranked list of alternatives (already ordered according to their relatedness with the user's preferences), decide which (small) set of items will be finally shown to the user. During this selection the system should tend to choose those items that are at the top of the initial list (which are the most accurate), but it should make sure that the selected items are different enough to show a varied set of recommendations.

This section presents the following methods, which in the next section will be evaluated in a Tourism recommender system and discussed in terms of diversity, accuracy and computational cost:

- Baseline-1 [*None*]: just select the top elements of the list, without evaluating their diversity.
- Baseline-2 [*Random*]: select randomly some elements of the list, without evaluating their diversity.

- *Quadratic*: select iteratively the element of the list with the best balance between accuracy and variety with respect to the already chosen items (Smyth and McClave, 2001).

- *Linear*: variation of the previous method in which a single analysis of the list is made, selecting those items that are different enough from the previously chosen ones.

- *Quadratic break*: variation of the previous method, in which the analysis of the list restarts from the first element each time that an item is selected.

- *Bounded quadratic*: variation of the previous *quadratic* method, in which only the initial elements of the list are taken into account in the selection process.

- New methods based on clustering (*clustering random* and *clustering quadratic*): variations of the *random* and *quadratic* methods in which the elements of the list are clustered (according to their semantic relatedness) before starting the selection process.

The following subsections describe each of these methods, giving an intuitive explanation and the high-level pseudo-code.

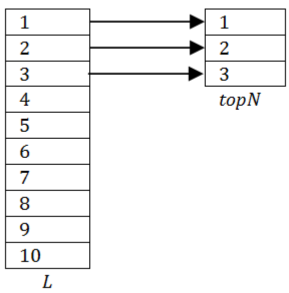
4.3.1. None

This method merely recommends the top N items of the ranked list of alternatives, without evaluating their diversity (see Algorithm 1). Thus, it will serve as a first baseline, as its results will have the maximum accuracy but the minimum diversity.

Algorithm 1. None
Input: L: list of items ranked by accuracy, N: number of items to recommend

Output: $topN$: list of N items to recommend

$n = 0$
while $n < N$ **do**
  $topN[n] = L[n]$
  $n = n + 1$
end while



4.3.2. Random

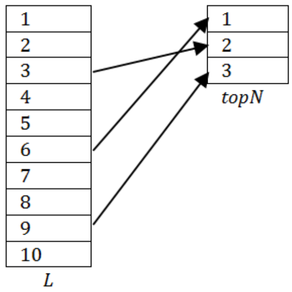
As shown in Algorithm 2, this method just selects randomly N items from the initial list, without taking into account their diversity. Thus, both the accuracy and the diversity of the results are unpredictable. This method will be considered as a baseline with respect to which the other diversification methods may be compared.

Algorithm 2. Random

Input: L: list of items ranked by accuracy, N: number of items to recommend

Output: *topN*: list of N items to recommend

```
1:  n = 0
2:  while n < N do
3:    topN[n] = pop random item from L
4:    n = n + 1
5:  end while
```



As shown in line 3, the method randomly pops an item from list L to be added to the *topN* list. It is not possible to choose the same item twice, since in the pop action the item is deleted from L.

4.3.3. Quadratic

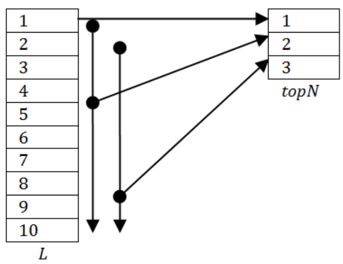
This method (see Algorithm 3) tries to find the elements that offer a best balance between accuracy and diversity (Ziegler et al., 2005; Smyth and McClave, 2001). In each iteration it loops the whole initial list to find the item that has the maximum combination of accuracy (i.e. the maximum score with respect to the user profile) and diversity with respect to the current *topN* list of selected items. A parameter λ, which ranges between 0 and 1, permits to adjust the desired level of diversity. If it is equal to 0, only accuracy will be considered (i.e. the first N elements of the initial list would be selected, as in the *None* method). If it is equal to 1, it would choose in each iteration the element that is more different from the already chosen ones, regardless of its position in the ranked list.

Algorithm 3. Quadratic

Input: L: list of items ranked by accuracy, N: number of items to recommend, λ: level of diversity

Output: *topN*: list of N items to recommend

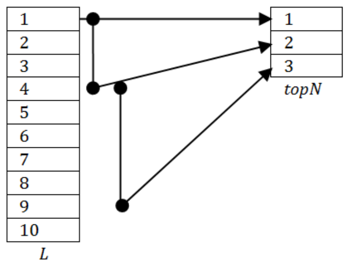
```
topN[0] = pop first item from L
n = 1
max = 0
while n < N do
    for each item i in L do
        d=1 – sim_list(i, topN)
        q = ( λ * d ) + ( (1 - λ) * weight of i )
        if (q > max) then
            max = q
            best_item = i
        end if
    end for
    topN[n] = pop best_item from L
    n = n + 1
end while
```



In line 1 the first item of the ranked list is moved to the *topN* list. This item is the one that has the maximum accuracy. Then the algorithm makes *N*-1 iterations of a loop. In each iteration the element of *L* that offers a best tradeoff between accuracy and diversity is moved to *topN*. In line 6 the algorithm computes the semantic distance (the inverse of the similarity measure shown in (10)) between each item of *L* and the whole set of elements already included in *topN*. Then, in line 7 this distance is combined with the weight of the item (i.e. the normalised score given to the item by the RS, which measures how well it fits with the user's preferences) to determine its overall score (which depends on the desired level of diversity). After having analysed all the items in *L*, the best one is added to topN (line 13) and the method proceeds to the next iteration.

4.3.4. Linear

This method tries to reduce the computational cost of *Quadratic*, which scans the whole list *L* in each iteration of the selection process. The idea is to make a single scan of the list. When an element that is different enough from those that have already been selected is found, it is added to *topN* and the system continues the analysis of *L* from that point (it does not start again from the beginning, as in the previous method). This behaviour is shown in Algorithm 4.



Algorithm 4. Linear

Input: *L*: list of items ranked by accuracy, *N*: number of items to recommend, λ: level of diversity

Output: *topN*: list of N items to recommend

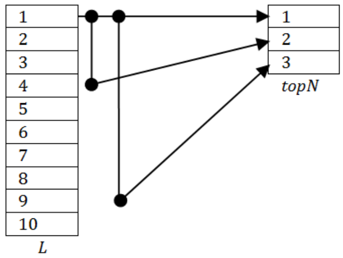
```
1:  topN[0] = pop first item from L
2:  n = 1
3:  while n < N do
4:      max_distance = 0
5:      for each item i in L do
6:          d=1 – sim_list(i, topN)
7:          if d > λ then
8:              topN[n] = pop item i from L
9:              n = n + 1
10:             if (n = N) then
11:                 break for
12:             end if
13:         else if d > max_distance then
14:             max_distance = d
15:             max_item = i
16:         end if
17:         if i is the last item of L then
18:             topN[n] = pop item max_item from L
19:             n = n + 1
20:         end if
21:     end for
22: end while
```

If an element of *L* is distinct enough from the elements already stored in *topN* (condition in line 7), it is immediately added to this list of results (line 8), and the analysis of *L* continues from that point. Notice that in this algorithm the diversity parameter λ is used as a minimum threshold for the distance that an item in *L* needs to have with respect to the items in *topN* in order to be selected. The lower is the desired diversity, the easier it will be for an element of *L* to be selected. The weight of the selected items is not directly taken into account at any moment.

In rare cases, if a very high diversity is required, it might be the case that, after completing a full analysis of *L*, the *topN* list does not contain yet *N* items. If the end of the list *L* is reached, the algorithm adds to *topN* the item that had the maximum diversity with respect to the list of results (line 18) and, if *topN* still does not contain *N* elements, it starts again to analyse *L* from the beginning. This extreme case will not be considered in the posterior study of the computational cost of this algorithm.

4.3.5. Quadratic Break

The *Linear* method certainly has a much lower computational cost than the *Quadratic* one, since it only makes a single scan of *L*. However, there are cases in which it may present counter-intuitive results. Consider the following example. After adding the first item of *L* to *topN* (line 1 in Algorithm 4), it may be the case that the first element that is different enough from this item is in the 10th position of *L*. After adding this item to *topN*, the algorithm looks (from the 11th position) which is the next item that is different enough from the two items already in *topN*. This item, which is the next one that should be added to *topN*, could be for instance in position 15. However, note that it might be the case that an item in a best position, for instance in position 5, has the same distance to the two items in *topN*. The reason is that, when item 5 was analyzed, it was only compared with the first item in *topN*, because the second item had not been added yet. This example shows that we may select items that have the same (or even worse!) diversity than other items that have a higher accuracy. In order to correct this behaviour, the *Quadratic Break* method goes back to the beginning of *L* every time that it finds an item dissimilar enough from the ones in *topN* (line 10 of Algorithm 5). Thus, the computational cost will be higher than the one of the *Linear* method, although it will not be as computationally expensive as the *Quadratic* one.



Algorithm 5. Quadratic Break

Input: *L*: list of items ranked by accuracy, *N*: number of items to recommend, λ: level of diversity

Output: *topN*: list of N items to recommend

```
1:  topN[0] = pop first item from L
2:  n = 1
3:  while n < N do
4:     max_distance = 0
5:     for each item i in L do
6:        d=1 – sim_list(i, topN)
7:        if d > λ then
8:           topN[n] = pop item i from L
9:           n = n + 1
10:          break for
11:       else if d > max_distance then
12:          max_distance = d
13:          max_item = i
14:       end if
15:       if i is the last item of L then
16:          topN[n] = pop item max_item from L
17:          n = n + 1
18:       end if
19:    end for
20: end while
```

4.3.6. Bounded Quadratic

The *bounded* version of the *Quadratic* method (Smyth and McClave, 2001) only takes into account the first *N* * *B* items of *L* (for instance, if the system wants to make N=10 recommendations and B –the *boundedness* factor- is set to 3, the 10 selected items will be taken from the initial 30 elements in *L*). Intuitively, the results will be more accurate but less diverse, although the computational cost will be heavily reduced because in each iteration only *B**N* elements will be analysed. The implementation of this method would be exactly like Algorithm 3, except that in the loop in line 5 it would not consider all the elements of *L* but only those in the first *B**N* positions.

4.3.7. Cluster Random

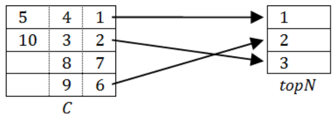
Aytekin and Karakaya proposed the idea of *clustering* the domain items to improve the diversity of the recommendations, by selecting items from different clusters (Aytekin and Karakaya, 2014). However, their clustering procedure was based on the *ratings* given by users; thus, it does not assure that the elements of a cluster are semantically similar (very different kinds of items could receive similar ratings). We propose to use this idea, but using a *semantically-based* clustering method. In this way, similar items will be in the same cluster and, if the RS picks up items from different clusters, they will probably be quite diverse.

The clustering of items is made offline using the well-known *k-means* algorithm (Hartigan and Wong, 1979). The process would be executed periodically to classify new items in clusters. The distance used to group items in each cluster is the *ontology-based semantic distance OSD* defined in section 4.2. The number of clusters k to be created is application-dependent.

The *Cluster Random* method, shown in Algorithm 6, picks up in each iteration the first element (i.e. the most accurate one) of a randomly selected cluster. The intuitive idea is that the results should be more varied than those of the pure *Random* method, because the elements in different clusters are semantically different. They should also be more accurate, since the selected items are the best ones of their clusters.

Algorithm 6. Cluster Random

Input: C: list of clusters C_j (in each cluster items are ranked by accuracy), N: number of items to recommend



Output: *topN*: list of N items to recommend

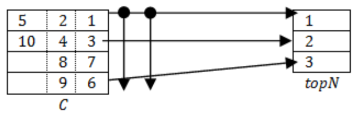
1: $n = 0$
2: **while** $n < N$ **do**
3: $topN[n]$ = pop item from list of random C_j
4: $n = n + 1$
5: **end while**

The algorithm takes as input the result of the clustering procedure (a list of semantically-related clusters C_1, C_2, C_3, ...). Each cluster contains a list of elements, ordered according to their relatedness to the user's preferences. In each of the N iterations a cluster is randomly selected and its first element is moved to *topN* (line 3). The same cluster could be chosen in more than one iteration (note that the number of classes could actually be smaller than N). The aim of this procedure is to select items that have a good accuracy but also offer a good degree of semantic diversity.

4.3.8. Cluster Quadratic

The idea of the pre-clustering procedure may also be applied to the *Quadratic* algorithm. In this case the computational cost will be heavily reduced, since the iterations are made on the list of clusters rather than on the original list of items, whereas the accuracy and the diversity of the results will be maintained.

The algorithm starts by moving the first item of the ranked list L to the *topN* list (line 1). Thereafter, the algorithm behaves as the *Quadratic* method (Algorithm 3); however, the iterations are made only over the first (i.e. best) items of each cluster. In line 6 the first item of each cluster iteration is considered, and a balanced score of its accuracy and diversity (with respect to the items in *topN*) is calculated (line 8). The element with the best score is selected in each iteration. The computational cost will be much lower than the one of the *Quadratic* method, since the inner loop only considers the k clusters, and not all the elements in L.



Algorithm 7. Cluster Quadratic

Input: L: list of items ordered by accuracy, *C*: list of clusters C_i (in each cluster items are ranked by accuracy), *N*: number of items to recommend, λ: level of diversity

Output: *topN*: list of N items to recommend

```
topN[0] = pop first item from L
n = 1
max = 0
while n < N do
    for p in 1..k do
        i = first item from cluster C_p
        d=1 – sim_list(i, topN)
        q = ( λ * d ) + ( (1 - λ) * weight of i )
        if (q > max) then
            max = q
            best_item = i; best_cluster=p
        end if
    end for
    topN[n] = pop best_item from C_p
    n = n + 1
end while
```

4.3.9. Temporal costs

In Table 10 we show the worst-case temporal cost of each of the methods described in this section. *N* is the number of items to suggest to a user (in a real case it could be in the 8-10 range). *L* is the size of the initial ranked list of items calculated by the RS. This size will depend on the database and on the capability of the recommender to filter out the items that do not fit well enough with the user's preferences, but the number of items could be very large (in the thousands). Therefore, *L* is the parameter that will penalise more heavily the temporal cost. The *Clustering* and *Bounded* methods try to avoid the repetitive analysis of all the elements in *L*. The cost of the clustering-based methods depends on the number of clusters (*C*). This number is application-dependent, but it would usually be between 10 and 30. The *Boundedness Factor* (*B*) will be analysed explicitly in the next section in a specific example, but it should not be a very high number if we desire an efficient bounded method.

Table 10. Table of temporal costs for each diversity method

Method	Cost
None	*O (1)*
Random	*O (N)*
Quadratic	*O (L*N*(N-1)/2)*
Linear	*O (L*N/2)*
Quadratic Break	*O (L*N*(N-1)/2)*
Bounded Quadratic	*O (B*N*N*(N-1)/2)*
Cluster Random	*O (N)*
Cluster Quadratic	*O (C*N*(N-1)/2)*

Quadratic and *Quadratic Break* are the methods with higher costs, since they depend on *L*N*(N-1)/2* in the worst case. However, notice that, in a real scenario, *Quadratic Break* will probably have a much lower running time, since it stops every iteration as soon as it finds an element that is different enough from the previously selected ones. The factor *N*(N-1)/2* appears in several methods because the length of *TopN* increases in each iteration, and we must add the cost of comparing an item with all the elements of *TopN* in each loop. In the *Linear* case it has been assumed that the selected elements are evenly distributed in *L*. All the methods whose cost depends on *L* are quite expensive, since *L* is supposed to be orders of magnitude larger than *N*, *B* or *C*. The *Bounded Quadratic* method reduces the cost with respect to the *Quadratic* one, because *B*N* should still be much smaller than *L*. *Cluster Quadratic* is much more efficient than *Quadratic* or *Quadratic Break*, because the number of clusters is much smaller than the number of recommendable items.

It also has to be taken into account that the clustering methods have the additional cost to execute the *k-means* algorithm on the list *L* to obtain *C* classes. The complexity of the algorithm can be noted as *O(LCT)* where *T* is the (usually small) number of iterations of the process. Since it is a time-consuming process (which, moreover, should be periodically repeated) it should be performed off-line.

4.4. Results

This section presents the results of the application of the previous diversification methods on the results offered by SigTur/E-destination, the recommendation engine for tourist trips detailed in chapter 5. This system combines several techniques to create ranked items based on the interests of the user, interests from similar users and the context of the trip (like the user location or the budget). Given a certain touristic activity to be evaluated, SigTur/E-destination computes a different score with each of these recommendation methods, which indicates if the item should be recommended to the user or not. These scores are aggregated into a single measure to obtain a final evaluation of each item, used by the system to decide the activities that fit better with the user. The aggregated score, which is normalised between 0 and 1, represents the level of accuracy of the recommendation of the item. Thus, the system calculates a ranked list of the activities that fit better with the user's preferences, demographic data and contextual information.

In order to assess the similarity between two activities, or the similarity between one activity and those that have already been selected, the semantic similarity distances defined in equations (9) and (10) were used with the tourism ontology detailed in chapter 5. Context has not been taken into account in the computation of the similarity (e.g. two History museums are very similar, even if they are located in very different geographical points).

A thorough study of the accuracy and the diversity of the results obtained with the approaches described in the last section is now reported.

4.4.1. Evaluation

We want to evaluate how the methods defined on section 4.3 influence the accuracy and the diversity of the recommendations provided by SigTur/E-destination. It will be considered that the recommendation process employed by the system is correct and it indeed returns a list in which the recommendable items are sorted according to their adequacy to the user. Thus, that initial list is taken to have a 100% accuracy. Each of the diversification methods will choose a subset of the items of the list, decreasing the accuracy but (hopefully) increasing the variety of the results. The final aim is to reach a satisfactory level of diversification with a minimum loss of accuracy (without incurring in a heavy computational cost). The size of the initial list is, on average, 872 elements. Clustering methods group these elements in 23 different clusters. The algorithm selects 8 items to be shown to the user.

The following measures are used to evaluate each of the methods:

- *Diversity*: it is a measure of the pairwise dissimilarity (*1-similarity*) between all the items in the *topN* list (the list of selected items). The similarity between two items is computed with equation (9). The final diversity is computed by applying the OWA aggregator on a vector containing all the pairwise dissimilarities.

- *Precision*: it is computed as the percentage of items in *topN* that are *relevant* for the user. An activity is taken to be *relevant* if it was assigned a minimum score of 0.7 by the recommender engine.

- F_{PD}: the F-measure is the harmonic mean of precision and diversity:

$$\frac{2 \times precision \times diversity}{(precision + diversity)} \tag{11}$$

4.4.2. Bound selection

The *Bounded Quadratic* method has a parameter that is the boundedness factor (the number of elements of the initial list that are considered for selection). The first step of the analysis is to determine, in an empirical way, which is the optimal value for this parameter.

The following figures (Figure 34, Figure 35 and Figure 36) show the diversity, precision and F_{PD} measure of the results provided by the *Bounded Quadratic* method for different bounds (even numbers from 2 to 50). Each figure shows the results taking into account four different levels of diversityλ (from 0.2 to 0.8), which is another parameter of most of the diversification methods shown in section 4.3. A profile with random preferences was used in this test.

As expected, the greater is the required diversity level, the more diversity and the less precision are obtained. It may be seen that, in this particular RS, the diversity and precision reach a stable level when the bound is around 16-18. The combined F_{PD} measure keeps increasing until the bound is 12-14, and then it also stabilizes. As the precision drops quickly with bounds larger than 8 (when high levels of diversity are considered), in the rest of the experiments reported in this section the bound has been set to this number.

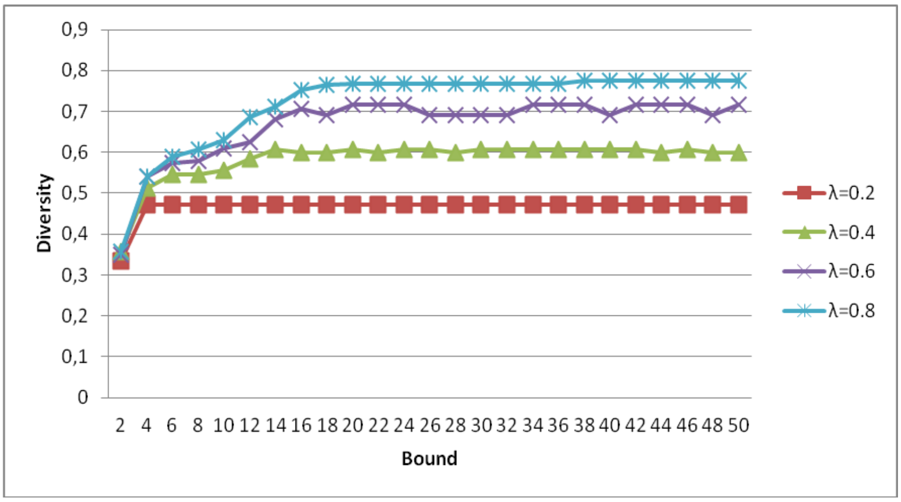


Figure 34. *Diversity* values on *Bounded Quadratic* method for each bound

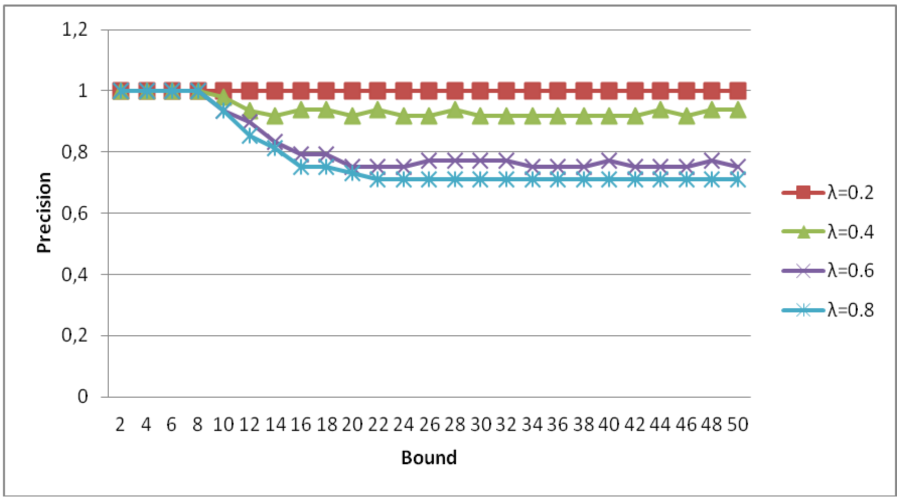


Figure 35. *Precision* values on *Bounded Quadratic* method for each bound

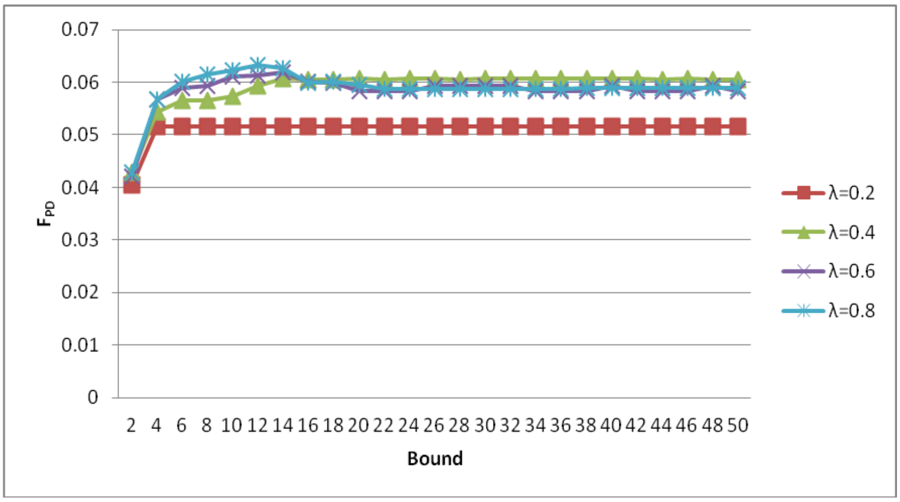


Figure 36. F_{PD} values on *Bounded Quadratic* method for each bound

4.4.3. Comparative analysis of the diversification mechanisms

In the SigTur/E-destination system the factor that has a stronger initial impact on the recommendations is the degree of interest on each motivation, explicitly given by the user in a questionnaire. This questionnaire allows to define the degree of interest in nine different motivations: beach, shopping, relaxation, leisure, culture, nature, gastronomy, sports and shows/events. The intuitive idea is that a tourist that sets high values on most of the motivations should be offered a very diverse list of recommendations, whereas a user that only chooses a few motivations is probably interested in visiting more specific places. Therefore, in the experiments shown in this section three different kinds of user profiles have been considered:

1. *General Profile:* the interests on the nine motivations are randomly set to values between 70% and 100%.

2. *Medium Profile:* the interests on five randomly selected motivations are set to random values between 70% and 100% (the remaining four motivations are given random interests lower than 30%).

3. *Specific Profile:* two randomly selected motivations are given random interests between 70% and 100%, and the other seven motivations are assigned random interests lower than 30%.

Figure 37, Figure 38 and Figure 39 show the diversity of the results offered by each of the methods for the three profiles, depending on the desired level of diversity λ. The *Quadratic* and *Cluster Quadratic* methods offer very good results. In the general and medium profiles they start to give diversified results for low values of λ, although for the specific profile they need a higher level of diversity. The *Bounded Quadratic* method offers similar results on the medium profile, but in the general and (especially) in the specific profile it offers lower levels of diversity, even when λ is high.

The reason is that the bound cuts off the items at the bottom of the initial list, which are the ones that could offer a high diversity. *Linear* and *Quadratic Break* give similar results, but they require a high level of diversity. Their curve is different from the one of the *Quadratic* and *Cluster Quadratic* methods because the meaning of λ, as described in the algorithms of section 4.3, is slightly different (in these latter methods it is the weight of the diversity with respect to the accuracy, whereas in the *Linear* and *Quadratic Break* techniques it is an absolute value of the required diversity). The performance of the *Random* and *Cluster Random* methods is not affected by the diversity level, but the diversity of their recommendations varies randomly. The diversity of the results offered by *None* does not depend on λ.

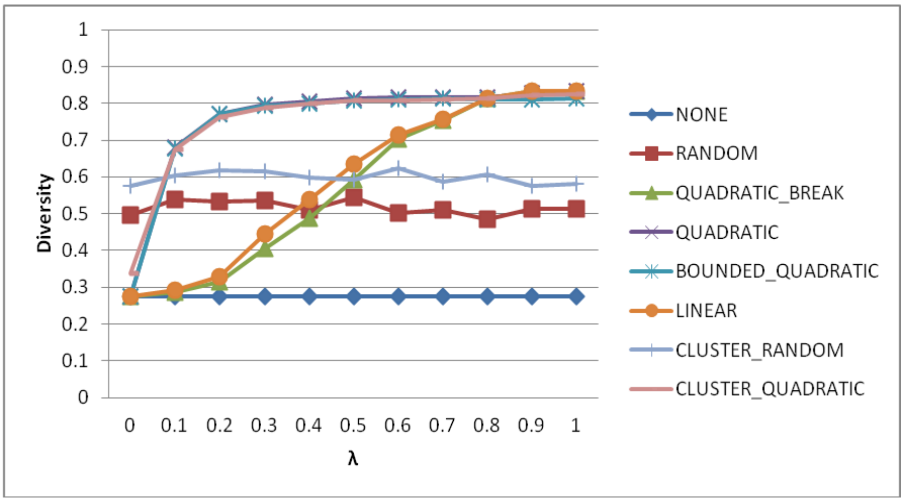


Figure 37. *Diversity* for the *General Profile*

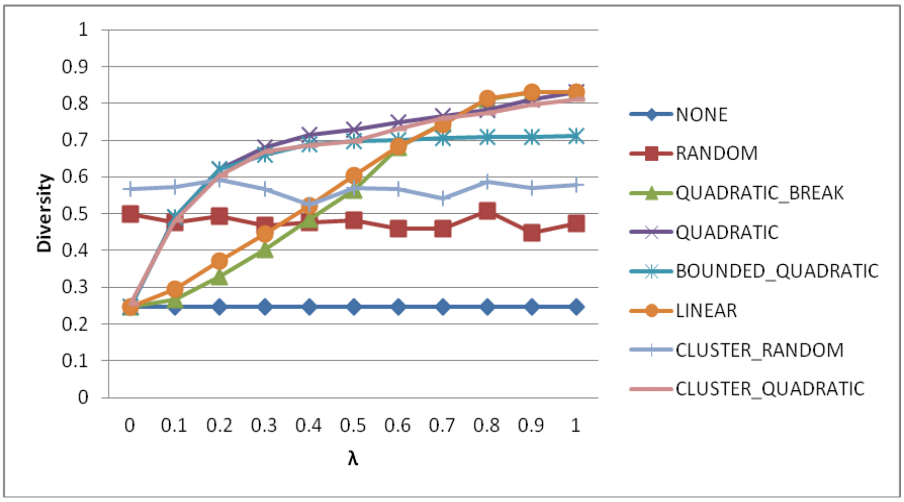


Figure 38. *Diversity* for the *Medium Profile*

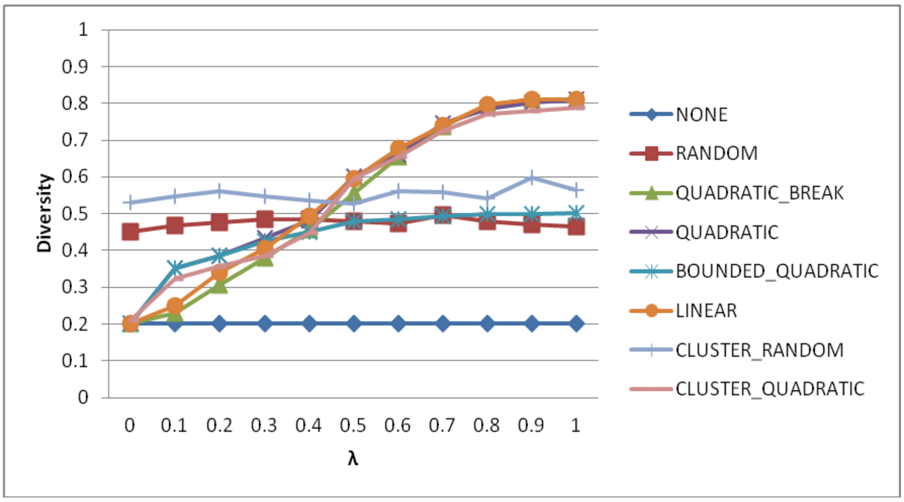


Figure 39. *Diversity* for the *Specific Profile*

Figure 40, Figure 41 and Figure 42 show the precision of the recommendations given by the different methods considering general, medium and specific profiles, respectively. The *Random* selection mechanism gives the worst results, as it merely suggests any item of the list. However, clustering the items before the random selection (*Cluster Random*) improves considerably the precision of the results, especially on the general profile. The reason is that items are clustered by similarity, and the best item (i.e. the most accurate) of the selected cluster is retrieved in each iteration. Two methods have a very high precision: *None* (which just returns the most accurate recommendations, without any consideration for diversity) and *Bounded Quadratic*. This method only considers the top *B*N* elements of the initial list to make the selection of the items to be recommended; if most of them have an accuracy over 0.7, the precision will be almost perfect. The remaining methods (*Linear*, *Quadratic*, *Quadratic Break* and *Cluster Quadratic*) reduce their precision when the value of required diversity is increased. If the profile is more specific, the precision decreases more quickly, even from low values of λ. The methods that are more influenced by the diversity level are *Linear* and *Quadratic Break*, whereas *Quadratic* and *Cluster Quadratic* are not so affected by high values of λ.

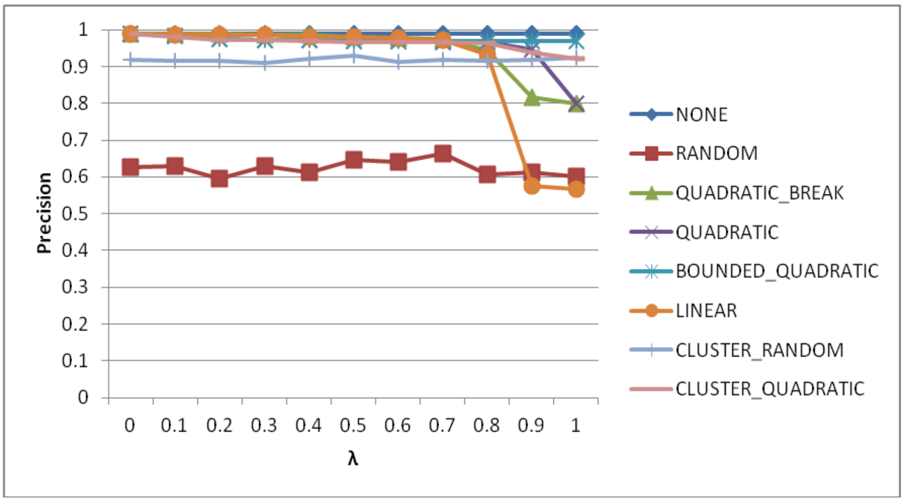


Figure 40. *Precision* for the *General Profile*

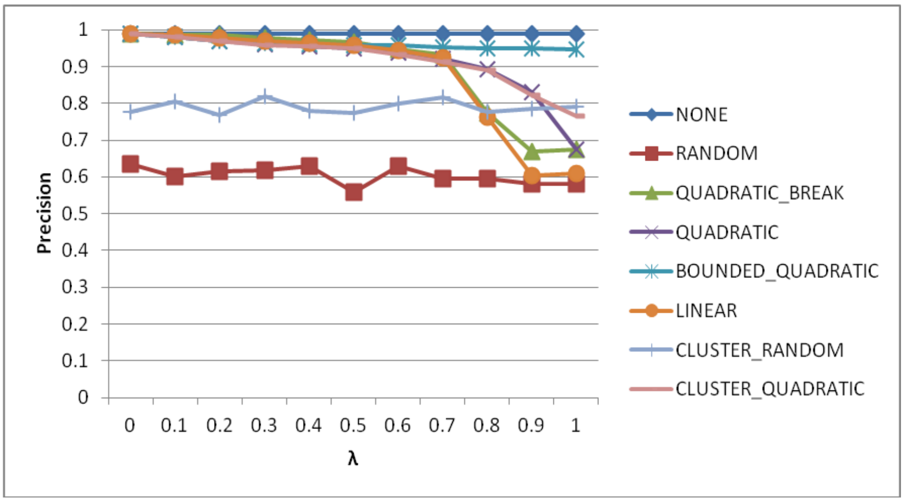


Figure 41. *Precision* for the *Medium Profile*

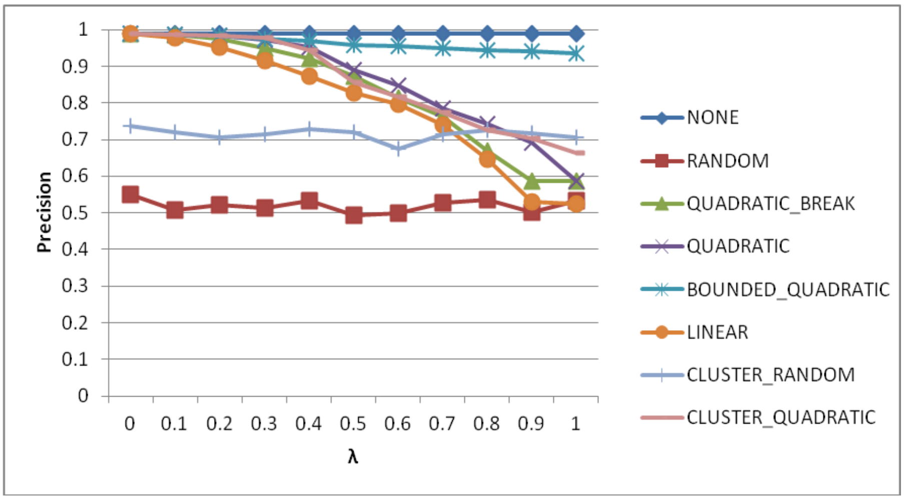


Figure 42. *Precision* for the *Specific Profile*

The previous figures have confirmed the intuition that, the higher is the value of the required diversity λ, the higher is the diversity and the lower is the precision of all the methods. As the objective is to have high levels in both dimensions, we are interested in analysing the behaviour of the F_{PD} measure, which provides a value that summarizes the global performance of the recommendation method. Figure 43, Figure 44 and Figure 45 show the results of the methods for the three kinds of profiles. Clearly the *None* and *Random* method offer the worst results. The former has a perfect precision, but its overall performance is heavily penalised by its lack of consideration of the diversity of the results. The latter does not guarantee either accuracy or diversity. As previously commented, a *clusterisation* of the items before the random selection improves the precision (and, therefore, the overall performance) of the method, especially for general profiles.

The method that seems to offer a best combination between precision and diversity across a wide range of required diversity levels is the *Quadratic* one. As seen in section 4.3, this mechanism analyzes all the options in each iteration and selects the one that offers a best compromise between these two perspectives. *Cluster Quadratic* and *Bounded Quadratic* also offer very competitive results, especially in the case of general profiles. This latter method does not decrease its performance for high values of λ, because the bound acts as a roof on the achievable degree of diversity. Finally, the performance of the *Linear* and *Quadratic Break* mechanisms is hampered by their lack of precision, especially in general profiles, because they select the first item that has enough diversity with respect to the previously chosen ones, without taking into consideration its accuracy.

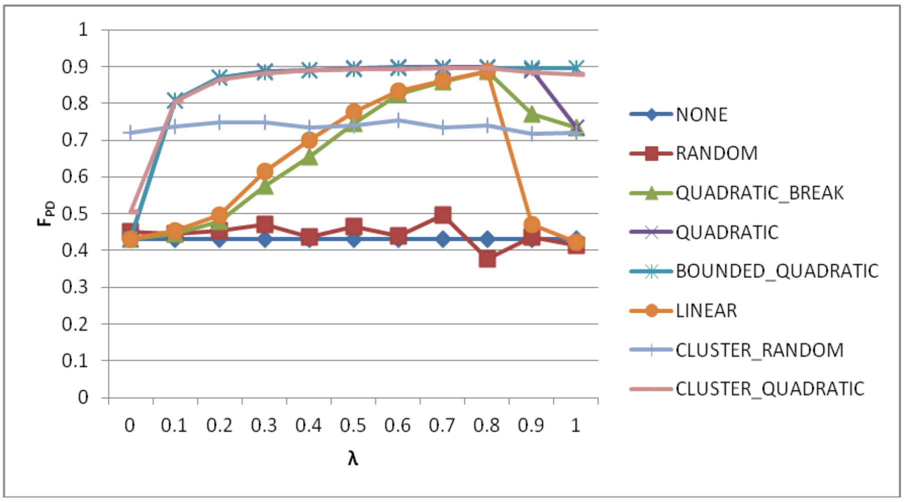


Figure 43. F_{PD} for the *General Profile*

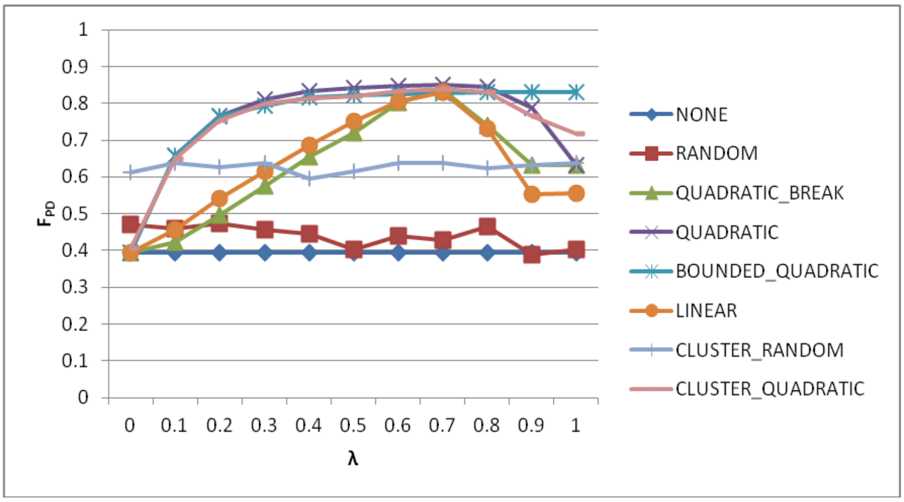


Figure 44. F_{PD} for the *Medium Profile*

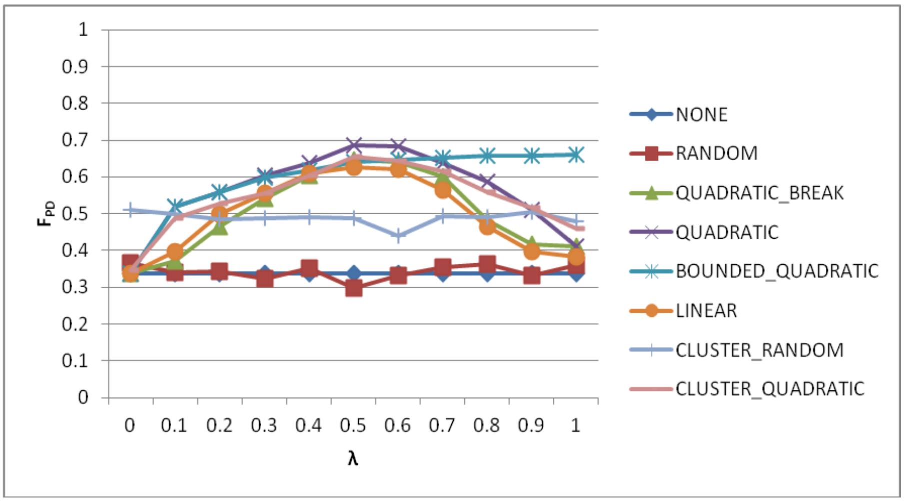


Figure 45. F_{PD} for the *Specific Profile*

Another way to evaluate the loss of accuracy with respect to the gain in diversity is the comparison between *Precision vs. Diversity*, shown in the plots depicted in Figure 46, Figure 47 and Figure 48 for the three profiles. The x-axis shows how the diversity increases as λ grows from left to right, whereas the y-axis shows the corresponding decrease in precision. The general profile maintains a very high accuracy until a high diversity degree (0.72) is required. The reason is that this profile models a user with a wide range of interests; thus, even if the system applies a high diversification, the user keeps receiving items that fit with his/her preferences. The impact of λ in the diversity of the results in the other two profiles is much higher, especially in the case of the specific profile. The *Quadratic* method is the one that offers a best performance for high diversity values, followed by *Cluster Quadratic*, *Linear* and *Quadratic Break*. *Bounded Quadratic* always offers a high precision but it does not reach relevant diversity values, especially if specific preferences are considered.

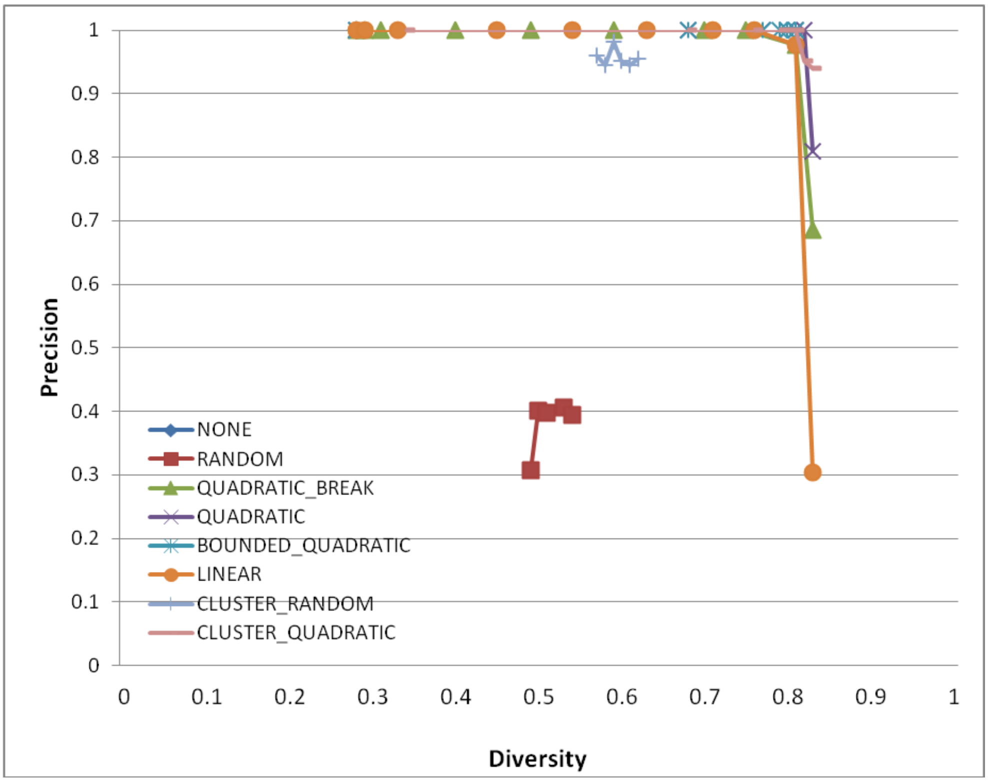


Figure 46. *Prec. vs Divers.* for the *General Profile*

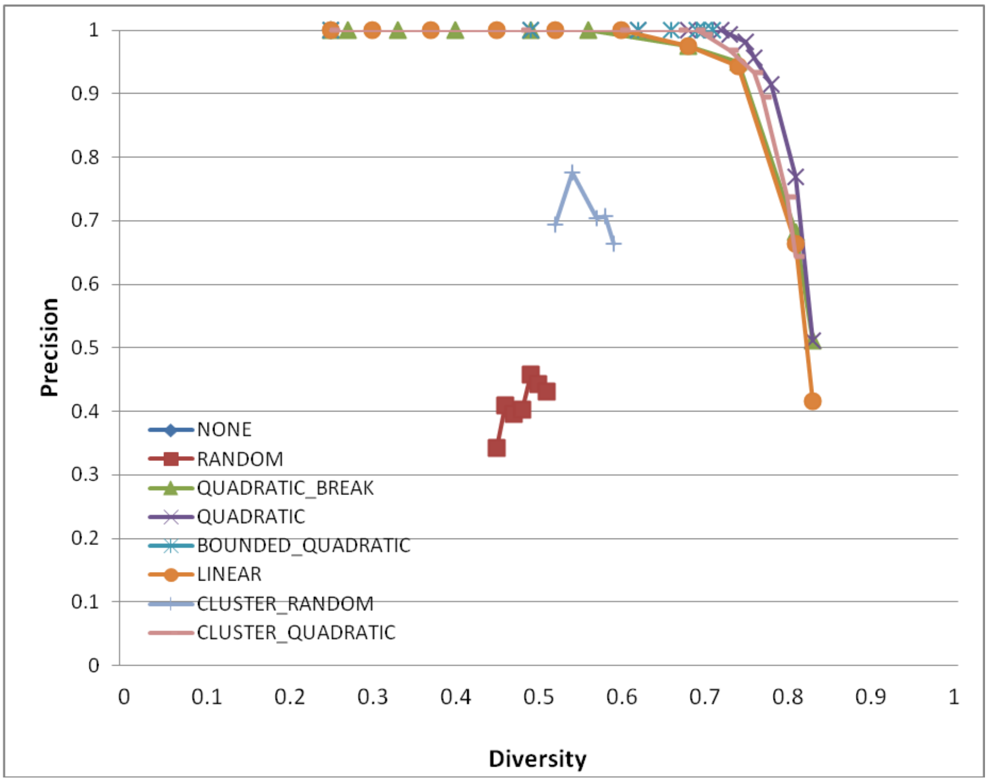


Figure 47. *Prec. vs Divers.* for the *Medium Profile*

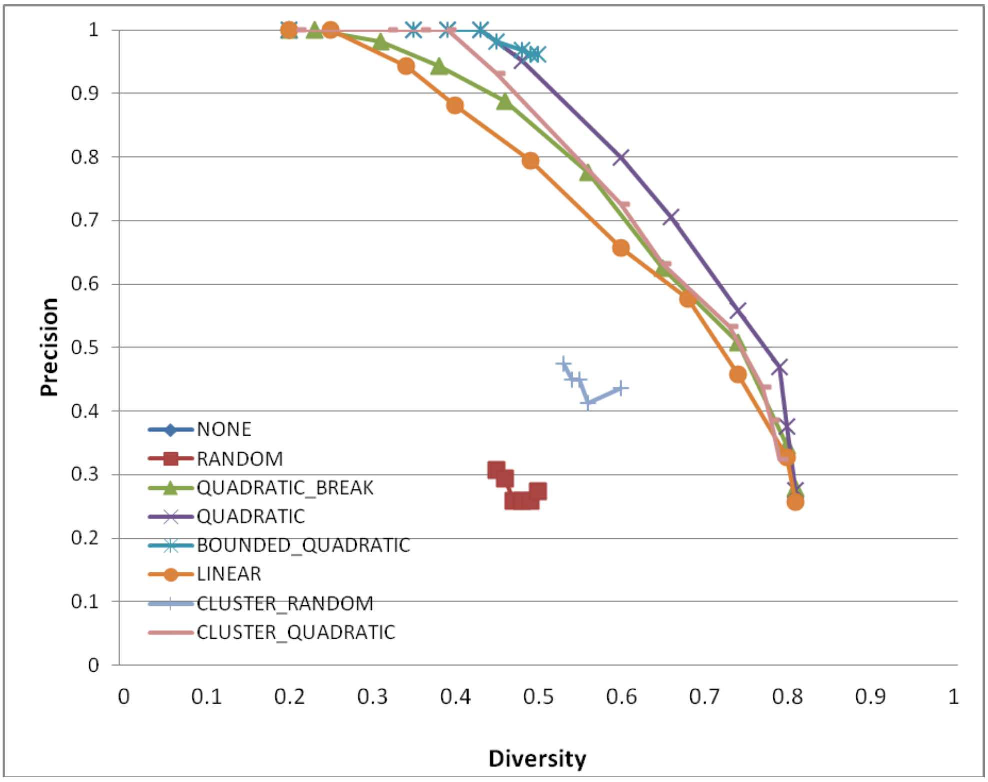


Figure 48. *Prec. vs Divers.* for the *Specific Profile*

The figures shown above may be used to automatically determine which value should be given to λ to obtain the best results for a particular user, depending on his/her degree of interest in the different travel motivations. It may be seen in the previous figures that the best value for a general profile should be around 0.7, whereas a medium profile gets the best results for values between 0.5 and 0.7 and a specific profile needs a low level of diversity (between 0.2 and 0.3) to offer an acceptable performance. Hence, the degree of diversity that the system should use depends on the kind of user, which can be determined by counting how many motivations the user is interested in. Therefore, the value of λ may be set dynamically with the following formula, where *#chosen_motivations* is the number of motivations in which the user has shown an interest above 30% and *#motivations* is the total number of available motivations (9 in SigTur/E-Destination):

$$\lambda = 0.25 + \left(\frac{\#chosen\_motivations}{\#motivations} \times 0.5 \right) \tag{12}$$

Finally, we show the results of the analysis of 270 user profiles with random motivation values. The parameter λ is dynamically set for each profile as described in the previous paragraph. Figure 49, Figure 50 and Figure 51 show, for each diversification mechanism, the averaged results for *Diversity*, *Precision* and F_{PD}, respectively. Table 11 details the values of such charts. The diversity in the initial results (without any selection process) is very low (0.24). A simple Random choice already doubles the diversity (0.5). There are 4 methods that offer a level of diversity between

0.57 and 0.65: *Cluster Random*, *Quadratic Break*, *Linear* and *Bounded Quadratic*. *Quadratic* and *Cluster Quadratic* are the ones that offer highest diversity with values of 0.72 and 0.70 respectively. All the methods offer a precision over 0.9, except *Random* and *Cluster Random*. *Bounded Quadratic* offers better results than *Quadratic*, *Quadratic Break*, *Cluster Quadratic* and *Linear* because the bound puts a limit in the achievable diversification, improving its accuracy. Looking at the global F_{PD} results, the 3 methods that offered more diversity have values around 0.8 (*Quadratic* (0.81), *Cluster Quadratic* (0.79) and *Bounded Quadratic* (0.78)). Two methods slightly exceed 0.7 (*Linear* and *Quadratic Break*), and even *Cluster Random* has a result well above the two baseline methods *None* and *Random*.

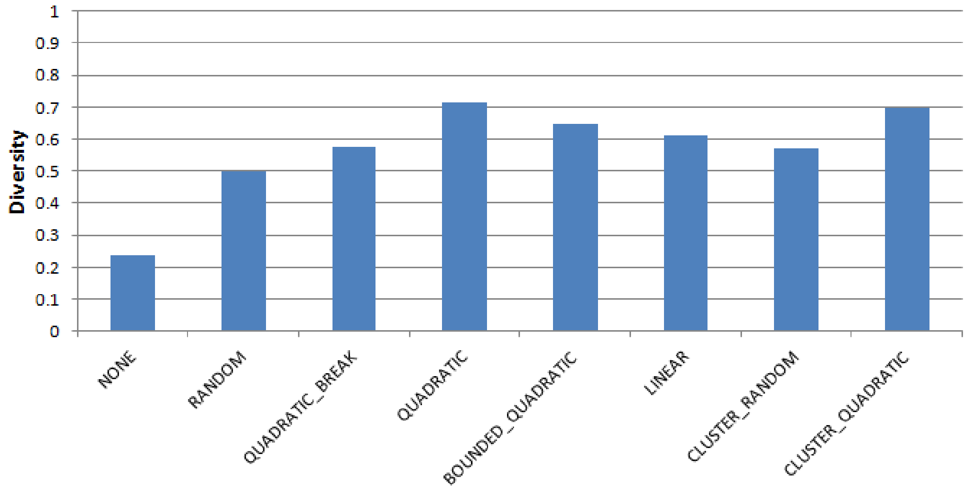


Figure 49. *Diversity* with dynamic λ

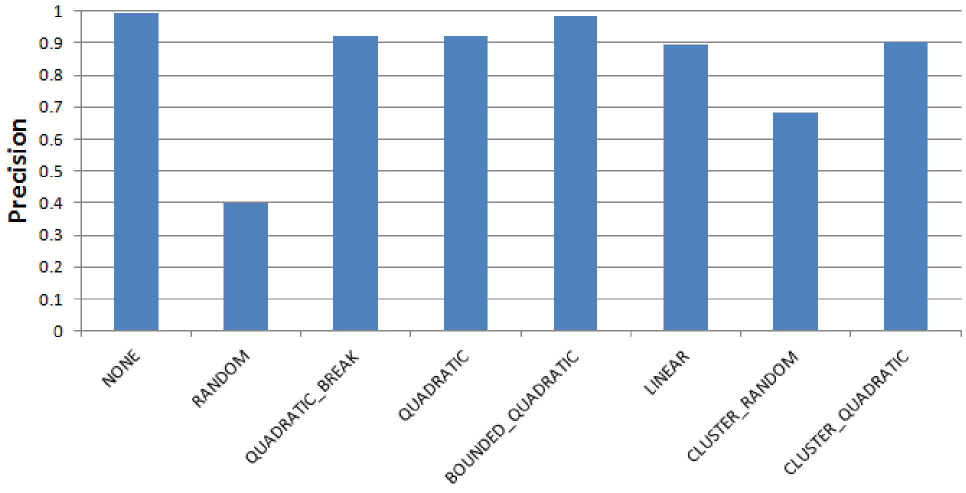


Figure 50. *Precision* with dynamic λ

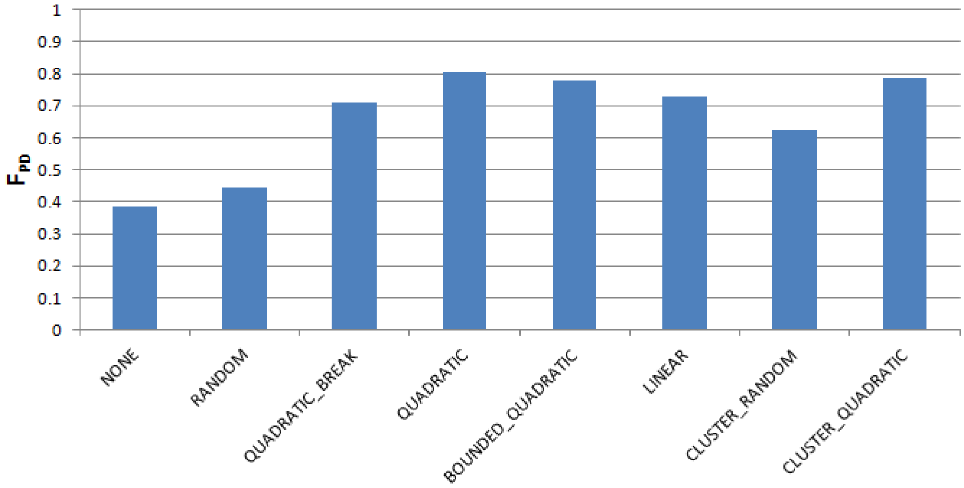


Figure 51. F_{PD} with dynamic λ

Table 11. *Diversity*, *Precision* and F_{PD} values for all methods

	Diversity	*Precision*	F_{PD}
None	0.24	1.00	0.39
Random	0.50	0.40	0.44
Quadratic Break	0.58	0.92	0.71
Quadratic	0.72	0.92	0.81
Bounded Quadratic	0.65	0.98	0.78
Linear	0.61	0.90	0.72
Cluster Random	0.57	0.68	0.62
Cluster Quadratic	0.70	0.90	0.79

These results must be compared with the time required by each of the algorithms. Figure 52 shows the number of iterations of each method, and Figure 53 shows the same information without the *Quadratic* method (which, as seen in section 4.3.9, has an extremely high computational cost). The cost of *Linear* and *Quadratic Break* also depends on the size of the initial list, hampering their performance. *Bounded Quadratic*, despite the bound, also has a very high cost. The new method proposed in this dissertation, *Cluster Quadratic*, seems the best overall alternative, since it provides almost the same performance level and it has a much lower computational cost. Note that these figures do not include the temporal cost of the clustering procedures in *Cluster Random* and *Cluster Quadratic*, which are assumed to be made off-line.

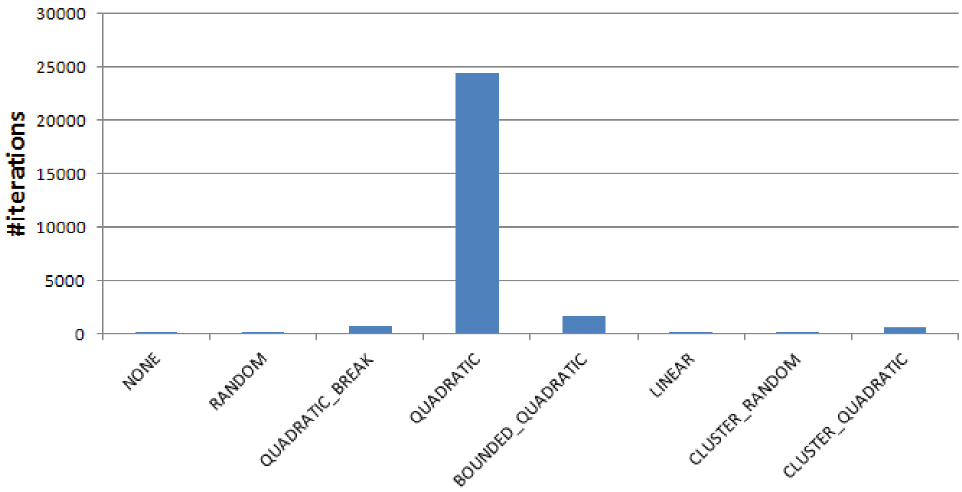


Figure 52. Number of iterations for each method

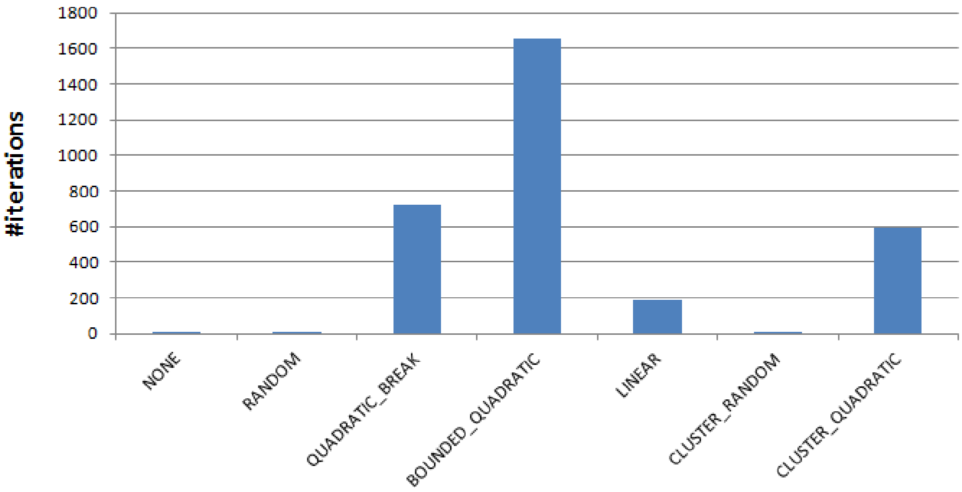


Figure 53. Number of iterations for each method (except Quadratic)

4.5. Conclusions and future work

In this chapter we have described a family of diversification methods, based on the selection of some items from an initial list of recommendations computed by the system. Some variations of previous methods and a new selection algorithm based on semantic clustering have also been proposed. All the methods have been thoroughly tested in the ontology-based personalised recommender of touristic activities that is detailed in chapter 5.

The results of the tests show that the *Quadratic* method is the one that gives the best combination of diversity and accuracy. The main reason is that it loops for all the items of the list to find the item that best combines both diversity and accuracy. However, it is not suitable to be run on real time since its computation costs are extremely high (see Figure 52). The remaining methods try to reach similar results more efficiently. For instance, a limitation on the number of items to loop is given on the *Bounded Quadratic* method. Despite the important time reduction with respect to the basic *Quadratic* method, it is still way more expensive than the rest of the methods (see Figure 53). *Lineal* and *Quadratic Break* try to reduce its computational costs without needing to find the best combination, stopping the selection process whenever they find an item that offers enough diversity. Finally, the novel *Clustering Quadratic* method reduces heavily the computation cost by pre-grouping semantically similar items. Then, the selection loop can be performed through the clusters, and not through the much longer list of items. Moreover, it may be argued that the clustering methodology is more scalable and adaptable to other datasets since the clustering process is based on the semantic similarities between items.

In this chapter we have also proposed to dynamically adapt the level of diversification depending on the initial general preferences of the user. Hence, for *generic* users, i.e. those that have a wide range of interests, the degree of diversity can be high since they are willing to accept more diverse items. On the other hand, in the case of those users that are interested on a more concrete set of topics, the degree of diversification should be much lower.

In the future work we want to explore the other two families of diversification mechanisms (see section 4.1). To evaluate those methods that integrate recommendation and diversity we plan to include diversification mechanisms within the recommendation algorithm of SigTur/E-Destination. The idea would be to include in the ranking process of each item some measure of serendipity (or unexpectedness), hoping that the inclusion of serendipitous results will increase the overall satisfaction of the user (a more explicit way to measure this output, either explicitly or implicitly, should also be devised). The study of the methods that offer aggregate diversity is also very interesting from the Tourism point of view, because Destination Management Organisations are very keen on diversifying the tourist offer and increasing the flow of tourists in the less popular and well-known attractions.

Chapter 5 – Intelligent Tourism Recommender System for the province of Tarragona

This chapter details the design and implementation of SigTur/E-Destination, a recommender system for tourists that want to plan a visit to the province of Tarragona. This system, which has been developed in collaboration with the Science & Technology Park for Tourism and Leisure, provides personalised suggestions using the techniques proposed through this dissertation. Such customization will produce a large number of different plans created for each particular tourist, which could be used to balance the tourism activity, spatially, thematically and financially, with important returns in terms of sustainable development at the destination.

The system offers a Web-based and a mobile interface that facilitates the user interaction and provides a better experience both in the travel preparation stage and during the trip. Concerning the information used by the recommender, it takes into account demographic data, the travel context (e.g. travel budget), 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.

SigTur/E-Destination combines many recommendation techniques, from the use of stereotypes (standard tourist segments) to content-based and collaborative filtering techniques. As will be shown in this chapter, the Artificial Intelligence tools used in SigTur/E-Destination include automatic clustering algorithms, *Multiple Criteria Decision Aid* (MCDA) methods, ontology management, semantic diversification 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 (as explained in chapter 3). 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.

The system includes *Geographic Information Systems* (GIS) 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. A GIS database was designed according to territorial singularities of the Tarragona region. It was also decided to build a specific ontology that fits the specificities of this territory. The design of this new domain ontology was inspired by 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.

This chapter is structured as follows. The next section explains the design of the system, focusing on the technical architecture and the creation of the Tourism domain ontology and its integration with the GIS database. The following sections explain how the system gathers explicit and implicit information from the user (section 5.2) and how this knowledge is used by different kinds of recommender algorithms (section 5.3). Section 5.4 explains how the suggestions provided by a variety of recommenders are aggregated to find out the final recommendations to be made to the user. After that, the functionalities that allow users to plan their trips are explained. Finally, the last section contains the conclusions of the chapter.

5.1. Design of the system

This section describes the architecture of SigTur/E-Destination and its main components. One of the basic knowledge structures of the system is a Tourism ontology, whose design and implementation are also commented. The system requires a large amount of spatial and non-spatial data associated with different resources and activities, so it is a logical choice to use a GIS for storing, managing, analyzing and visualizing these data. The processing tasks related with the GIS and its integration with the Tourism ontology are also explained in this section.

5.1.1. Architecture of the SigTur/E-Destination system

Figure 54 depicts the general architecture of SigTur/E-Destination. All the modules, which have been fully 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 client application that shows information on maps and lists that are very easy to manage.

The client side of the system is accessible through a Web browser and mobile devices. For the Web version, the application has been built using the HTML[27] and jQuery[28] languages, which permit to enhance the user experience with dynamic pages as if they were using an application. The connection with the server has been achieved with Java Server Faces[29], which is a standard component-oriented user interface (UI) framework for the Java EE platform. For the mobile version, as commented in chapter 2, ad-hoc apps for both Android[30] and iOS[31] platforms have been developed. These apps communicate with the server with a RESTful (REpresentational State Transfer) API (*Application Programming Interface*). Data packets are sent in the JSON[32] format, which allows the transformation of structured data into plain text.

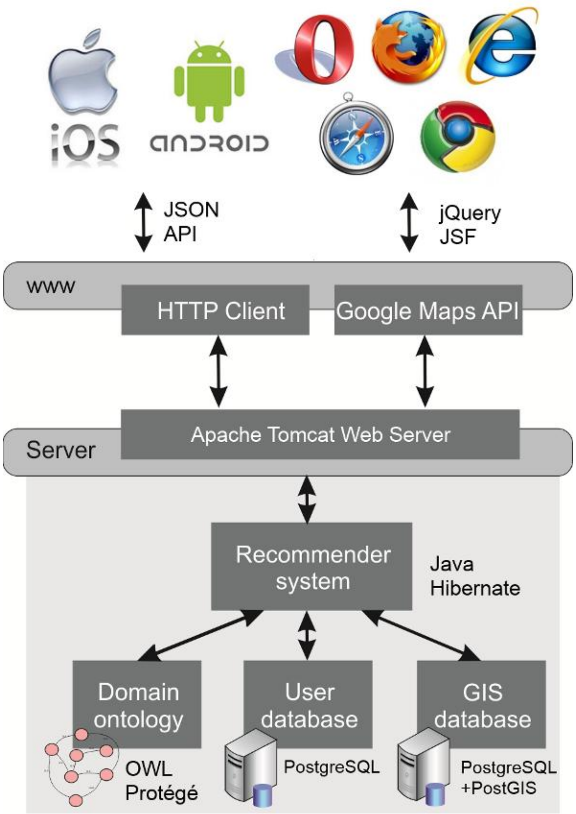


Figure 54. General architecture of the system

The representation of the geographical resources in maps has been achieved using the Google Maps API[33]. This set of functions 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.

---

27 http://www.w3schools.com/html (last access February, 2015)
28 http://jquery.com/ (last access February, 2015)
29 http://www.javaserverfaces.org/ (last access February, 2015)
30 http://developer.android.com/index.html (last access February, 2015)
31 https://developer.apple.com/technologies/ios/ (last access February, 2015)
32 http://json.org (last access February, 2015)
33 https://developers.google.com/maps (last access February, 2015)

On the other side there is a Java-based server, whose core is the recommender system. It handles the interaction between all the modules and manages the user profile dynamically, updating its state after each user action. This allows the recommender system to take into account the behaviour of the user and provide more accurate results. The system employs two databases to store information. One of them is a GIS database that 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[34] and tourist resources are stored with the PostGIS[35] extension that allows running GIS processes, such as querying the information available within a geographical boundary box. Database connections are managed by the Hibernate[36] framework with a spatial extension that handles geographic data. Some modifications have been applied to Hibernate to improve the pool of database connections. In order to process spatial functions over tourist resources, such as computing the distance between two points, the JTS Topology Suite API[37] has been used. Databases are not only used to store 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.

5.1.2. Tourism ontology

As described in chapter 3, *domain ontologies* contain the definition of the basic concepts, relationships, properties and instances of a given domain. They define areas of common understanding between multiple actors, easing their interoperability and permitting a high-level communication (Berners-Lee et al., 2001).

In the last decades the Tourism sector has developed catalogues and taxonomies to facilitate the management of information in this field. Lately, an effort to generate global standards has been made in order to ease the exchange of data between Tourism agents. This is the case of the Thesaurus on Tourism and Leisure Activities[38] defined by the World Tourism Organization (WTO).

Different Tourism ontologies have been developed in the last years. 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). Harmonise[39] was one of the first ontologies that aimed to face the interoperability problems of Tourism, focusing on the exchange of data between

34 http://www.postgresql.org/ (last access February, 2015)
35 http://postgis.refractions.net/ (last access February, 2015)
36 http://hibernate.org/ (last access February, 2015)
37 http://www.vividsolutions.com/jts/ (last access February, 2015))
38 http://www.wtoelibrary.org/content/m7434p/ (last access February, 2015)
39 http://www.thesaurus.com/browse/harmonize (last access February, 2015)

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 1,000 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), was created to establish a shared structure for multi-modal and multi-lingual Tourism question answering. The DERI e-tourism ontology (Hepp et al., 2006), developed in the OnTour research 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 (Ruíz-Martínez et al., 2011). cDOTT (*Core Domain Ontology for Travel and Tourism*, (Barta et al., 2009)), developed in 2009, was based on the Harmonise 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. (Buján et al., 2013) designed the tourism ontology TourExp[40] that enhanced the existing ones by defining not only general concepts but also the restrictions and needs of tourists profiles. (Sriharee, 2014) developed different ontologies that contained touristic concepts (e.g. Museum, Church, Island, Palace, etc.) and designed a procedure for the automatic association of articles to them.

In SigTur we need to represent the main tourism and leisure resources in the GIS database. These resources are characterized not only by the "3S" (Sea, Sand, Sun) tourism, predominant in the region of Tarragona, but also by the distinctive features of the territory, such as eno-gastronomy, cultural heritage, or leisure (Anton-Clavé, 2010). Therefore, the GIS database was designed according to these territorial singularities and, consequently, the system required a specific ontology that fitted perfectly with the GIS database. That is why we decided to design a new domain ontology following the principles of the thesaurus of the World Tourism Organization but adjusting it to these specificities. The ontology, which was manually created, represents up to 203 connected concepts in 5 hierarchy levels. As shown in Figure 55 and Figure 56, the ontology is structured around eight main concepts (red circles in Figure 55) that constitute the first level of the hierarchy: *Events*, *Nature*, *Culture*, *Leisure*, *Sports*, *Towns*, *Routes* and *ViewPoints*. 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*.

[40] http://tourexp.morelab.deusto.es/ont/tourexp (last access March 2015)

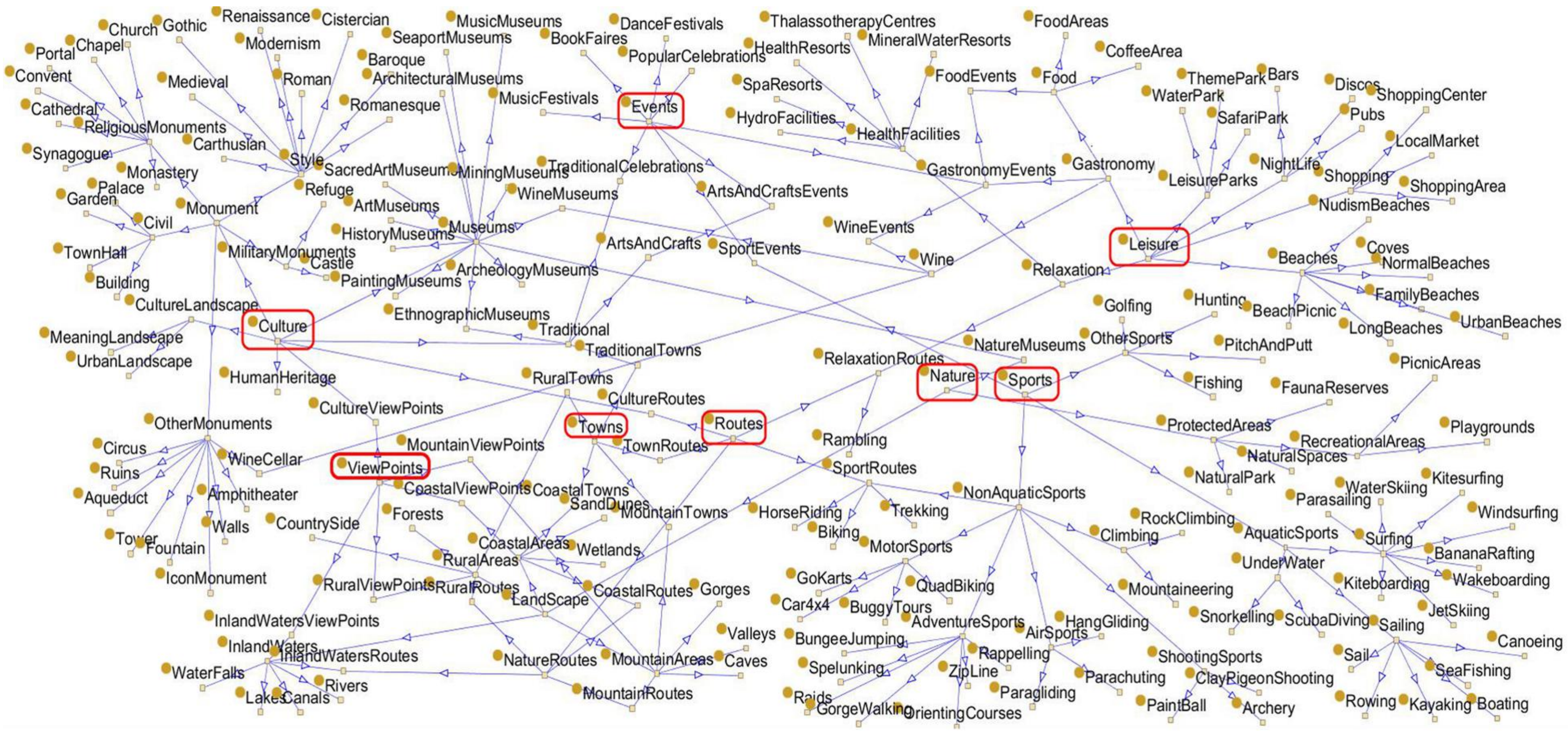


Figure 55. The Tourism ontology of the system

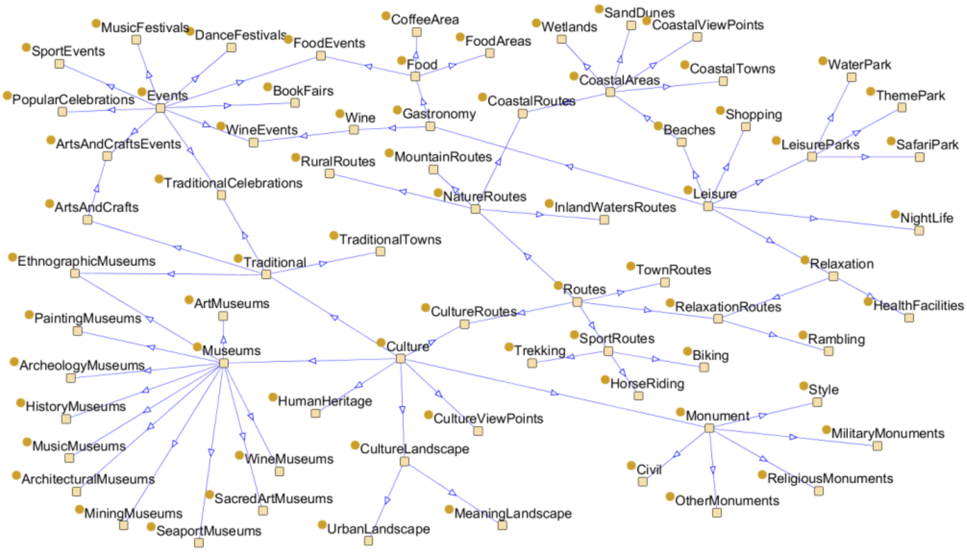


Figure 56. Portion of the Tourism ontology of the system

The ontology has been developed using the Thesaurus of the World Tourism Organization as a reference guide to represent the touristic and leisure activities in the Tarragona 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 Science & Technology Park for Tourism and Leisure. The level of detail in each part of the ontology depends on the set of activities available in the 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 were 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 later. Each activity is tagged with one or more ontology concepts, which are leaves (or low level nodes) in the hierarchy. For instance, the Roman Amphitheatre of Tarragona is tagged with the following concepts: *HistoryMuseums*, *Roman*, *HumanHeritage*, *Romanesque* and *Amphitheatre*. 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 Figure 54).

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 Protégé[41] editor and it is represented in the OWL language. Jena[42] is the Java Semantic Web framework used at the core of the recommender system. It provides tools to manage the ontology and to apply inference mechanisms based on rules.

5.1.3. Geospatial features and GIS database

A first step to develop SigTur/E-Destination was to collect various data sets of tourism resources (leisure activities, cultural heritage, natural spaces, sport activities, routes and events) of the Tarragona province to build the GIS database. This information was spread in different government administrations; therefore, the first task was to request these data sets. Most of them were obtained from Diputació de Tarragona, although an important part was provided by Generalitat de Catalunya. The acquired data sets were in multiple formats: Shapefiles, GPS formats and mostly spreadsheets. There was an extensive work of converting formats before uploading them to the GIS database. Regarding spreadsheets, there was an additional task of geo-coding, since these types of files are not geo-referenced. However, most of the activities had an address field that permitted to obtain their coordinates. In the cases in which data about tourist resources were not available, they were manually generated. Therefore, there was an exhaustive task of documentation and digitization before adding new activities.

The activities of the GIS database of SigTur/E-Destination are grouped into six categories: *leisure*, *sports*, *culture*, *nature*, *events* and *routes*. The last two play a cross-cutting role, since they can be related to any of the other categories. Items associated to the ontology concepts *towns* and *viewpoints* are always stored in one of these categories (for instance, an item tagged as *cultureviewpoints* or *traditionaltowns* would be stored in the *culture* category). *Leisure* contains five entities (equivalent to tables or map layers): beaches, theme parks, spa centers, shopping areas and nightlife areas. The data of these entities have been added to the database with special care, performing an exhaustive documentation task, since they are the main tourist attractions in Tarragona. *Sports* have been classified in two subcategories: aquatic and non-aquatic. *Culture* includes two entities: cultural heritage assets and museums. They are stored in different tables since the structure of their information is relatively different. *Nature* contains two entities: natural spaces, which encompass all the natural spaces protected by law, and the recreational areas contained within these spaces. *Events* include temporary activities (such as fairs, festivals, traditional celebrations, and so on) that can be programmed throughout the year in any of the other categories. Finally, *routes* include three entities that can also be

[41] This work uses the stable release 3.4.7 of Protégé editor available at http://protege.stanford.edu (last access February, 2015)
[42] For more information: http://jena.sourceforge.net (last access February, 2015)

related to the other categories: walking routes, biking routes and driving routes. Figure 57 illustrates the schema of the GIS database.

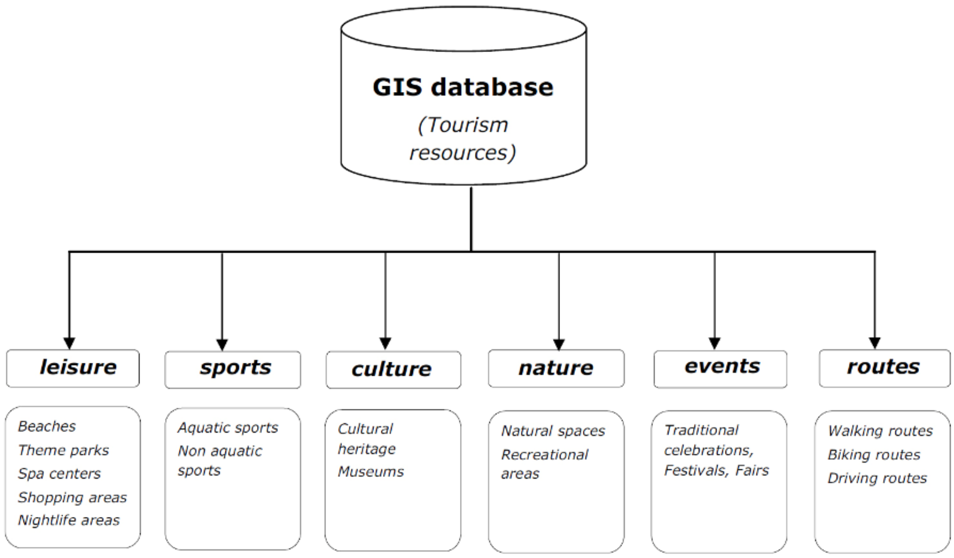


Figure 57. GIS database schema

The geographical entities of a GIS database have geometric properties that can be modelled by the measurements, properties and relationships of points, lines, angles and surfaces. Two types of geometric data types are prevalent: *Raster* data and *Vector* data. Raster data are represented with an array of points, where each point represents the value of an attribute for a real-world area. Vector data include points, lines and polygons, all of which are representations of the space occupied by real-world entities (Baumann, 1994). The GIS database of SigTur/E-Destination belongs to the Vector data type, including points and lines. Except routes, the rest of the entities are composed by points. Routes are formed by lines (or multilines), since they store the tracks of each itinerary.

Currently, the GIS database contains over a thousand resources. Nevertheless, there is still a considerable ongoing work on adding new resources and updating the existing ones. In any case, the GIS database has been designed in order to easily support these future additions and updates. Besides, the structure of the tables is as similar as possible, containing mostly the same fields, which facilitates the management of the database and massive operations. Additional fields were added just in the cases in which tourist resources required specific information. The edition of geo-referenced data was made with the Open Source software Quantum[43], as shown in Figure 58. This tool enables to create and edit geometric points and lines.

[43] http://www.quantum.com/ (last access February 2015)

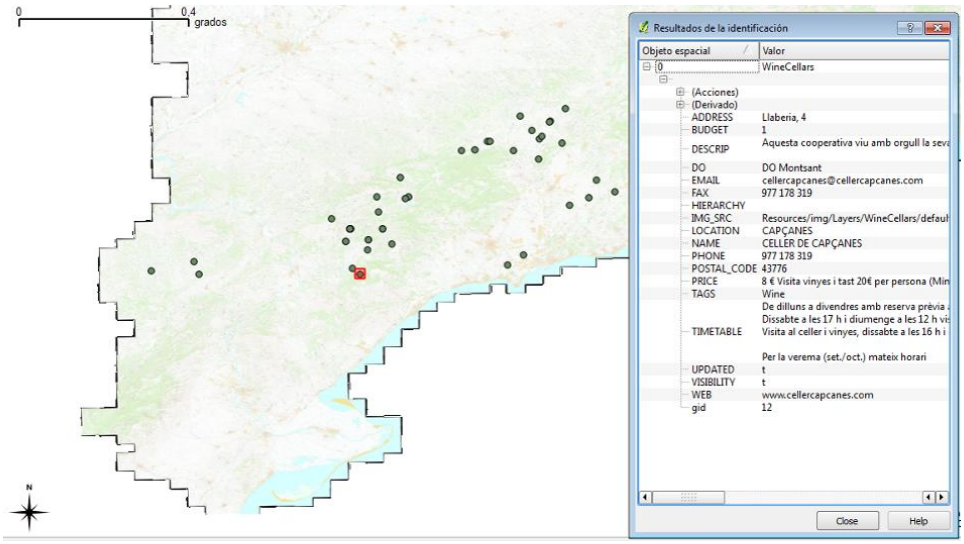


Figure 58. Geo-coding resources with Quantum

5.1.4. *Integration of the Tourism domain ontology and the GIS database*

As detailed above, the GIS database and the ontology domain have a similar structure, as the main categories are the top concepts of the ontology (except v*iewPoints* and *towns*, which link different categories with their ontology relationships). Each activity of the GIS database is represented by at least one concept in the domain ontology. For instance, as there are three types of museums in the GIS database (archaeology museums, history museums and anthropology museums), there are three concepts in the domain ontology, located as subclasses of the museum concept. Besides, each museum-related resource stored in the GIS database has a tag with the concept of the domain ontology. These tags facilitate the classification of each item and, moreover, they allow the interaction between the GIS database and the domain ontology, enabling the system to yield recommendations properly. Currently, the number of concept definitions of the ontology is higher than the activity types in the GIS database. This fact provides efficiency to the system, since it makes it easier to add new activity types to the GIS database and to reuse the ontology data model in other scenarios.

5.2. User profile management

The SigTur/E-Destination recommender system manages a user profile that is composed by two parts: (1) a static one, which is a vector with demographic and travel information and (2) a dynamic one, represented with an instantiation of the Tourism ontology, which contains the user's degree of interest on each type of activity. For instance, if the current user of the system likes visiting 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.

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 discover the user's preferences, the application acquires both *explicit* and *implicit* information from him/her. The former includes the specification of the travel motivations and the rating values given by explicit evaluations of items. The latter is obtained from the observation of the actions of the user on the system, such as requesting more information about a certain activity or adding it to the travel planner. The next subsections give more details on these two types of user feedback.

5.2.1. 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 Tarragona 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, leisure, culture, nature, gastronomy, sports and shows/events. Each of these motivations corresponds to a concept stored as a class in the Tourist ontology that may either be at the top level of the ontology, such as *Leisure* or *Culture*, or at lower levels, such as *Beaches* or *Shopping*. Even though the concepts *Beaches* and *Shopping* are children of the *Leisure* concept, we decided to ask independently about the three motivations due to their importance on the survey analysis. Figure 59 shows the interface used by the tourist to enter the degree of relevance (0-100%) of each of these motivations. These values are stored in the ontology of the user to initialize 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 of the session (see form in Figure 60). These data include information about the country of origin of the user, other people the user travels with (the allowed values are shown in Table 12), the location of the accommodation, the type of accommodation (allowed values are also shown in Table 12), 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 (travel dates) or to locate the recommendations into a given geographical area (near the chosen destination).

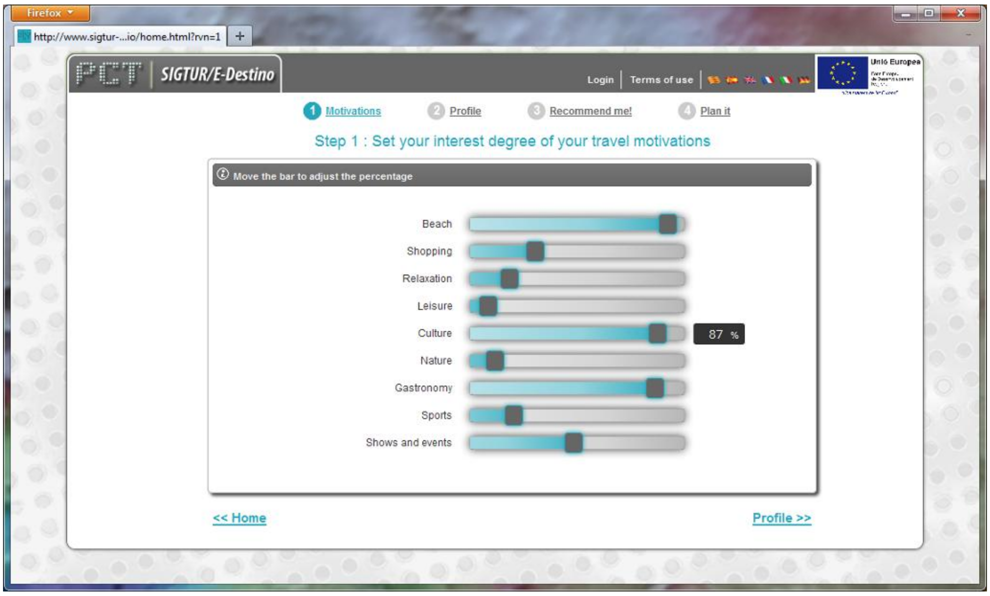


Figure 59. Initial form to discover the user's travel motivations.

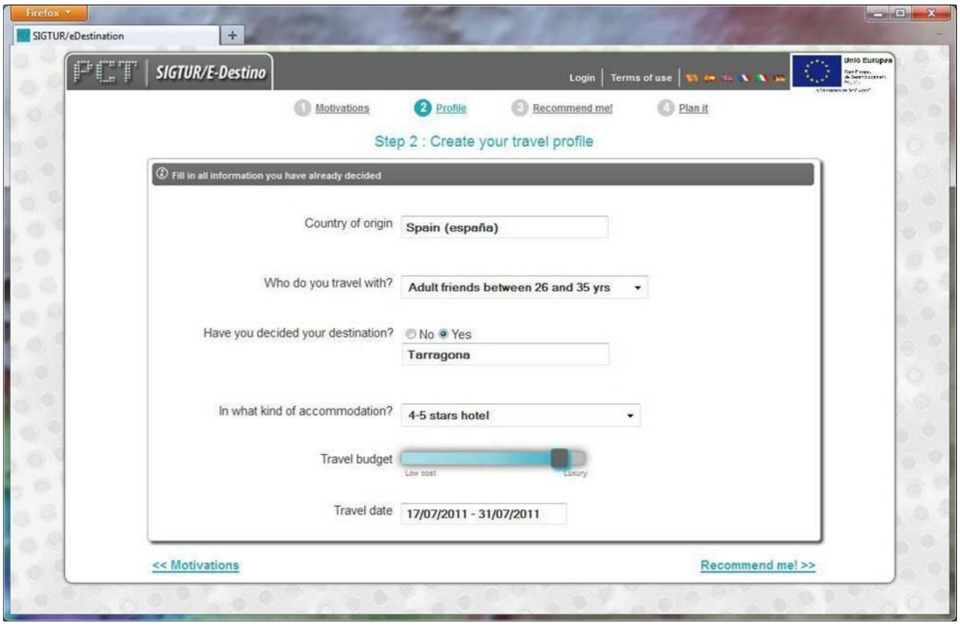


Figure 60. Form to obtain details about the travel.

The selection of the demographic questions proposed in Figure 60 is based on a study of previous data collected from tourists of the same region. We considered 30,000 questionnaires filled in by tourists between 2001 and 2009 with the aim of discovering what kinds of activities they visit (depending on the characteristics of the users). The study was conducted by the Costa Daurada Tourism Observatory[44] and the characteristics asked at the questionnaire where among others: country of origin, age, profession, sex, travel group composition (allowed values shown in Table 12), type of accommodation used (allowed values shown in Table 12), motivations (open answers) or transport means used during the trip. After that, each

[44] http://www.pct-turisme.cat/cat/innovacio_fetcd.html (last access March 2015)

visitor had to indicate whether he/she had performed at least one activity of the following types during the stay: beach, relax, health and care, leisure events, nature and culture activities, rambling, shopping, events, sports or nightlife.

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. 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). This statistical model is used to predict a binary response of a categorical variable based on other independent variables. We have computed it with the SPSS software[45] processing the whole dataset of the questionnaires, obtaining the discrimination weight of each variable (tourists' characteristics). The variables with higher weight were selected as demographic questions for the system: country of origin, travel group composition, type of accommodation and motivations, while other variables from the questionnaire were discarded due to their low discrimination value, such as age, sex, profession, social class or the number of previous visits to our region.

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, in which they express explicitly their degree of satisfaction. Users may rate activities with an integer value between 1 and 5, where 5 corresponds to the best.

Table 12. Allowed values for visiting groups and accommodation types

Criterion	Allowed values
Travel group composition	With children between 0 and 5 years old
	With children between 6 and 12 years old
	With children more than 12 years old
	With adult relatives less than 35 years old
	With adult relatives more than 36 years old
	Adult friends less than 25 years old
	Adult friends between 26 and 35 years old
	Adult friends more than 35 years old
	Senior group
	School
	Alone
	Business and others
Accommodation type	4-5 stars hotel
	3 stars hotel
	1-2 stars hotel
	Apartment rented through an agency
	Rented apartment
	Camping
	Own home (second home)
	Family-friends home

45 http://www-01.ibm.com/software/analytics/spss/ (last access March 2015)

5.2.2. 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 the information about the user's preferences and its recommendations. This information, which is implicit in the user behaviour, is commented in this section.

Once the user obtains a list of recommendations (the way in which the system produces the recommendations is explained in the next section) the user is able to make several actions on the proposed activities. The system is able to infer the user's interests by capturing and analyzing these actions. This process 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 may select those activities he/she is interested in and add them to a *travel plan*. Other actions the user is able to make on activities are to request more detailed information on a specific event, to ask for activities geographically close to the currently selected one or to obtain activities that are similar to the current one. Section 5.5 describes how the user can perform all these actions, which can be considered as 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 explained in the next section.

5.3. Recommendation techniques

This section explains how the system predicts the degree of interest of the user on each type of activity, that is, how the ontology-based profile is maintained and exploited. The aim is to suggest a ranked list of activities that are interesting to the user and varied enough. The following sections explain how the system employs content-based and collaborative recommendation techniques, as well as contextual parameters, to adjust the recommendations to the user's needs.

5.3.1. Content-based recommendation

Content-based recommenders (Pazzani et al., 2007) are based on a direct matching between the features of the activities to be recommended and the interest of the user 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 Figure 54). Each of the activities is labelled with a list of concepts belonging to the Tourism ontology introduced in section 5.1.2. The basic aim of the recommender system is to associate a degree of preference to each concept of the ontology according to different information that the system can gather; let us denote the preference score as $S \in [0..1]$. With 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 $CL \in [0..1]$ on that value, as was discussed in chapter 3. This confidence level depends on the evidences that have led to the computation of the interest degree on that particular concept.

5.3.1.1. Travel motivation

To estimate the preferences, the initial information provided by the user is the motivation of the travel, given in the form shown previously on Figure 59. The levels of interest on the nine possible motivations are 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 stores a score value of S=0.85 to the ontology concept *Beaches* and the confidence level in that concept is set to CL=1.0 since that information has been provided directly and explicitly by the user. Afterwards, the system performs the downwards propagation explained in section 3.4.1.2 to establish initial preference values throughout all ontology concepts, and then the system is able to start providing the first recommendations of items based on these motivations (process detailed in section 5.4). Note that in the general preference initialization process described in section 3.4.1.1 the initial preferences were set at the first-level concepts of the ontology. In SigTur some motivation-related concepts are located in lowers levels (e.g. *Beaches*), and moreover, they can be descendants of other motivations (e.g. *Beaches* is a child of *Leisure*). In this case the *Leisure* preference values are not propagated downwards to *Beaches* and its descendants, since the user has already defined explicitly his/her level of motivation on that concept. However, the system indeed uses this relationship during the upwards propagation process (explained next), to predict that a user that gives positive feedback about beaches is also providing positive feedback on leisure. In SigTur the parameter α is reduced to 0.15 for the downwards propagation since the ontology applied in this system has more hierarchy levels than in the example in section 3.4.1.2.

5.3.1.2. Interaction of the user with the system

As commented in section 5.2.2, the analysis of the actions of the user in the system also provides implicit information on his/her interests, which can be used to refine the degree of preference 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 (e.g. handling directly the travel plan is more relevant than merely asking for more information).

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 confidence values for implicit actions are set lower than the ones for explicit actions, since they are considered less accurate (Kelly and Teevan, 2006). Table 13 shows the range of possible score (*s*) values and the default confidence level (*w*) 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.

Table 13. Scores and confidence levels for different kinds of explicit/implicit information

Action ID	Action type	*s*	*w*
action1	Add activity to travel planner	1.0	0.5
action2	Remove activity from travel planner	-1.0	0.5
action3	Request detailed information about an activity	1.0	0.3
action4	Request activities similar to the current one	1.0	0.2
action5	Request activities near the current one	1.0	0.2
action6	Rating of an activity	[-1.0, 1.0]	1.0
action7	No actions on a recommended item	0.0	0.15

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, as was explained in section 3.4.2. The list of user actions (with their associated scores and weights) has been adapted to this particular recommender system (from the basic list shown in Table 8 in section 3.4.2) although the updating mechanism (equations (2) and (3)) is generic enough to be usable in any domain. The system updates automatically during the user session the preference and confidence values for all the ontology concepts associated to the recommended activities with which the user interacts.

5.3.1.3. 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. The ontology structure may be leveraged 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 (e.g. someone interested in *Archeology Museums* 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. This process, detailed in section 3.4.2, has two steps: *upwards propagation* (in which the interests on the ancestors of the modified leaves are updated) and *downwards propagation* (in which the preference and confidence on other descendants of these ancestors are also updated).

5.3.2. Collaborative recommendation techniques

Collaborative filtering techniques are recommendation methods based on the opinions of a set of users about the items available in the domain. 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 the analysis of 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. Serendipity 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 interactions provided by the users (two users are similar if they performed similar interactions 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 interacted with any item, the first kind of similarity is applied. When the user has already interacted with a certain number of recommended activities, the second kind of similarity takes more relevance.

In order to perform a user-based collaborative recommendation it is necessary to have a way to compare 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 motivational attributes, which is explained in the next subsection.

Due to its scalability in computation time, the K-means algorithm has been applied to make the different clustering processes in the system, which will be commented in sections 5.3.2.2, 5.3.2.3 and 5.3.2.4 (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 (section 5.3.2.1) 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 the clustering is applied with four different purposes: to obtain a basic initial set of tourist segments (5.3.2.2), to obtain classes of users with similar demographic characteristics (section 5.3.2.3) and to obtain classes of users with similar interactions on items (section 5.3.2.4) or interactions on ontology concepts (section 5.3.2.5). In fact, as will be explained later, the system stores different interest scores and confidence values for both activities and ontology concepts depending on the set of clusters. First, from a tourist segment it can be obtained the interest values of generic concepts of the ontology. Then, with a group of similar demographic attributes it is obtained the interests on each activity (section 5.3.2.3) and the interests on ontology concepts (as explained at the end of section 5.3.2.5). From the group of users with similar interactions on activities (section 5.3.2.4) it is obtained the interests on each activity. Finally, from the group of users with similar interactions on ontology concepts (section 5.3.2.5) it is obtained the interests on each ontology concept. All these data are finally aggregated into a unique preference and confidence value for each activity, as will be described later in section 5.4.2.

5.3.2.1. Measure of similarity between two users

In order to create groups of similar users and assign new users to a group, the system has to measure the similarity between users based on their characteristics. The values that are considered in the comparison process are the travel motivations (Figure 59), the travel group composition, the accommodation type and the country of origin (Figure 60).

As it has been said in section 5.2.1, before defining the similarity measure, a logistic regression analysis was applied on a set of 30,000 hand-filled questionnaires to obtain the degree of relevance of each attribute with respect to the discrimination of the travel activities performed by users. The discrimination relevance values for the selected variables are shown in Table 14. It may be noticed that the composition of the travel group is the most relevant factor.

Table 14. Discrimination values of demographic and motivational criteria

Attribute	Relevance
Travel group composition	0.37
Accommodation type	0.33
Country of origin	0.23
Travel motivation	0.07

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 $x=(x_1,x_2,\ldots,x_{12})$ 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 $R^n \to R$ 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:

$$\mathrm{OWA}(a_1,\ldots,a_n) = \sum_{j=1}^{n} w_j b_j \tag{13}$$

, where b_j is the j^{th} largest 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, 1996) as

$$Q_\alpha(r) = r^\alpha \tag{14}$$

$$w_i = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right) \tag{15}$$

, giving α the value 2 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. Hence, using α=2 the generated weight vector with n=9 (*W*) is: [0.012, 0.037, 0.062, 0.087, 0.111, 0.136, 0.160, 0.185, 0.21].

As an example, let us measure the similarity between the two users shown in Figure 61. We first measure for each motivation the inverse of the distance between the related preferences (e.g. in the *Shopping* motivation the result is 1-abs(0.9-0.8)=0.90). These 9 values are stored in descendent order in a vector *b*. In the example of Figure 61 this vector is [1.00, 0.97, 0.95, 0.94, 0.90, 0.90, 0.85, 0.80, 0.37]. Thereafter we apply the OWA operator (13) to the vectors *b* and *W* generated previously, giving a final result of 0.77 that represents the similarity in terms of motivations between the two users.

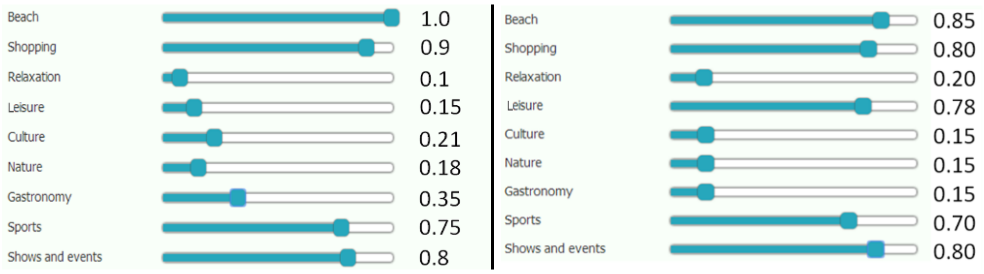


Figure 61. Travel motivations of two users *u* (left) and *v* (right).

After that, this evaluation of the similarity with respect to the travel motivations is combined with the comparison of the demographic features using the *Logic Scoring of Preferences* (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 employed to obtain the similarity between two users *u* and *v* is the following:

$$sim(u,v) = (w_1 x_1^r + w_2 x_2^r + w_3 x_3^r + w_4 x_4^r)^{1/r} \qquad (16)$$

In this expression, *r* has been set to 0.5 to specify a weak conjunction. The values w_1, w_2, w_3 and w_4 are the relevance weights from Table 14 for travel group composition, accommodation type, country of origin and travel motivations, respectively. x_1 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, 0.5 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*.

Following the previous example, let us assume that user *u* is a Spanish group of 25 years old friends that have rented an apartment, and user *v* is a French group of friends with the same characteristics. Thus, in this case x_1=1 (same travel group), x_2=1 (same accommodation), x_3=0 (different nationality) and x_4=0.77 (similarity of their motivations). Therefore, the final similarity between these users, *sim*(*u*,*v*), given by equation (16), is $(0.37 \times 1.0^{0.5} + 0.33 \times 1.0^{0.5} + 0.23 \times 0^{0.5} + 0.07 \times 0.77^{0.5})^{1/0.5} = 0.58$.

This similarity measure is used in different steps of the recommendation process, as explained in the following sections.

5.3.2.2. Estimating the interests from similar segments of tourists

A common problem in collaborative recommender systems is the initial lack of users. To solve it, 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 Tarragona is used. Therefore, the system is initially enriched with the preferences associated to tourist segments obtained from a survey of 30,000 questionnaires conducted in this area between 2001 and 2009. As was explained in 5.2.1, we take as attributes the country of origin, the travel group type, the accommodation type and the travel motivations. Users could employ a free list of keywords to express their main motivations, whereas the group types and accommodation were chosen from the options shown in Table 12. Tourists also had to explain which kinds of activity types they had performed during their stay (the available options are beach, sports, relaxation, shopping, etc.).

An automatic clustering process, based on the well-known *k-means* algorithm, was applied to the set of tourist responses, using the measure of similarity described in the previous section). 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), which quantifies both the separability between categories and the homogeneity within categories, considering different numbers of clusters and cut levels in the hierarchy. Afterwards, a *prototype* was calculated for each segment (Table 15 shows 10 of them). The value selected for the demographic data is the value with more selections in the group. Note that the same segment group may contain for example tourists that have different nationalities but are similar in other demographic data. In the attribute associated to motivations there is vector that contains the percentage of tourists that chose the same keyword (Table 15 only shows the most common keywords: *Beach* (*Bch*), *Relaxation* (*Rlx*), *Leisure* (*Leis*), *Culture* (*Cult*) and *Nature* (*Nat*). The attribute related to types of activities is also a vector, which contains, for each kind of activities, the percentage of cluster members that have performed it. The kinds of activities preferred by visitors were *Beach* (*Bch*), *Relaxation* (*Rlx*), *Leisure* (*Leis*), *Shopping* (*Shp*), *Sports* (*Spts*) and *Night Life* (*NL*). Table 15 shows the values of 10 prototypes, highlighting in dark orange and dark green their most relevant motivations and types of activities. For instance, prototype 83 (the first one on the table) grouped basically young Spanish tourists that stay on 3-stars hotels, travel with friends and are mainly interested in going to the beach. On the other hand, we can see that most of these tourists, apart from enjoying the beach, also like shopping (70% of them) and going out during the night (50% of them).

Table 15. Prototypes of 10 tourist segments

Demographic data				Motivations					Type of activities done					
Proto-type ID	Accomm-odation	Group type	Country of origin	Bch	Rlx	Leis	Cult	Nat	Bch	Rlx	Leis	Shp	Spts	NL
83	Hotel 3*	Friends < 25 years old	ES	65	5	18	14	8	80	46	45	70	3	50
4	Hotel 3*	Senior	ES	9	0	0	100	0	24	76	34	95	4	0
74	Own Home	Adult family more than 35 years old	FR	75	0	0	2	100	85	94	6	85	52	12
31	Camping	With children more than 12 years old	FR	73	4	3	9	18	95	57	51	79	15	9
41	Hotel 4-5*	With children more than 12 years old	ES	100	6	25	19	5	92	71	55	82	5	2
81	Own Home	Adult friends more than 35 years old	ES	24	0	0	0	100	34	94	3	88	25	0
93	Camping	Adult friends between 26 and 35 years old	ES	67	2	0	17	3	78	44	25	42	15	47
76	Agency apartment	Adult friends between 26 and 35 years old	UK	30	3	0	14	0	78	47	35	80	45	42
45	Hotel 3*	With children between 0 and 5 years old	ES	100	11	26	12	8	93	71	48	81	7	8
37	Camping	Adult family more than 35 years old	GE	79	0	3	16	60	85	68	8	86	16	1

When a new user logs into 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. Since each type of activity of the questionnaire can be associated to an ontology concept, the system assigns the average value of the most similar prototype as the score of the related concept. For example, when a user that is similar to the prototype 81 (sixth row on Table 4) enters the system, the scores of ontology concepts are filled with these activity type's values: *Beach* (0.34), *Relaxation* (0.94), *Leisure* (0.03), *Shopping* (0.88), *Sports* (0.25) and *Night Life* (0). The confidence level associated to each concept is the similarity measured with equation (16) between the current user and the selected prototype. Since the concepts obtained in this method are generic within the ontology, the system executes the same downwards propagation, explained in section 3.4.1.2, of such values to the lowest level concepts.

5.3.2.3. Estimating preferences from users with similar demographic characteristics

Since the basic 100 segments contain only generic types of activities that tourists may be interested in visiting, it is necessary to have a way of obtaining preference values on more precise types of activities (lowest level concepts) and, at the end, on particular activities to recommend. In order to do that, the system has to find out which users are similar (from the demographic and motivational points of view) to the current one. The equation (16), which takes into account the four aspects mentioned in Table 14, is used for this purpose.

Thus, when a new user arrives, he/she first specifies his/her demographic data and travel motivations (Figure 59 and Figure 60). 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 equations (2) and (3), the score and the confidence level for each activity, given the actions that have been performed on the activities by the members of the cluster. In addition, the similarity between the user and the cluster is multiplied by the confidence level (e.g. if the similarity between the user and the cluster is 0.9 and the *CL* of an item is 0.8, the final *CL* for such item will be 0.8 x 0.9=0.72).

The idea is to apply periodically the clustering procedure on the full set of users stored in the database, in order to take into account new users. This periodicity can typically be weekly, but it will depend on the number of new users of the system. The number of clusters was initially set to 100, taking as a reference the number of segments, since we considered that the kinds of tourists that use the application should be similar to the types of tourists discovered in the segment surveys. However, this fact has to be proved in the future when the volume of users registered at the system is high enough.

5.3.2.4. Estimating preferences from users with similar interaction with the system

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 correspondence with his/her demographic data and travel motivations. However, after the user interacts with the system and manipulates the recommended activities, SigTur/E-Destination is able to perform more accurate recommendations by finding other users that performed similar actions (from Table 13) on the same activities (like viewing or adding the same activity to the plan). 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 made more actions. Note that, in this case, users with similar actions can have very different demographic values (travel motivations, group composition, country of origin, etc.). This kind of similarity between two users u and v ($\omega_{u,v}$) is computed using the *Frequency-weighted Pearson Correlation* (FPC) (Breese et al., 1998):

$$\omega_{u,v} = FPC(u,v) = \frac{\sum_{i \in \chi_{u,v}} \lambda_i (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sum_{i \in \chi_{u,v}} \lambda_i (r_{u,i} - \bar{r}_u)^2 \sum_{i \in \chi_{u,v}} \lambda_i (r_{v,i} - \bar{r}_v)^2} \tag{17}$$

In this expression $\chi_{u,v}$ are the activities that have been manipulated by both *u* and *v*, $r_{u,i}$ is the weighted aggregated score given to an activity *i* by user *u* (with the weights shown in Table 2)

$$r_{u,i} = \frac{\sum_{a \in A_i} s_a w_a}{\sum_{a \in A_i} w_a} \tag{18}$$

and $\bar{r}_u$ is the weighted mean of all the actions made by user *u* on all the activities

$$\bar{r}_u = \frac{\sum_{a \in A} s_a w_a}{\sum_{a \in A} w_a} \tag{19}$$

The factors λ_i, defined in (20), have been included to increase the variety of the recommendations (Breese et al., 1998). This parameter takes into account the log-ratio of all users that have performed actions on an activity *i* as well as the ratios of the actions of *u* and *v* on activity *i* with respect to all the actions they have made. Thus, activities that have received fewer actions by all users have a higher relevance. On the other hand, activities with more actions performed by both *u* and *v* have a higher relevance. *U* is the weight accumulated from all the actions of all the users within the system and U_i is the accumulated weight of the actions of all the users on a particular activity *i*. R_{ui} corresponds to the accumulated weight from the actions of user *u* over activity *i*, and R_u is the accumulated weight of all the actions of the user on all items. R_{vi} and R_v are the same values for user *v*. The first factor in equation (20) was suggested in (Breese et al., 1998), whereas the other two factors are novel.

$$\begin{aligned} &\lambda_{i1} = \log\frac{|U|}{|U_i|}, \lambda_{i2} = \exp\frac{|R_{ui}|}{|R_u|}, \lambda_{i3} = \exp\frac{|R_{vi}|}{|R_v|}, \\ &\lambda_i = \lambda_{i1}\lambda_{i2}\lambda_{i3} \end{aligned} \tag{20}$$

Once this activity-based clustering has been obtained, the process to follow is the same one that was explained in the previous section: the system finds the cluster that is more similar to the current user, and then it can measure the preference and confidence level on the activities based on the actions done by the members of this cluster (using (2)). The similarity between the user and the cluster is multiplied by the CL of the values obtained for each activity.

5.3.2.5. Ontology-based collaborative recommendation

Collaborative recommendations based on user interactions on items require that each user performs a large number of actions on the system (such as viewing an item or adding it to the travel plan) to predict accurate recommendations. However, the probability of two users interacting with the same items in large data bases is relatively small (as happens in previous section). It is even more difficult in datasets where suggestions highly depend on geo-location (as will be explained in the next section), because users that visit different areas of the same city will probably not receive the same items, even if they have similar interests. To overcome this drawback we analyze not the actions on individual items but the actions made on activities of the same type (items associated to the same ontology concepts). In fact, we keep the two methods, the one that measures similarities on concept interactions that allows discovering new items, and also the method that measures similarities on item interactions (previous section) that allow more accurate recommendations when they are available.

For example, imagine a tourist that wants to visit Falset and adds to the travel plan the modernist wine cellar “Celler cooperatiu de Falset”, which is labelled with the tag *ModernistCellar*. Later, another user going to Valls adds to the trip the modernist wine cellar “Vinícula de Nulles”, also labelled with the same *ModernistCellar* ontology concept. The cities of Falset and Valls are both in the province of Tarragona but they are 57 km. away, so it seems clear that those users will probably not be recommended the same items and hence we could not calculate any similarity based on their interactions with the same activities. However, we may say that indeed they have similar tastes, since both of them are interested in visiting modernist cellars.

Therefore, the system performs a new clustering of users that are grouped depending on the concepts associated to the items they have manipulated. Hence, the users in the same group will have interacted with the same kind of items, even if they have not interacted exactly with the same activities. The classification process is analogous to the one described in the previous section (using the same distance function) although, in this case, we consider the concepts of the involved items and not the items themselves. The similarity between two users u and v based on the concept-level interaction ($\omega_{u,v}$) is made with the previous equation (17), although now the expression $\chi_{u,v}$ refers to the concepts linked to the activities that have been manipulated by both u and v. Therefore, in both equations (17) and (18) the term i refers to an ontology concept instead of an activity.

Afterwards, when a new user enters the system and interacts with some items, the system looks for the cluster that has more similar interactions on the concepts associated to those items. Once the cluster has been found, the average preference and the confidence level on the ontology concept with which users have interacted are measured (equation (2)). The overall idea is that leveraging the classification with ontology concepts (instead of activities as it is done in previous section) in collaborative methods can increase the discovery of new activities that have not been extensively interacted with because they are not located in the most popular locations.

This ontology-based preference assessment is also applied on the groups of users with demographic similarities. In this case, we do not generate a new clustering, but we obtain another preference measure from the demographical clusters already created (section 5.3.2.3). That means that the system finds the most demographically similar cluster and uses the interactions provided by its members to calculate the preference and confidence levels on the lowest level ontology concepts they have interacted with.

5.3.3. Context-aware recommendation

The decision making process of tourists when they are planning a trip is a complex task that is affected by both "*internal*" factors, such as personal motivations, interests and past experiences, and "*external*" factors, like advices, location of the visiting places or even the weather forecast (Swarbrooke and Horner, 2007). *Context-aware recommender systems* can handle these external factors, providing recommendations that fit better with the user's constraints. Whereas traditional recommender systems focus basically on the user's ratings and interests, context-aware recommenders take also into account *contextual* information, such as the user's current location, the available time or the weather conditions (Gavalas and Kenteris, 2011). This contextual information can be gathered explicitly or implicitly. For instance, the user's location can be set by him/her explicitly specifying an area of interest; however, currently, with the extended use of mobile devices, the user's location is widely gathered automatically using their sensors.

Context-aware recommendations can apply two types of filters. *Pre-filter methods* remove from the set of items to be considered by the recommender those ones that do not fit with the user's needs (for instance those tourist activities that are farther than 50km from the user's location). On the other hand, *post-filtering methods* order the items to be recommended depending on their distance to the current location of the user. Although the spatial location of activities is certainly important, we believe that a more extensive modelling of the context of the visitor must be made to provide more accurate recommendations. Therefore, we do not only take into account the user's location, but also his/her travel budget, whether the travel group includes children, the specific dates of the trip, the spoken language and even possible promotions associated to particular activities. In the following subsections we detail how we manage these contextual aspects. For each item and each of these factors we will calculate a score between 0 and 1 that specifies how well each item matches the user's needs with respect to that factor. These scores will be used to calculate the final ranking of the activities, as will be explained in section 5.4.

5.3.3.1. Location

The initial location of the user is obtained through the web form shown in Figure 60, in which he/she will indicate the city to be visited. In the mobile version of the system, the location can be gathered also with the GPS location. During the pre-filtering process, the system filter outs those items that are farther away than the maximum distance the user is willing to move. On the other hand, all items that are within the specified distance will have more relevance as more closer they are to the user. To do so, we measure the score (S) of an item i for the user u given the distance between them ($dist(u,i)$), using the following formula:

$$\left.\begin{array}{ll} S = 1 & \text{if } dist(u,i) < R_{inner} \\ S = 0 & \text{if } dist(u,i) > R_{outer} \\ S = 1 - (dist(u,i) / MAX\_DISTANCE) & \text{if } R_{inner} <= dist(u,i) <= R_{outer} \end{array}\right\} \quad (21)$$

In this expression *MAX_DISTANCE* is a parameter (given by the user) that indicates the maximum distance that he/she is willing to move around, R_{inner} is $0.05 \times MAX\_DISTANCE$ and R_{outer} is $0.9 \times MAX\_DISTANCE$. The R_{inner} threshold has been defined to avoid giving a high score only to those items that are exactly at the city centre. For instance, if the maximum distance is set to 50 Kms., all the items within a radius of 2,5 Kms. will have the maximum score.

5.3.3.2. Travel budget

The travel budget is a parameter specified by the user at the initial form (Figure 60) that estimates qualitatively the amount of money the user intends to spend on the travel. This level is set with a slide bar (Figure 60) that gives values from 0 to 1, giving 0 to a low cost trip and 1 to a luxury trip. We avoid asking the user about the precise amount of money he/she intends to spend, because this question could be rude and, moreover, we should specify more clearly the aspects to be included in the cost (travel, accommodation, food, cultural and leisure activities, etc.). With the slide bar we can distinguish between visitors that are not very much worried about the price of the activities (those with high budget values) and those that will prefer to do free or cheap activities (those with low budget values).

The system stores the price in € of each activity. This quantity is used to calculate the *price level* of each activity, which is 0 if it is free, 1 if the price is over the *max price*, and otherwise the ratio between its price and *max price*. *Max price* is the maximum price of the chosen items in the database or a maximum value of 50 €. In order to obtain a *budget score S* for each item we compute the inverse of the difference between the *travel budget level* and the item *price level* as follows:

$$S = 1.0 - (abs(price\_level - travel\_budget\_level)) \quad (22)$$

With this approach, users travelling at low cost will have higher possibilities to receive free items since these items will obtain a higher budget score. On the other hand, luxury travellers will tend to receive costly items, since they are the ones that can afford them. This score is only used in a post-filtering stage, since we do not want to filter out completely those activities that do not fit with the general financial level of the visitor (a low-cost traveller could decide to make an exception and visit an expensive activity, whereas affluent visitors could certainly enjoy free activities).

5.3.3.3. Travelling with kids

Kids change the adults' way of life, especially on holiday's time. Kids need to be carefully taken into consideration when planning a trip, since they will be bored (and, therefore, they will disrupt the enjoyment of the trip by their parents) if most of the activities are not oriented towards their consumption.

In our case, some of the activities recommendable for the system have been labelled as "good for kids" (activities specially focused on them or that at least have some children-oriented section). Whenever a tourist travels with kids (indicated by the form in Figure 60), all these items will have a *children score* of S=1, and the remaining items will have a score of 0. In the case of tourists without kids, all the items will have S=0 in this score.

5.3.3.4. Trip dates

Tourists are normally rather restricted on their travel dates, due to the constraints imposed by jobs, means of transport, dates suitable to all the group members, etc. Therefore, the dates of the trip are heavily taken into account in the recommendation, especially in the case of temporal events. It is certainly not useful to suggest events that are not available during the trip interval. However, it may be stimulating to suggest specific events that are only a few days before or after the planned dates, giving the visitors an opportunity to modify their travel dates to enjoy them. Hence, we apply a pre-filtering step to filter out those events that are more than five days before or later the trip dates. In the post-filtering phase the system computes a *temporal score* for all the activities, which is 1 (if they are within the trip dates) or the inverse of the division between the number of days outside the trip divided by 5 (the maximum temporal window considered by the system).

5.3.3.5. Spoken language

The language spoken by the tourists is also a handicap whenever they are visiting some places, especially those that need some explanation such as history museums or monuments. Therefore, for all items that have guided explanations, we have specified in which languages they are available. It seems reasonable to suggest activities that users will understand before those that they will not. All items that include explanations have a *language score* (1 if the guide is available in the user's language, 0 otherwise).

5.3.3.6. Promoted items

Destination management organisations are usually worried about the spatial control of the flow of visitors, which usually tend to visit the central areas and the most popular places, overcrowding them. The SigTur/e-Destination system gives the opportunity to these entities to promote quality items that for some reason are not well known (for instance, a new museum, a temporal exhibition or a unique fair). For each item, they can set a value between 0 and 1 to specify the *degree of promotion* of a particular item (the higher the score, the more promoted it will be).

5.4. Aggregation of several satisfaction criteria

In the previous sections we have presented several methodologies that can be used to discover which activities match better with the user's interests and needs. From them, we can construct different criteria that can help the system to make the final recommendation. The first criterion is obtained from the analysis of the semantic content of the activities, taking into account the motivations of the current user and his/her interaction with the system. A second criterion comes from the use of several collaborative recommendation techniques, mainly based on previous users that have some similarities. These two criteria intend to measure the interests of the user on the different activities that are available, however, whenever visiting a place some other factors of the trip may affect the final decision to visit one place or another, like its location, budget or the language used. These other contextual criteria must be included in the recommendation algorithm in order to decide suggesting one activity over another. In this section we propose a hybrid recommendation method that integrates all those criteria in order to find the alternatives that best satisfy all preferences and context factors. In this section, we first show how we aggregate the information about preferences on the tags of the activities (content-based and collaborative) and then, how these preference scores can be used in a multi-criteria recommender system together with the context information.

5.4.1. A content-based semantic preference criterion

A first element that the system will use to make the recommendation is the evaluation of the activities on the basis of their descriptive tags, using the knowledge stored in the ontology. Although the use of ontologies in recommender systems is quite new, we can find different approaches in the literature (see section 3.3). There are different ways of representing the user profile in an ontology, as it has been presented in section 3.2.2.1. Depending on how we store the preferences in the profile, the techniques to evaluate the suitability of the alternatives are different.

In the usual approach the user profile is represented as a vector of features that contains the degree of interest of the user in each concept. In this case, each feature can be interpreted as a different partial criterion that can be used to evaluate an alternative. The goal is then to calculate an overall interest score for a certain alternative. The simplest approach consists in using an aggregation operator to combine the user ratings on the concepts that define a certain alternative. For instance if the alternative corresponds to a museum associated to the concepts {'Archeology', 'family', 'Roman Empire'} the ratings of the user for these 3 concepts are obtained from his profile and are aggregated. The most usual aggregation operator is the arithmetic average (f.i. (Sendhilkumar and Geetha, 2008; Garcia et al., 2011; Yang et al., 2010)).

If there is some additional information on the preferences, the average can be calculated with weights. For example, in (Codina and Ceccaroni, 2010) the confidence levels associated to the ratings are used as weights. In (Hagen et al., 2005) the authors also consider a membership degree of the alternatives to the different categories as a weight associated to the features. In this model, alternatives are instances of more than one class of the ontology and each instantiation has its own membership degree. Synonymous terms are also considered in the aggregation.

Some authors select an optimistic (or a pessimistic) approach to aggregate the partial ratings. This can be done by taking the maximum (or minimum) of the values (Sieg et al., 2007) or by using the summatory (or product) (García-Crespo, 2009). In this case, for each particular application the degree of simultaneity and replaceability of the aggregation can be modeled using linguistic quantifiers (e.g. "most of the features are fulfilled", "at least half of the features are fulfilled").

Another possibility, as discussed deeply in (Cantador, 2008), is to employ the classical voting rules defined in the *Social Choice* field. They propose the use of these techniques to find the global profile of a group of users. Basically, they identify two different approaches: the combination of the individual preferences of the members of the group, and the combination of the ranked item lists obtained from the recommendations derived from personal profiles. In both cases, well-known voting rules such as *Borda Count*, *Plurality Rule* or *Approval Voting* can be applied (García-Lapresta et al., 2010).

A more sophisticated multi-criteria decision making method called *Analytical Hierarchical Process* (AHP) (Saaty, 1980) has also been applied to aggregate semantic information. AHP has four stages: (1) construct a decision matrix including the value of each criterion for each alternative; (2) construct a pairwise comparison matrix of the criteria; (3) derive the relative weight of the criteria from the comparison pairwise matrix; and (4) compute the rank of each alternative based on the derived relative weight. Ontologies can be used in the first stage as in (Huang and Bian, 2009). In this approach the value of an alternative depends on the estimation of the preferred activities of the user, which is stored in the ontology-based user profile. In (Niaraki and Kim, 2009) ontologies are used to define a family of criteria and sub-criteria. The aim is to obtain several criteria for natural disaster modelling based on an ontology-driven architecture, and to combine these criteria together in a unique function using an ANP method (the *Analytical Network Process* is an extension of AHP that does not assume independence among the criteria).

Some recommender systems annotate semantically each alternative with a subset of concepts of the ontology, which are treated as descriptive keywords. Similarly the users are also associated to a list of concepts that define the type of things they are interested in. For example in (Lamsfus et al., 2010) the classes of the ontology define archetypes of tourists, like cultural, sportive or adventurous. In this model, similarity measures are used to calculate the matching between the user profile and the profile of an alternative. A typical measure is the cosine similarity between the two vectors (Lamsfus et al., 2010; Jiang and Tan, 2009; Sendhilkumar and Geetha, 2008; Bhatt et al., 2009). A correlation measure has also been applied to measure similarity in (Albadvi and Shahbazi, 2009) or (Middleton et al., 2009). The work reported in (Shoval et al., 2008) proposes a set of rules to measure the similarity between the two vectors, distinguishing *perfect match* if the same concept appears both in the user and item profiles, *close match* if the concept in the user's profile is more general than the one in the item's profile by one level (his parent) or viceversa, and *weak match* if there is a two-levels difference. Ontology-based semantic similarity measures are also used in (García-Crespo et al., 2009). This kind of functions have been defined in the field of Computational Linguistics and permit to compare two terms from a conceptual point of view by exploiting the taxonomical and semantic relations represented in the ontology. In (García-Crespo et al., 2009) a feature-based similarity algorithm is applied, using several ontologies as reference.

The similarity-based multi-criteria decision aid method TOPSIS has been also used in semantic recommender systems. It is based on the principle that the ideal solution should have the maximum similarity to the best possible solution and the minimum similarity to the worst one. The best solution would be the one with the best performance value on each criterion, and the worst solution would be the one with the worst performance value on each criterion (i.e. the combination of all the worst ratings). In (Zheng, 2011), the recommendation is done on the basis of two scores of the alternatives. The first one measures the cosine similarity on the semantic annotations of the user and the alternative, whereas the second one is given by the TOPSIS method, which is used to calculate an overall utility value for each alternative with respect to its characteristics (not including the ontology concepts). The two values are aggregated with a weighted average. Another approach, proposed in (Yang et al., 2010), filters the set of alternatives using the ontology. An expert system based on these rules is implemented (using standard inference methods). The rules compare the user profile and the description of the alternatives, based on the semantic information stored in the ontology. Then, the alternatives found with the rules are ranked using TOPSIS by analyzing their characteristics (not including the semantic information of the ontology). Notice that, in both cases, the knowledge provided by the ontology is not integrated in the TOPSIS method, but used in a separate stage of the process.

In this work, as it has been presented before, for each user we store a membership degree μ_c that may be understood as the user satisfaction degree with the concept c. In addition we have a confidence level *CL* on the satisfaction estimated value. The satisfaction degree and confidence of each concept are calculated as explained in section 5.3.1, using:

1) The user's motivations (Figure 59).

2) The actions of the user over the activities shown by the system (Figure 63, Figure 64 and Figure 65).

3) The explicit rating of the activities that have been visited (Figure 71).

After calculating these scores for the concepts in the ontology, we take each activity of the database, we find the scores (and *CL*) of its tags and calculate the aggregated suitability score using the OWA for those concepts with a minimum *CL* of 0.2. As it has been said, by defining an appropriate weighting vector, we can establish different aggregation policies (from conjunctive to disjunctive). In this case we also use a linguistic quantifier to define the weights (eq. (15)) with a value of α=2. For a certain activity with *t* tags we calculate:

$$S = \mathrm{OWA}(S_1,...,S_t) = \sum_{j=1}^{t} w_j b_j \text{ and } CL = \mathrm{OWA}(CL_1,...,CL_t) = \sum_{j=1}^{t} w_j b_j$$

, where S_i is the score of the i^{th} concept of this activity, S_i is the confidence of the i^{th} concept and b_j is always the j^{th} largest S_i.

For example, imagine a biking route tagged with the following concepts and their corresponding scores: *Biking* (*S*=1.0, *CL*=1.0), *RuralRoutes* (*S*=0.7, *CL*=0.4) and *CultureRoutes* (*S*=0.6, *CL*=0.7). Since the three concepts have a CL higher than 0.2, all of them will be used to the aggregation. Hence, using the formula (15) the weighting vector (*W*) with α=2 is: [0.111, 0.333, 0.556]. The vector of scores *b* is created in descend order with each corresponding *S*, which results as: [1.0, 0.7, 0.6]. Measuring eq. (13) with *W* and *b* results to a final score (*S*) of 0.67 for the current activity. In order to calculate the final *CL* we also measure the OWA operator, however, in this case the order of *b* is given by the related score, which is: [1.0, 0.4, 0.7]. The final result of *CL* for the activity is 0.63.

5.4.2. A collaborative preference criterion

In SigTur/E-Destination the following preference scores and confidence values can be obtained based on each collaborative information method (section 5.3.2):

1) The preferences about ***ontology concepts*** given by the similarity between the user and the predefined ***tourist segments*** (section 5.3.2.2).
2) The preferences about ***activities*** given by the similarity between the user and clusters of users with similar ***demographic characteristics*** (section 5.3.2.3).
3) The preferences about ***activities*** given by the similarity between the user and clusters of users with similar ***interactions on activities*** (section 5.3.2.4).
4) The preferences about ***ontology concepts*** given by the similarity between the user and clusters of users with similar ***demographic characteristics*** (classification of users explained in section 5.3.2.3 and preference values obtained as described at the end of section 5.3.2.5).
5) The preferences about ***ontology concepts*** given by the similarity between the user and clusters of users with similar ***interactions on ontology concepts*** (section 5.3.2.5).

Notice that we have methods that measure interests on ontology concepts (methods 1), 4) and 5)) and methods that measure interests directly on activities (methods 2) and 3)). The aim of the system is to aggregate all this values for each activity, however, it is necessary first to aggregate the interests of ontology concepts from the methods 1), 4) and 5). In these three cases the system calculates a preference score and a confidence level for each concept of the ontology using the information of each cluster. Note that we may not assign a preference score to all the concepts of the ontology, but only to those that appear in the different clusters of users. In order to compute the final preference value (S) and a confidence level (CL) for concepts c of the ontology, we compute the weighted mean of the scores given by each of the available clusters A, as follows:

$$CL(c) = \frac{\sum_{i \in \mathrm{A}} CL_i(c)}{|\mathrm{A}|} \tag{23}$$

$$S(c) = \frac{\sum_{i \in \mathrm{A}} \mu_i(c) CL_i(c)}{\sum_{i \in \mathrm{A}} CL_i(c)} \tag{24}$$

Once obtained the preference values for the available concepts of a certain activity, the system aggregates them, as was made for the content-based method in the previous subsection with the OWA operator of all tagged concepts.

In summary, we have obtained an OWA aggregated value for a particular item given by the values on concepts in methods 1), 4) and 5). However, we still have the interest values directly to activities given by the methods 2) and 3). The final preference for each activity is measured with a weighted mean of these three indicators (OWA aggregation and methods 2) and 3)) with the same equations (23) and (24) presented in this subsection, where *A* are these three indicators and *c* is the activity to measure.

As an example, let us consider the same biking route given in last section (labelled with concepts *Biking*, *RuralRoutes* and *CultureRoutes*), to measure its preferences for a French family group older than 35 years old that have their own home. First, the most similar segment group is found, which is the prototype 74 (third row in Table 15). From the available list of prototype preferences (columns *Type of activities done* from Table 15), *Sports* is the only one that is ancestor (in the ontology hierarchy) of the item's labelled concepts. In fact, the preference values for *Sports* for this user given by prototype 74 is *S*=0.85 with a *CL*=0.98 that will be downwards propagated (section 3.4.1.2 with α=0.15) to the only one affected descendent concept *Biking*, with the following results: *S*=0.85 and *CL*=0.53.

Thereafter, the system looks for the cluster of users with similar demographic characteristics and obtains the preference of concepts that those users has interacted with. For such concepts we obtain in this method the following results: *Biking* (*S*=0.84, *CL*=0.67), *RuralRoutes* (*S*=0.72, *CL*=0.56) and *CultureRoutes* (*S*=0.45, *CL*=0.43).

Finally, the system looks for the cluster of users with similarities in concept interactions. The results for the tagged concepts obtained from this cluster are: *Biking* (*S*=0.91, *CL*=0.78), *RuralRoutes* (*S*=0.78, *CL*=0.72) and *CultureRoutes* (*S*=0.32, *CL*=0.56). Thereafter, those indicators have to be aggregated with eq. (23) and (24) which gives the following results:

$$CL(biking) = \frac{0.53 + 0.67 + 0.78}{3} = 0.66$$

$$S(biking) = \frac{0.85 \times 0.53 + 0.84 \times 0.67 + 0.91 \times 0.78}{0.53 + 0.67 + 0.78} = 0.87$$

$$CL(RuralRoutes) = \frac{0.56 + 0.72}{2} = 0.64$$

$$S(RuralRoutes) = \frac{0.72 \times 0.56 + 0.78 \times 0.72}{0.56 + 0.72} = 0.754$$

$$CL(CultureRoutes) = \frac{0.43 + 0.56}{2} = 0.495$$

$$S(CultureRoutes) = \frac{0.45 \times 0.43 + 0.32 \times 0.56}{0.45 + 0.56} = 0.369$$

Once obtained the preference for the concepts of the item, the system aggregates them with the OWA operator where the vector *b* with the preference scores *S* is [0.87, 0.754, 0.369] and with *CL* values is [0.66, 0.64, 0.495]. Then the OWA operator gives *S*=0.55 and *CL*=0.56 for the "biking route" as a result of the collaborative methods based in concepts (methods 1), 4) and 5)) in this example.

Finally, the system aggregates with the OWA operator two other preference values directly to the activity given by the methods 2) and 3). First, the most similar cluster based on demographic attributes gives *S*=0.8 and *CL*=0.58. Then, the most similar cluster based on actions on items gives S=0.87 and CL=0.67. The final result for the activity *a* aggregating the OWA operator and the methods 2) and 3) is:

$$CL(a) = \frac{0.56 + 0.58 + 0.67}{3} = 0.6$$

$$S(a) = \frac{0.55 \times 0.56 + 0.8 \times 0.58 + 0.87 \times 0.67}{0.56 + 0.58 + 0.67} = 0.75$$

5.4.3. A new approach to hybrid recommendation using the ELECTRE outranking decision aiding method

Although the semantic content of the activities and the collaborative evaluation are two very important criteria, the final decision of the tourist is usually very much influenced by some context factors (section 5.3.3). In SigTur/E-Destination recommender system we consider up to 8 criteria to make the final proposal of activities to each user: Content similarity, Collaborative similarity, location, budget, kids, calendar, language and promotion.

These 8 criteria have quite different meaning and role with respect to the selection of the most appropriate alternatives. For each of them, we have explained how to calculate a suitability score *S* that has to be maximized in the activities recommended. However, performing a simple average of the scores among all variables would not be very appropriate, as they are considering different dimensions of the recommendation problem. For this reason, we have studied the application of advanced Multiple Criteria Decision Aiding (MCDA) techniques (Figueira et al., 2005). A multi-criteria approach is interesting for recommender systems aimed at finding the most suitable alternatives for each user. This is recognized in the recent literature about multi-criteria recommender systems that consider several different descriptors that have to be aggregated, as can be seen in these surveys (Adomavicius et al., 2011; Shambour and Lu, 2011; Valls et al., 2013).

The main objective of this stage is to rank the alternatives according to the 8 suitability criteria. There are two main approaches to ranking in

MCDA: utility-based methods and outranking methods. In the utility-based approach, an aggregation function is defined to merge the score given by each criterion into an overall score. Then, using this overall score the alternatives can be ranked from the highest to the lowest score. This approach is based on the unanimity and the dominance principles. A large literature on aggregation operators can be found, each one with its own mathematical conditions (Torra and Narukaw, 2007).

On the contrary, outranking methods build a binary relation $a\delta b$ that means: "*a is at least as good as b*". The credibility on this relation is calculated using voting-like techniques (inspired in Social Choice models). In that way, each criterion g_i is "voting" in favour or against $a\delta b$, depending on the performance of a and b in g_i. The underlying idea of introducing the outranking methods is to copy the natural decision making procedures of the people, thus avoiding some strong mathematical hypotheses of the aggregation operators in utility methods. Outranking methods are characterized by the limited degree to which a disadvantage on a particular criterion may be compensated by advantages on other criteria in comparison to utility methods that allow trade-offs between criteria (Pirlot, 1997). Therefore, the outranking approach is a generalization of the dominance relation. However, the outranking relation is richer because the unanimity property and the dominance relation are weakened, so that not all criteria must be in favour of $a\delta b$, to establish this relation as certain, but only sufficient evidence is required.

Although outranking MCDA methods have been successfully applied in many decision making problems, they have not been exploited in recommender systems yet (Valls et al., 2013).

In this thesis we propose to use the ELECTRE methodology (Roy, 1991), which is one of the two most important outranking methods in MCDA (Figueira et al., 2005). ELECTRE is based on doing a pairwise comparison between alternatives for each criterion. For every pair of alternatives, a outranks b if a outperforms b on enough criteria of sufficient importance, and a is not outperformed by b, by having a significantly inferior performance on any single criterion. This is formalized into two tests:

- Concordance test: measures the strength of the coalition of criteria that support the hypothesis "*a is at least as good as b"*
- Discordance test: measures the strength of evidence provided by some criterion against this hypothesis

Each alternative a is evaluated on a set G of n criteria g_i, $i=1,\ldots,n$. In our case, we have that G is formed by the list of 8 criteria mentioned above. For each of them, we have the score S calculated from different evidence, as explained in the previous sections. This score will be used to compare the alternatives. For the content-based and collaborative criteria, only those scores that have a minimum confidence level of 0.2 will be taken into account in order to assure a minimum certainty of the criteria.

ELECTRE method uses a weight w_i to express the relative importance of criterion g_i. Note that this weight must be interpreted as the voting power of each criterion when evaluating the outranking relation. The higher the weight, the more important is the comparison made in this criterion. Thus, the weights of criteria do not represent substitution rates as in the case of compensatory aggregation operators. The weight has been set with the help of a team of tourism experts with the values specified in Table 16. Although we could consider interesting to let the users specify the importance of each criteria, we think that a good approach would be to implicitly acquire the weight by analysing the user behaviour or with the evaluation of users with the system. We give this issue as a future aspect to take into consideration.

There are several methods within the ELECTRE family (Figueira et al., 2013). We propose to use the one based on pseudo-criteria. A pseudo-criterion permits to model the uncertainty associated to the values using some thresholds. We will consider three thresholds when comparing a pair of alternatives for the *i*-th criterion:

- Indifference (q_i): is a difference beneath which the decision maker is indifferent between two alternatives for the criterion *i*. Alternative *a* is weakly preferred to alternative *b* in terms of criterion *i* if g_i(a) > $g_i(b) + q_i$

- Preference (p_i): is a difference above which the decision maker strongly prefers an alternative over all for the criterion *i*. Alternative *a* is strictly preferred to alternative *b* in terms of criterion *i* if $g_i(a) > g_i(b) + p_i$

- Veto (v_i): blocks the outranking relationship between alternatives for the criterion *i*. Alternative *a* cannot outrank alternative *b* if $g_i(b) \geq g_i(a) + v_i$

The values of these thresholds have been set for each criterion, taking into account the level of uncertainty and veto power that we want to give to each one (see Table 16). For instance, between content-based and collaborative, the former thresholds are more strict than the latter because we consider that the collaborative scores are more uncertain. Veto is only applied to content and location criteria because a bad performance on these two criteria must not be compensated with good performance in the rest. Moreover, kids and language are set to 0 in indifference and 1 to preference thresholds since their range of possible parameters are only limited to 0 and 1 (see section 5.3.3.3 and 5.3.3.5 respectively). Note that the budget parameters are defined as a function of the budget of each user, making a stricter decision when the user chooses a low cost budget.

Table 16. ELECTRE parameters for each criterion

Criterion	Range	Indifference q	Preference p	Veto v	Weight
Content-based	[0..1]	0.15	0.3	0.8	0.4
Collaborative	[0..1]	0.2	0.5	-	0.2
Location	[0..1]	0.1	0.3	0.8	0.4
Budget	[0..1]	0.3 x budget	0.5 x budget	-	0.2
Kids	[0,1]	0	1	-	0.2
Calendar	[0..1]	0.1	0.3	-	0.1
Language	[0,1]	0	1	-	0.2
Promotion	[0..1]	0.1	0.3	-	0.2

Giving the previous thresholds we measure the concordance (c_i) and the discordance index (d_i) for each pair of alternatives *a*, *b* in each criterion, as follows:

$$c_i(a,b) = \begin{cases} 1 & \text{if } g_i(a) \geq g_i(b) - q_i \\ 0 & \text{if } g_i(a) \leq g_i(b) - p_i \\ \dfrac{g_i(a) - g_i(b) + p_i}{p_i - q_i} & \text{otherwise.} \end{cases} \tag{25}$$

$$d_i(a,b) = \begin{cases} 1 & \text{if } g_i(a) \leq g_i(b) - v_i \\ 0 & \text{if } g_i(a) \geq g_i(b) - p_i \\ \dfrac{g_i(b) - g_i(a) - p_i}{v_i - p_i} & \text{otherwise.} \end{cases} \tag{26}$$

Given the previous measures, for each criterion we can compute an overall concordance *conc*(*a*, *b*) that determines the level of the hypothesis "*a is at least as good as b"*:

$$conc(a,b) = \frac{1}{W} \sum_{i=1}^{n} w_i c_i(a,b) \tag{27}$$

Finally, the discordance index is applied to evaluate the final credibility between two alternatives *aδb* with the following formula:

$$a\delta b = \begin{cases} conc(a,b), & \text{if } d_i(a,b) \leq conc(a,b) \forall i \\ conc(a,b) \displaystyle\prod_{d_i(a,b) > conc(a,b)} \frac{1 - d_i(a,b)}{1 - conc(a,b)}, & \text{otherwise} \end{cases} \tag{28}$$

Using the outranking relation *aδb*, we can build a valued graph of credibility, where nodes are the alternatives and the arcs indicate the existence (and degree) of an outranking relation. Figure 62 shows the outranking relations between three alternatives.

ELECTRE methodology defines different exploitation procedures for reaching a decision based on this outranking graph (Figueira et al. 2013). We propose to use a ranking technique known as Net Flow Score (NFS). For each alternative we calculate have two evidences: strength and weakness. The strength of alternative *a* is defined as the sum of the credibility values of the output edges to the node *a*. The weakness of alternative *a* is defined as the sum of the credibility values of the input edges to the node *a*.

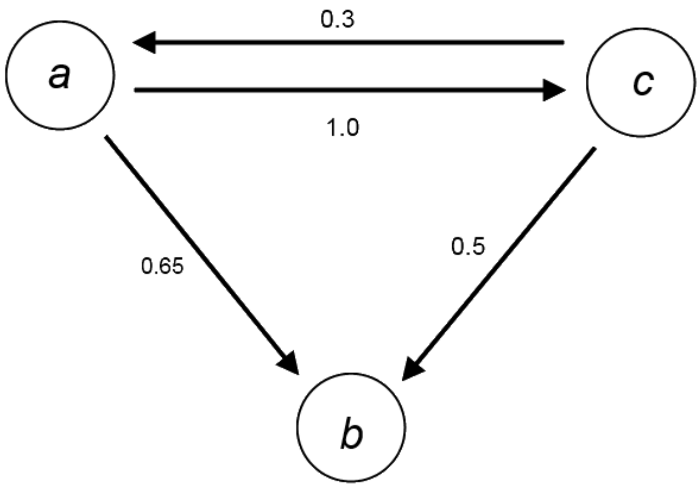


Figure 62. Graph of credibility indexes between alternatives

The Net Flow Score (*NFS*) is defined as the strength minus the weakness. The higher the *NFS*, the better, because the alternative is able to outrank many others and it is only outranked by few ones. The *NFS* value permits to sort the alternatives in descending order. For the example in Figure 62 *NFS* for each alternative results as:

$NFS(a)$ = 1.0+0.65-0.3 = 1.35;

$NFS(b)$ = -0.65-0.5 = -1.15;

$NFS(c)$ = 0.3+0.5-1.0 = -0.2;

Thereafter, using *NFS* we can rank the items. In this example we obtain the following ordered list: *a*, *c*, *b*.

ELECTRE in practise with an example

We have shown how ELECTRE can be used to order items based on several criteria. Now, a complete example with two different user profiles is shown to illustrate this procedure. The first profile is a group of English tourists that travel with kids, who have relevant interest on museums and buildings and some interest on wines, and have indicated a budget level of 70%. The second profile is a Spanish couple with a budget level of 10%, interested in wines and historical buildings. In both cases, they are willing to move at maximum 50 Km from the centre of a city and their main motivation is culture. Imagine we want to suggest the best activity from a set with three options with characteristics shown in Table 17.

The information given in this table describes objectively each of the three alternatives. In this example, each activity is tagged with a single concept (*HistoryMusem*, *WineCellar* and *Building*).

Table 17. List of alternatives to be recommended

Items	Concepts	Distance	Price	Good for kids	Opened	Language guides	Promotion
1	HistoryMuseum	12 Km	4 €	No	All year	CA, ES	0.5
2	WineCellar	5 Km	12 €	Yes	All year	CA, ES, EN, FR	0.5
3	Building	6 Km	Free	No	All year	CA, ES	0.3

The first step is to compare them with the user profile in order to evaluate the suitability of each item for each user. Then, a new table is constructed with the subjective suitability score for each criterion. The results of this step are shown in Table 18 (for the first profile) and Table 19 (second profile). The first two columns (*CB* and *CF*) show the measured score *S* given by the aggregation methods explained in section 5.4.1 and 5.4.2 respectively. The first user has some relevant in the *HistoryMuseum* and *Building* whereas medium interest in *WineCellar*. The second user has a high interest on *Building*, medium interest on *WineCellar*, but really low interest in *HistoryMuseum*. On the other hand, CF method gives high interest in *HistoryMuseum* and *WineCellar*, and low interest on *Building* for the first user. For the second user, CF gives high interest in *HistoryMuseum* and medium interest in *WineCellar* and *Building*. The rest of the columns are the scores of contextual factors evaluated for each profile as described in section 5.3.3. The cost depends on the budget, where items that have some cost are more suited for the user with more budget level. Alternative 2 has been set to score of 1 for tourists that travel with kids. Location, calendar and promotion have no differences among these two users. Both profiles are willing to move the same distance from the same place. Regarding the trip dates factor, the items are opened the whole year, and then it is not affected. Promotion is independent on user characteristics. Finally, for the spoken English user, the criterion language has only one item (number 2) with good score, the one that gives guides in English (EN).

Table 18. Criterion scores evaluated for each alternative (first user profile example)

Items	CB	CF	Location	Cost	Kids	Calendar	Language	Promotion
1	0.75	0.7	0.76	0.633	0	1	0	0.5
2	0.4	0.8	0.9	0.7	1	1	1	0.5
3	0.7	0.1	0.88	0.3	0	1	0	0.3

Table 19. Criterion scores evaluated for each alternative (second user profile example)

Items	CB	CF	Location	Cost	Kids	Calendar	Language	Promotion
1	0.1	0.8	0.76	0.767	0	1	1	0.5
2	0.6	0.5	0.9	0.1	0	1	1	0.5
3	0.9	0.6	0.88	0.9	0	1	1	0.3

After the evaluation of each criterion for each user, the process continues with the calculation of the concordance indices for each pair of alternatives using the thresholds *q* and *p* given in Table 16. The results are shown in Table 20 and Table 21 for each user. Each row corresponds to a different pair of alternatives and each column measures the concordance degree for each criterion (equation (25)). The last column gives the overall concordance value (equation (27)).

Table 20. Pairwise concordance among alternatives (first user profile example)

Comparison	CB	CF	Location	Cost	Kids	Calendar	Language	Promotion	Conc.
1 vs. 2	1	1	0.8	1	0	1	0	1	0.75
2 vs. 1	0	1	1	1	1	1	1	1	0.79
1 vs. 3	1	1	0.9	1	1	1	1	1	0.98
3 vs. 1	1	0	1	0.119	1	1	1	0.5	0.75
2 vs. 3	0	1	1	1	1	1	1	1	0.79
3 vs. 2	1	0	1	0	0	1	0	0.5	0.53

Table 21. Pairwise concordance among alternatives (second user profile example)

Comparison	CB	CF	Location	Cost	Kids	Calendar	Language	Promotion	Conc.
1 vs. 2	0	1	0.8	1	1	1	1	1	0.75
2 vs. 1	1	0.67	1	0	1	1	1	1	0.86
1 vs. 3	0	1	0.9	0	1	1	1	1	0.66
3 vs. 1	1	1	1	1	1	1	1	0.5	0.95
2 vs. 3	0	1	1	0	1	1	1	1	0.68
3 vs. 2	1	1	1	1	1	1	1	0.5	0.95

Looking at both tables we can see how the concordance in cost criterion gives for the second profile better results at items with low or free cost, as shows the concordance of 3 vs. 1 and 3 vs. 2, where 3 is the item with free access; and 1 vs. 2 where item 1 is much cheaper than item 2. For the first profile, items that has some cost have better results. Kids criterion is only relevant for the tourists that travel with kids, giving that the comparison against item 2 (the one with "good for kids" tag) has good score when measuring its concordance against others. Note that when comparing two items that does not offer any attention for kids (items 1 and 3), the concordance is also 1 as there is no difference between them, and hence we can say that both are "as good as" the other in terms of kids criterion. Similarly, the language has no effect on the second profile, since they understand guides of all items, whereas for the first profile, item 2 is better than the others in this term. There are also some differences in the scores obtained by each item in CB and CF criteria, because the motivations of the two profiles are quite different. Promotion, calendar and location have no differences between these two users.

After calculating the concordance, which represents the majority opinion, we have to find the discordant criteria, which are against the majority opinion. Discordance is only possible if this criterion has a concordance of zero for a certain pair of alternatives. In this case the preference and veto thresholds are used (Table 16).

- For the first profile, we find low discordance (0.1) when comparing 2 vs. 1.

- For the second profile, there are two discordance situations: 1 vs. 2 and 1 vs. 3 for CB given by the low preference score. The former results a discordance of 0.4 and the latter of 1.

The next step calculates the credibility index by reducing the overall concordance in the cases of high discordance. For the first profile the discordance has no effect since the value given (0.1) is lower than its concordance (0.79). For the second profile, discordance on 1 vs. 2 (0.4) is also lower than its concordance (0.75) and has no effect. However, in 1 vs. 3 the difference is larger than v=0.8 so the item 1 is vetoed to be ranked first than 2 giving a discordance of 1, that is higher than its concordance (0.66). Then its credibility index is set to 0. In the rest of cases, there is no veto. The credibility values of the outranking relation are shown in Table 22 and Table 23.

Table 22. Outranking matrix for the first user profile

GRAPH	1	2	3
1	-	0.75	0.98
2	0.79	-	0.79
3	0.75	0.53	-

Table 23. Outranking matrix for second user profile

GRAPH	1	2	3
1	-	0.75	0.0
2	0.86	-	0.68
3	0.95	0.95	-

The final step consists in calculating the Net Flow Score for each item from the outranking matrix. The values in each row correspond to output edges, counted as strengths, whereas the columns correspond to input edges, indicating weakness. Then, to compute the NFS for each item, we add the strength minus the weakness, obtaining the following results:

- The English group of tourists give the following values for each item: *NFS*(1)=0.19, *NFS*(2)=0.31 and *NFS*(3)=-0.49. These results suggest that the best item to show for this user is item 2. Although the interest for this item is not as high as other items (the difference is a bit higher than *p* but further from *v*), other contextual parameters positively compensates such difference on the interest, increasing the attractiveness of this item for this group offering activities for kids and the guide with their spoken language.

- The results for the Spanish couple are: NFS(1)=-1.06, NFS(2)=-0.15 and NFS(3)=1.21. Given these results it is clear that the best item is number 3 due to its free cost and they have a low budget but also because of their high interest on it. We have to note that item 1 has a really bad NFS due to the discordance given by its low interest on this item.

Summarising, the system uses the ELECTRE method explained before to calculate a score (*NFS*) for each item in the database that is used as a parameter to sort the alternatives to be suggested for each profile given by several heterogeneous criteria.

5.4.4. Diversity lists

As presented until now, we have a recommendation method that is able to evaluate a set of touristic activities and rank them according to a user profile. The system uses 8 criteria, being the Content and the Location the most relevant ones (with higher weight and veto power). The Content (CB) depends on the scores given to ontology concepts, initially obtained from the travel motivations specified at the initial form (Figure 59). Then, imagine that a tourist sets the motivation beach at 100% of interest, nature and sports at 60% and the rest of motivations with values lower than 30%. With those motivations it would be logical to suggest firstly some beaches, but also some nature and sports activities. However, if the system follows the procedure explained in last section, it may happen that the N first items suggested are only beaches. This is because *indifference q* and *preference p* thresholds specified for CB criterion is 0.15 and 0.3 respectively (see Table 16), the score for the concept beaches will be 1.0 whereas the rest of the concepts will be set with scores lower than 0.7, and hence, when comparing items, those tagged with the concept *Beaches* will also have a higher concordance index respect the others. Only the distance will be able to penalize beaches that are far away from the user destination and then allowing suggesting other items, such as nature or sport activities. This situation is shown in the example provided in Table 24 for a tourist that travels to Salou city, willing to move at most 15 Km with the motivations specified previously. As noticed, there are only beaches on the list, so it is clear that this suggestion will not satisfy the user's expectations. The *Precision* of the list given by NFS is 1 since they all give values higher than 0.7.

Table 24. List of recommendations for a user that visits Salou willing to move 15Km and with the following main motivations: *beaches* (100%), *nature* (60%) *and sports* (60%)

Name	Type	Tags	Distance (Km)	NFS
Platja de Llevant	Beaches	FamilyBeaches, UrbanBeaches, AquaticSports	0.73	1
Cala de Llengüadets	Beaches	Coves, UrbanBeaches	1.31	1
Platja Llarga de Salou	Beaches	NormalBeaches	1.56	0.99
Platja Capellans	Beaches	NormalBeaches, UrbanBeaches,AquaticSports	1.01	0.98
Cala Font	Beaches	Coves, UrbanBeaches	2.74	0.97
Cala Crancs	Beaches	Coves	3.45	0.96
Platja de Ponent	Beaches	NormalBeaches, UrbanBeaches, AquaticSports	1.99	0.95
Platja del Cap de Sant Pere	Beaches	FamilyBeaches, UrbanBeaches	4.72	0.93

If the user wants to discover new places, he could even ask for more items through the pagination of the list. Table 25 shows the second page with new items ranked with the NFS score. In this second list, the user now receives another type of activity (Sailing), however all sport activities are catamaran trips, which will be quite boring for the tourist.

Table 25. Second list of recommendations for a user that visits Salou willing to move 15Km and with the following main motivations: *beaches* (100%), *nature* (60%) *and sports* (60%)

Name	Type	Tags	Distance (Km)	NFS
Platja de VilaFortuny	Beaches	NormalBeaches,UrbanBeaches	5.21	0.92
Excursió amb catamarà a vela. Tram de Cambrils a Salou	Sailing	Sail	0.68	0.91
Excursió amb catamarà a motor pel litoral de la Costa Daurada i Terres de l'Ebre	Sailing	Sail	0.68	0.91
Excursió amb catamarà a motor. Tram de Salou fins a Calafat	Sailing	Sail	0.68	0.91
Excursió amb catamarà a vela. Tram de Cambrils a L'Hospitalet de l'Infant	Sailing	Sail	0.68	0.91
Excursió amb catamarà a motor. Tram de Salou a Cambrils	Sailing	Sail	0.68	0.91
Plata de la Pineda	Beaches	NormalBeaches, UrbanBeaches, AquaticSports	4.51	0.9
Platja de l'esquirol	Beaches	NormalBeaches, UrbanBeaches, AquaticSports	5.92	0.86

Due to the problem mentioned above, it is proposed to use the *quadratic clustering* diversity algorithm detailed in section 4.3.8. This procedure aims to overcome the problem of showing too similar items to the same user, in order to increase his/her satisfaction on the suggested activiteis. Table 26 shows the list of suggestions after the diversity algorithm has been applied. This list firstly suggests three beaches, then four different kinds of sport activities (not only sailing as the previous suggestions did) and a natural space. In this case, the suggestion it is clearly attractive, since it permits to discover different kind of places with still high interest for the user. In fact, the precision still remains in 1 (NFS of suggested items are higher than 0.7), but the diversity (measured as explained in section 4.2) has notably increased from 0.12 (of the previous list) to 0.55.

Table 26. List of activities applying a diversity algorithm

Name	Type	Tags	Distance (Km)	NFS
Platja de Llevant	Beaches	FamilyBeaches, UrbanBeaches, AquaticSports	0.73	1
Cala de Llengüadets	Beaches	Coves, UrbanBeaches	1.31	1
Platja Llarga de Salou	Beaches	NormalBeaches	1.56	0.99
Excursió amb catamarà a motor. Tram de Salou fins a Calafat	Sailing	Sail	0.68	0.91
Karting Salou	Motor Sports	GoKarts	2.46	0.8
Lumine Golf PortAventura	Golf	Golfing	2.61	0.8
Busseig. Cap de Salou	Under Water	Snorkelling	3.12	0.78
Sèquia Major	Natural Spaces	NaturalSpaces	3.75	0.76

5.5. Planning functionalities

This section summarizes the whole user interaction from the point in which the system acquires the initial information about the general user preferences to the point where he/she downloads the desired travel plan to a mobile device.

Whenever the user accesses the web site, he/she is firstly asked to fill up the motivations form (Figure 59), as explained in section 5.2.1. In this step, the user will indicate the level of interest in each motivation. The second step is to provide the user demographic data and the context of the travel (Figure 60). The chosen variables, whose rationale was explained in section 5.2.1, are: country of origin, travel group composition (allowed values shown in Table 12), type of accommodation used (allowed values shown in Table 12), destination, budget level and trip dates. Once the user has set up the preferences and travel information, he/she can proceed to the next step in which the system can start suggesting the first items based on this initial data (following the recommendation process explained in the previous sections). Suggested items are shown in a web page, as illustrated in Figure 63. The list of suggested activities is located at the left panel of the page and it shows for each item its name, a brief description and a thumb image. At the right side of the screen there is a map that geo-locates all these activities with the icon associated to its main type (beaches, shopping, museums, etc). Whenever the user moves the mouse over an activity (either from the list or the map), its border and its icon are highlighted with a magenta colour, which allows relating each activity to its spatial position. The user may move an activity to the travel plan by selecting its checkbox of the list. When an item is added to the plan its related graphical icon turns into a suitcase, thereby providing an easy view of the location of the chosen activities. The list of items is paginated showing *N* items (in this example *N*=6, but this number can be internally configured), allowing the user to navigate to the next page with new *N* items.

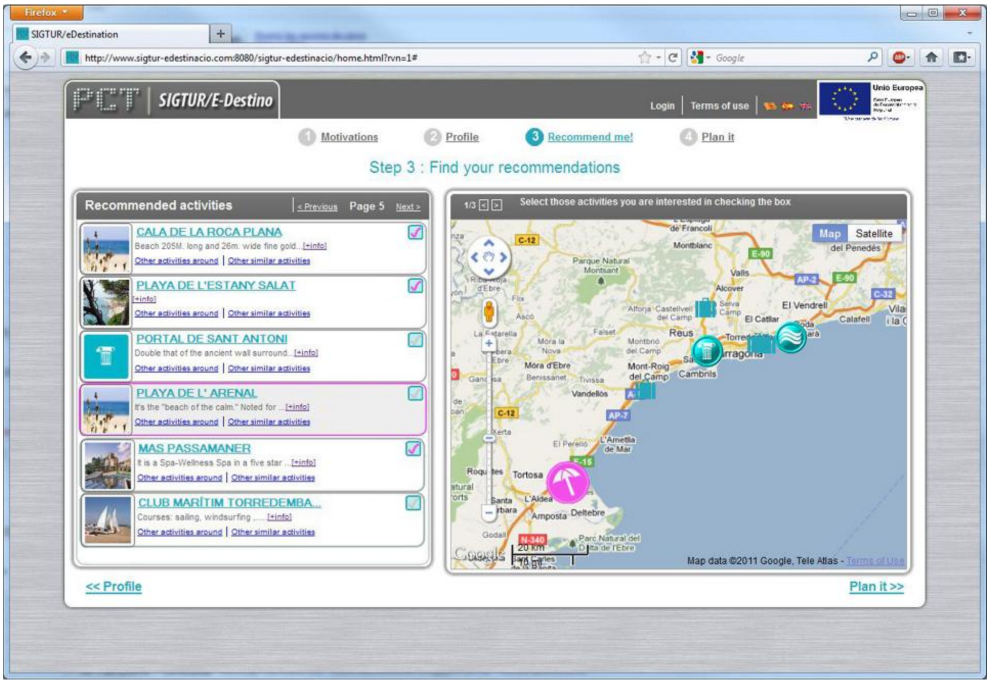


Figure 63. Screenshot of the recommendation pane.

Another action the user can perform is to request more information on a specific activity by clicking on the *+info* button located at the right side of its description. Figure 64 and Figure 65 show pages where the left panel displays detailed information of a particular activity (pictures, complete description, observations or main services). In addition, the user can request more activities that are near the current one (as shown in the map of Figure 64) or that are similar to the current one (as shown in the map of Figure 65).

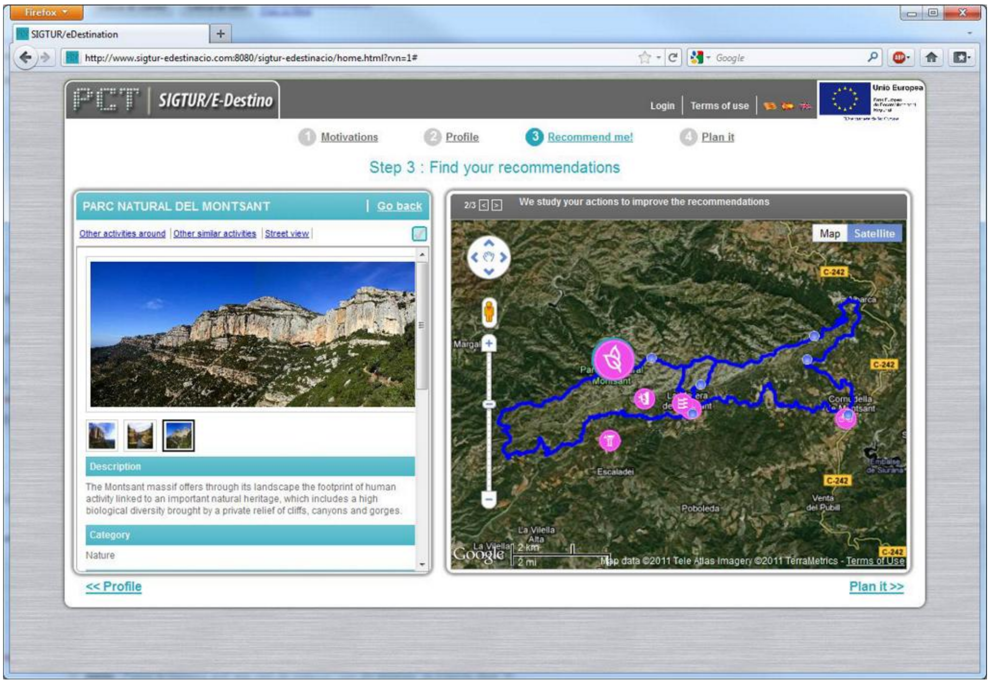


Figure 64. The left panel displays the detailed information of an activity, and on the map it is possible to see the activities near it.

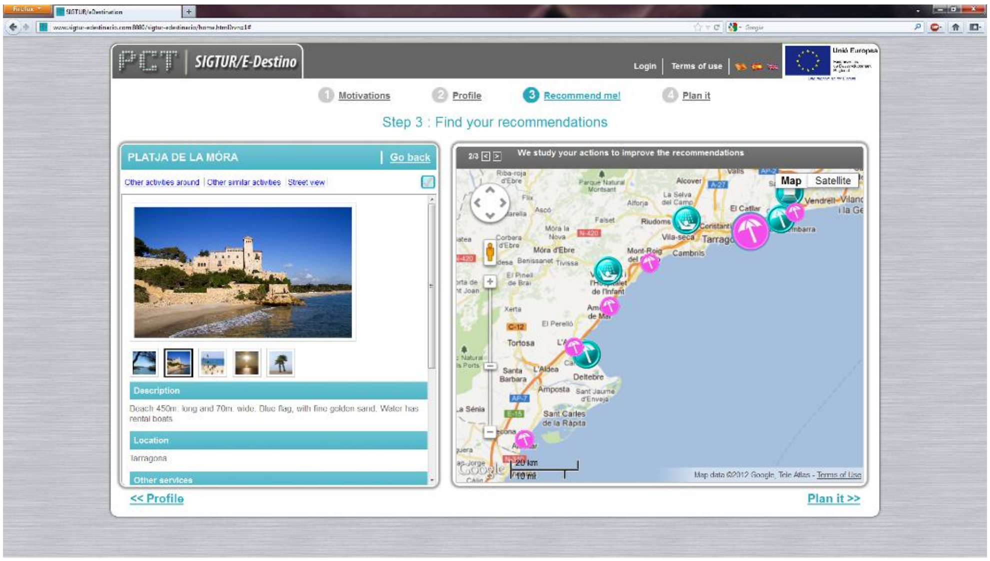


Figure 65. After the user asks for activities similar to a particular one, they are represented on the map with the magenta icons.

Once the user has checked the recommended activities, he/she can proceed to the last step to plan the trip. Figure 66 shows the page where the user can schedule the route with the chosen activities. The user has to drag and drop the cell of the desired item to one of the days of the trip. For each day the visiting order of the activities can be rearranged. Route directions and approximate times to move from one activity to another are also shown on the map. Finally, whenever the user is satisfied with the scheduled plan he/she can download in a PDF file all the information about the activities to visit each day and their driving directions.

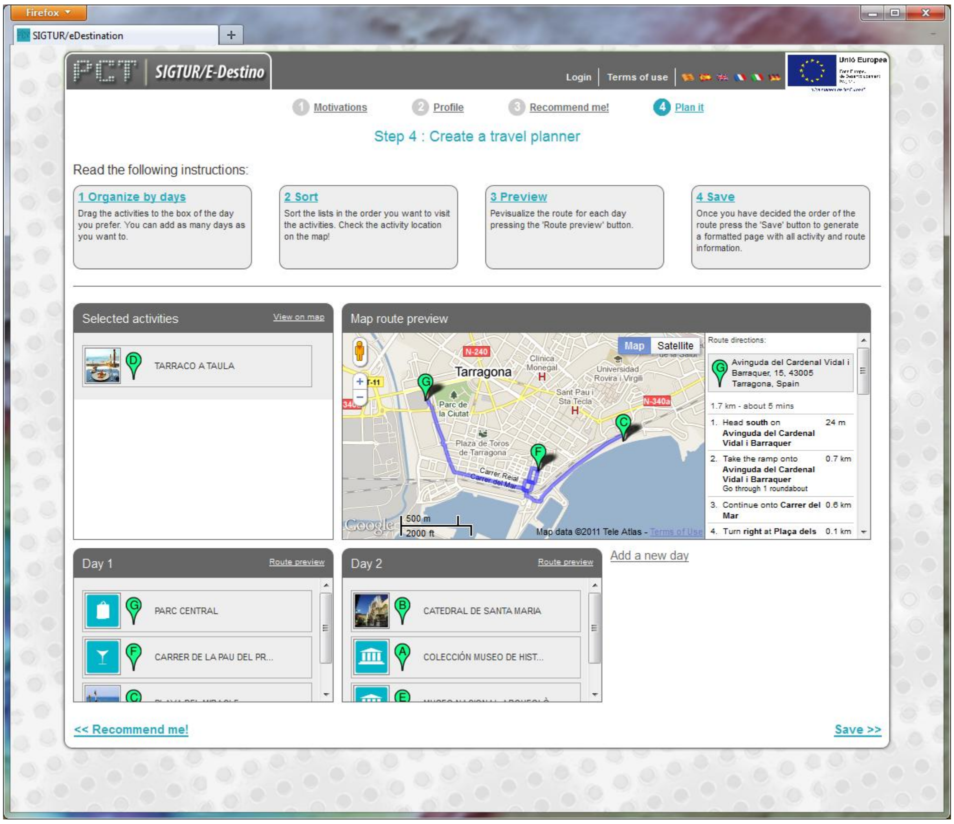
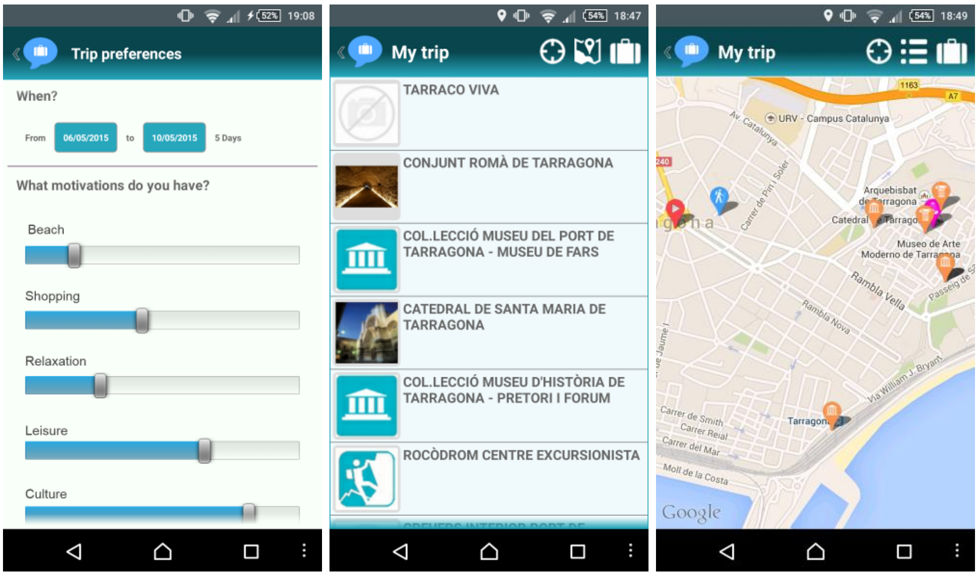


Figure 66. Travel planner.

On the other hand, users with smartphones can download the trips to their devices and either follow or modify the route *in situ*. Native apps have been developed to be run on Android[46] and iOS[47] platforms as a front-end layer that connects to the recommendation engine through the Web server (as explained in section 5.1.1). When using a registered account, user's trips are synchronised, i.e. any plan can be changed either from the Web or the mobile versions and the results will be automatically updated on all platforms. The mobile interface has almost the same functionalities than the website. However, as the size of the screen is highly reduced, all the information cannot be shown at the same time and, therefore, the screen shows only the list or the map of recommendations, or the list or map of the designed trip plan. When opening the app the user has to login or sign up an account, and automatically all his/her trips are downloaded to the mobile phone. In addition, all the content of the activities (only text data) is also downloaded and stored on the mobile phone, and hence the app can be run offline. This is especially useful for tourists that do not enjoy international data roaming or for rural areas without good phone coverage. In this offline mode the user will not have access to pictures, because it would be too time consuming to download them. If the user has not created any trip from the Web site, he/she can create a new one filling up the user profile (Figure 67 (a)). This profile is the same as the one of the Web version, with the exception that the user may select a destination name for the localisation or

[46] http://developer.android.com/ (last access March 2015)
[47] https://developer.apple.com/technologies/ios/ (last access March 2015)

set his/her current location, captured with the GPS. After that, the app connects to the server and sends the profile information, and the server executes the recommendation engine and returns a list of identifiers ranked for such user that the app will print in a list and a map (Figure 67 (b) and (c)).

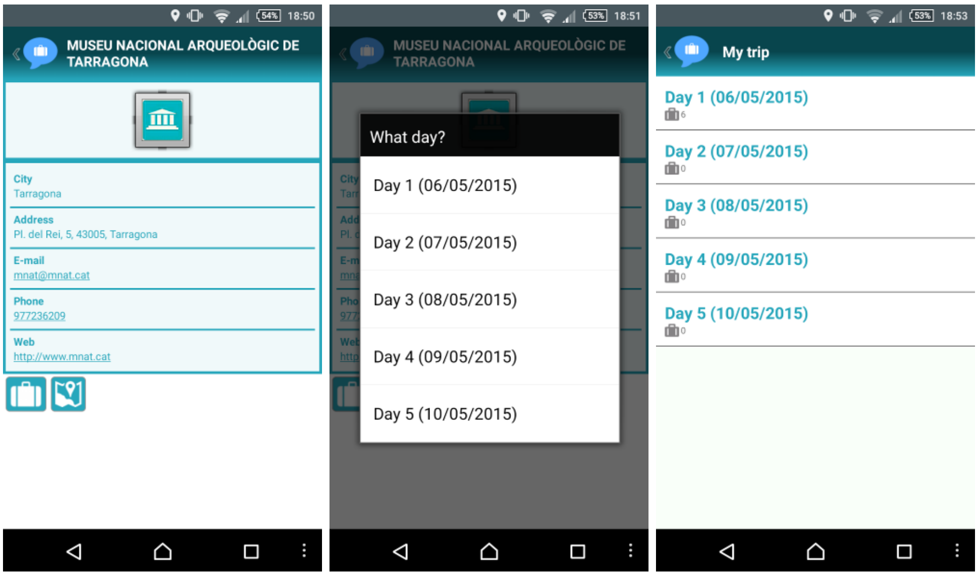


(a) (b) (c)

Figure 67. App screenshots: a) trip preferences; b) list view of suggestions; c) map view of suggestions

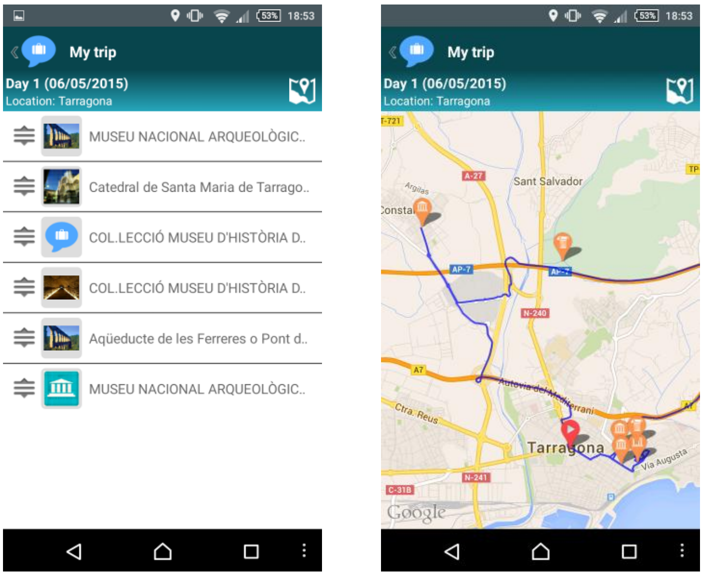
The user can see more information of a particular activity (Figure 68 (a)) and add it to the trip plan by pressing the travel suitcase button and choosing the desired day of the trip (Figure 68 (b)). These user actions are also recorded and sent to the server in order to implicitly learn his/her preferences, as was done in the Web version. The trip plan is displayed daily (see the list of days in Figure 68 (c)) in a list and a map with the route to follow (Figure 69 (a) and (b)) where the user can rearrange the order of the activities by moving them up or down. The user can also choose to receive the driving directions to reach a particular activity (obtained from the Google Maps app). Unlike the Web version, the user session does not finish at any point since he/she can follow the route. For instance, if he/she has more time available to visit more places it can request to the app to append to the route new nearby places.

Whereas the Web version of the system is well suited to prepare a plan, the main advantage of the mobile version is that it can be used whenever the user is already visiting the activities. In this way, users can discover new places near their location or change the route plan if necessary. In addition, the capabilities of modern mobile devices (camera, GPS) open the door to another way of displaying data, *Augmented Reality* (AR). When a user focuses with the camera on some place, the screen can enhance the real image with richer data. For example, Figure 70 shows how an image and the name of the point of interest appear at the screen when the mobile is facing it.



(a) (b) (c)

Figure 68. App screenshots: a) activity information; b) action of adding an activity to a day of the trip; c) list of days of the trip



(a) (b)

Figure 69. App screenshots: a) list of ordered activities planned for one day; b) route of the planned day printed on a map

Figure 70. App AR screen

Three days after the trip the system automatically sends an email to the user with a link to the page where he/she can rate each activity and write a short review about it (Figure 71).

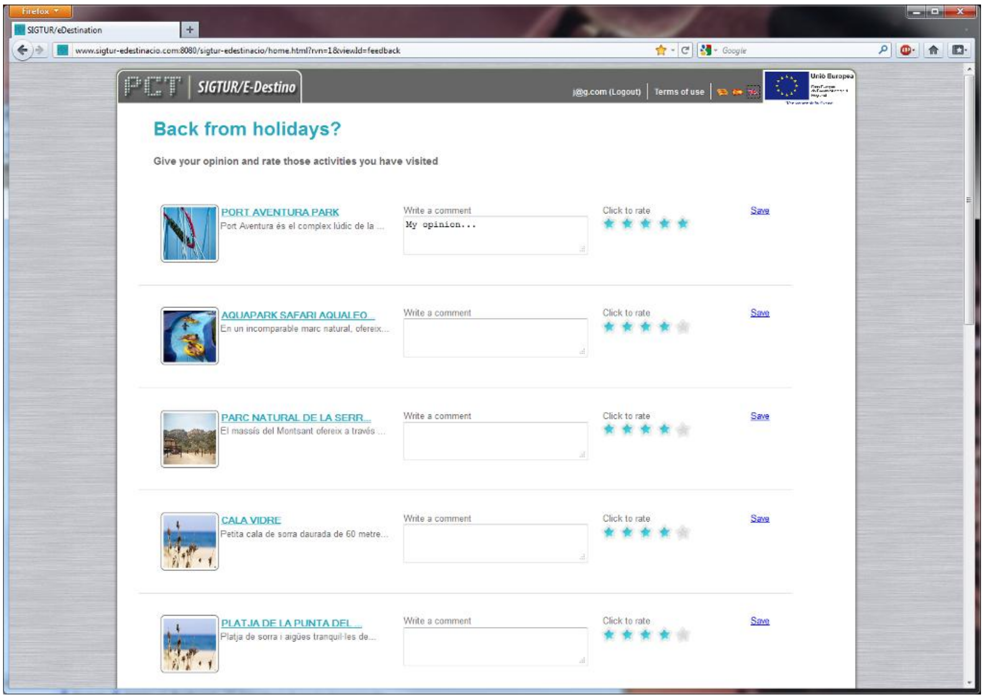


Figure 71. Explicit evaluation and comment of activities the user has visited.

5.6. Conclusions

Recommender systems are important tools in the provision of personalised 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 and mobile-based interfaces of the presented system allow planning activities before and 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, an 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. The framework for managing uncertain preferences explained in chapter 3 was successfully applied in the recommender system. Concerning the employed recommendation techniques, the system considers as much information as possible to provide an accurate recommendation, including demographic and travel data, trip motivations, tourists' segments, the actions of the users on the platform, classes of users with similar tastes or demographic attributes, and last but not least, the context of the visit, such as the location or dates of the trip. All this information has been aggregated with a Multi Criteria Decision Analysis technique (ELECTRE) that permits to rank items based on different heterogeneous variables. In addition, it was demonstrated that applying the cluster quadratic algorithm explained in chapter 4 can increase the attractiveness of the recommendations with diversified alternatives.

Concerning the future work, the main objective is to make the SigTur/eDestination site available online for any user and obtain as much information as possible from their use. Therefore, we will be able to evaluate the recommendations by analysing their behaviour on the system. For instance, we can discover if users choose the first items of the list of recommendations or how many pagination actions they perform. With a thorough analysis of this behaviour, we would be able to accurately adjust the parameters that are applied in each recommendation methodology to satisfy better the expectations of users. Or even better, we could find an algorithm that automatically personalizes the values of the parameters for each user profile, as was done in chapter 4 with the dynamic adaptation of λ based on the user motivations. We will also have to analyse the demographic attributes of the users periodically every year in order to define the number of clusters of users in the demographic method (section 5.3.2.3) because the type of tourists may change overtime (e.g. the number of tourists from emerging countries such as China may grow).

A nice feature of this recommender system is that it can be easily adapted to specific domains or geographical areas, as will be shown in the next chapter. Quantitative and qualitative results of the evaluation of SigTur/e-Destination will be also presented.

Chapter 6 – Test and validation

Chapter 5 explained the design and implementation of SigTur/E-Destination, which was created *ad-hoc* for the purpose of suggesting activities to tourists visiting the Tarragona province; thus, it covers many different kinds of tourism and leisure activities. However, it could be useful to have the possibility to adapt the system so that it can focus on a more specialised set of activities. Therefore, the system has been implemented to be easily adaptable to any other geographical area or domain. That is why the recommender engine and the database have been designed to be easily reusable in different domains, areas or user profiles. Then, using the same source code and structure of the database, it is possible to add new functionalities that can be run only on a particular adaptation. The interface of the system and the communication with the server may be fully reused, just changing their HTML structure, logos and colours.

This genericity has been proved with the adaptation of the base system to the suggestion of eno-touristic activities, as will be described in section 6.1. Another adaptation has also been tested for the specific area of Costa Daurada and Terres de l'Ebre (the coastal area of the province of Tarragona) as detailed in section 6.2. This process has involved a specialisation of the domain ontology, the user profile and the activity types. The following section of this chapter provides two validations of the SigTur/E-destination system (a theoretical one based on the analysis of the recommendations made to stereotyped users, and a practical one based on the assessment of the system by real users). The last section concludes the chapter.

6.1. Eno-SigTur

This section explains the adaptation of the system to the *enotourism* domain, which has led to the development of a new product called *Eno-SigTur*.

6.1.1. Enotourism

The selection of the enotourism domain to adapt the system reflects a fundamental challenge for the competitiveness and sustainability of the tourism offer in the province of Tarragona, which is one of the most important areas of the Spanish coast for "sun and beach" activities. Faced with the growing instability of the sector, this area has taken on the challenge of differentiation and innovation. The managers of the destination opted for the diversification of supply, developing alternative and complementary products which are available in less known landscapes, located outside the overcrowded tourist areas. This is the case of wine tourism (or enotourism) that counts on high quality resources.

The region of Tarragona is characterized by a remarkable specialization in the wine sector. The number of Protected Designations of Origin (PDO) in wine is quite high for a relatively small territory, with 5 own PDO and 3 shared with the neighbouring provinces. 8 of the 10 districts of the province of Tarragona are included in some PDO. Although each region has unique elements, there are assets that add value to the whole territory, as the "Cathedrals of Wine" (modernist monumental cellars) and popular events linked to the symbolic heritage of wine-making, which attempts to increase value through the "Pais del Vi"[48] (Wine Country) brand.

Despite the huge potential of wine tourism, this brand is not widely known by international tourists, which see the province of Tarragona mainly as a seaside destination. For this reason, manager destinations are considering new strategies for boosting tourism in rural areas, both in Tarragona as in the whole of Catalonia (Anton-Clavé, 2009), which have a direct impact on the wine sector.

6.1.2. Enotourism information systems

Recommender systems or planning routes for tourists have mainly focused on major cities (e.g. www.triporg.org, www.citytripplanner.com), so this technology has not reached other tourist areas with the same intensity, as is the case of wine. In this section some applications focused on wine tourism, accessible via the Web or via mobile, are discussed (Table 27 summarizes their main features). It can be seen that most of these applications were developed to make a simple promotion of tourist destinations. They are Web pages with a list of hotels and restaurants, but they do not offer a recommendation service according to the users preferences. Other systems are recommenders of wines, given a certain user profile. Moreover there are not any recommendations of itineraries or suggestions of other activities (not specifically wine-related) that may be also of interest to the tourist.

Most of the applications shown in Table 27 are GIS that show geo-referenced information. Some of them support mobile applications (apps), which display information taking into account the position of the visitor, gathered through GPS. Most mobile applications show a list of resources requested by type (winery, restaurant, etc.) that are near the visitor. Realtur is the only one that displays information using augmented reality.

[48] http://www.paisdelvi.com/ (last access March 2015)

Table 27. Analysis of available Enotourism information systems

Product	Web	App	GIS	Routes	Wineries	Events	Restau-rants Hotels	Local info.
Vins et Tourisme en Bourgogne (vins-tourisme-bourgogne.com)	●	●	●		●	●	●	
Pesquisa de Vinhos y Rotas do Vinho (infovini.com)	●		●	●	●	●		●
Wine Regions of Victoria-VicWineries (visitvictoria.com)	●	●	●	●	●	●	●	●
Finger Lake Wine Country (fingerlakeswinecountry.com)	●	●	●		●	●	●	●
Vin Vaudois (vins-vaudois.com)	●	●	●		●	●	●	●
Realtur [Android app]		●	●		●		●	●
Visit Napa Valley (legendarynapavalley.com)	●	●	●		●	●	●	●
The Wine Hub (thewinehub.com)	●				●	●		●
Vinho Verde (vinhoverde.com)	●		●	●	●		●	●

Apart from the selection of specific activities, tourists are usually interested in planning a travel route (one or several days). Most of the analyzed applications do not allow building a tourist itinerary automatically. However, some of them permit to select activities and create a schedule manually. These tools require the user to check the schedules and availability of visits and to calculate the time needed for visiting each item and driving among them. Other applications only provide pre-defined static routes by type and location. For example the Portuguese *Rotas Do Vinho* has already established several routes, but they cannot be personalized and varied, while *VicWineries* offers the possibility to organize an itinerary choosing from all tourist towns in the Australian region of Victoria.

Finally, we note that some applications can be used to discover wine-related events, such as activities of cultural and creative nature (exhibitions, fairs, etc.). This component is very interesting for visitors, since it allows them to have a richer tourist experience in the region.

6.1.3. Adaptations from SigTur/E-Destination

The adaptation of the recommender system to enotourism aimed to provide a personalized service for planning routes or trips facilitating the discovery of other activities of the territory, for both visitors with low knowledge of the wine geography in the region and also expert wine tourists.

Eno-SigTur can display personalized wine-related information to those visiting the province of Tarragona. This information is focused on wine-related activities (visiting wineries, wine landscapes, wine tasting, etc.) but it also includes other cultural and leisure activities that can complement a wine trip, such as visits to museums, monuments or natural itineraries. It also includes information on accommodation and restaurants.

Aside from the domain adaptation of the system, it has been a good opportunity to improve its design. A home page that can be rapidly configured to any domain with different content and corporate colours has been developed. For *Eno-SigTur*, this page (see Figure 72) explains how to use the system, allows users signing up or logging in, and permits to select the preferred language. The initial form to build the user profile has been changed to one page, with more basic user characteristics as shown in Figure 73. This change has reduced the time needed for the user to provide this information. This reduction of factors has been done for this specific domain, but the system can be easily configured with different user profiles satisfying any needs.

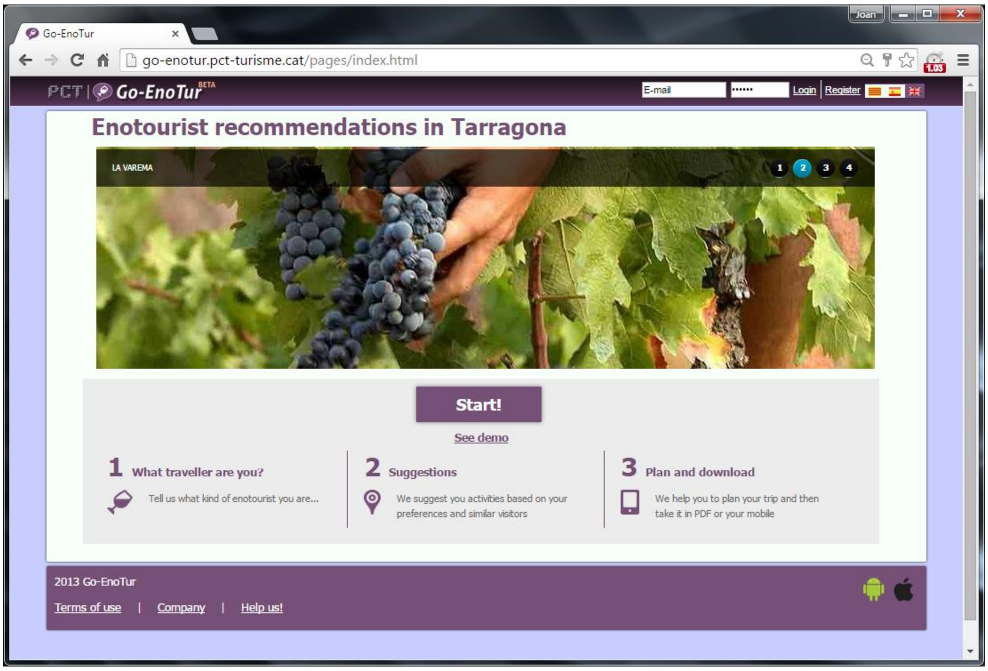


Figure 72. Eno-SigTur home page

Together with wine tourism experts from the Science & Technology Park for Tourism and Leisure it has been decided to reduce the user profile parameters to five travel motivations, country of origin, type and size of the travel group, trip dates and destination. The five motivations chosen are somehow related to wine tourism: *Culture*, *Nature*, *Sports*, *Health and Care*, and *Leisure and Entertainment*. The traveller group has been simplified to more generic profiles: alone, with family or friends, with couple, senior group and business. The user may select if they travel with kids or not, which allows to combine any group type with the kids option. In this domain, some activities are related to the visit of wineries or museums in rural areas that in many cases are small businesses in which the person that makes the wine is the same person that guides the touristic visits. Whereas some big wineries are ready to accept large groups of tourists (for instance 30 people) and they have professional guides, the small ones are more focused on small groups such as couples and they offer a more authentic visit with guides that are working daily in the process of wine making. That is why we have added a new parameter to allow users to specify the number of adults travelling with the group, which is used by the system to personalize the suggestions taking into account this factor.

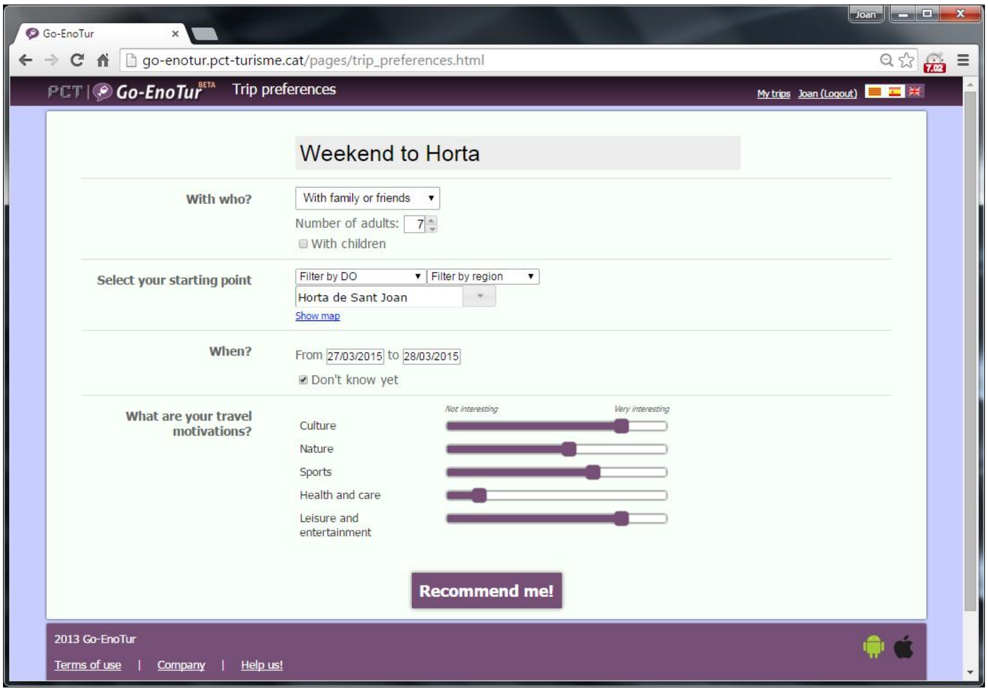


Figure 73. Eno-SigTur initial form

In *SigTur/E-Destination* the user chooses the wished items and then, in a new web page, he/she can organize the route for each day. In *Eno-SigTur* we have enhanced the functionalities of the system by allowing users to receive and create the plan at the same time, which permits to discover new resources near the planned route. As can be seen in Figure 74, the web site shows the list of suggestions at the top with four classification tabs (enotourism, other activities, where to eat?, and where to sleep?), the list of planned activities for one day at the right side, and the map with the suggestions and the planned route. This allows the user to discover new activities that come across the planned route, hence enriching the touristic driving path with interesting places to see or visit.

The user can directly transfer the preferred activities with the drag and drop action to the day panel of the trip plan. He/she can move each item to change the order in which places should be visited during the route. On the other hand, the system can be asked to sort the chosen items to create an optimal path (in terms of distance among all the items). We have used the Google directions API[49] to order and display the driving route that reaches the selected items. The approximate travel times and visiting times are also shown to the user, so that he/she can plan accurately the visit. When the user selects a place to sleep, then the route starts at the location of the chosen accommodation. It is also important to note that some wineries have special opening and closing dates, especially those that are small enterprises, where

[49] https://developers.google.com/maps/documentation/directions/ (last access March 2015)

they normally only open during weekends. To avoid planning a place on a certain date in which it may be closed, the system shows an appropriate warning to the user. Finally, as in *SigTur/E-Destination* whenever the user is satisfied with the planned route he/she can download a PDF file with the details of each activity to visit and the driving directions to follow the route or use the app version of the system that automatically downloads the trips created on the website.

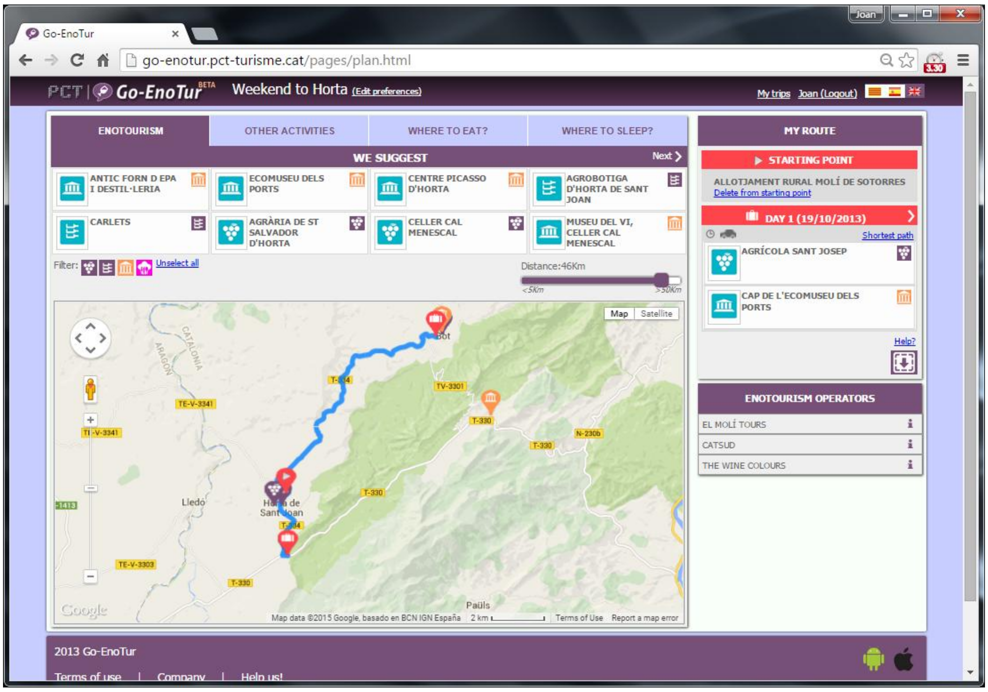


Figure 74. Eno-SigTur interface: list and map of suggested activities and planned route for each day of the trip

The mobile apps have also been adapted to the enotourism domain, mainly changing the interface, logos, colours and the variables of the user profile. Figure 75 and Figure 76 show the main screens of Eno-SigTur for the Android and iOS platforms respectively.

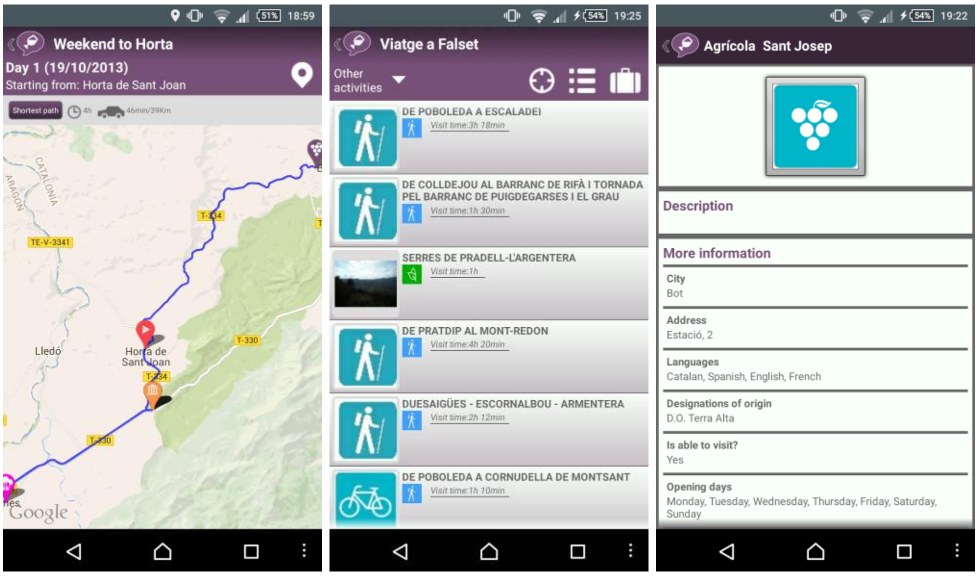


Figure 75. Eno-SigTur app screens of the Android version

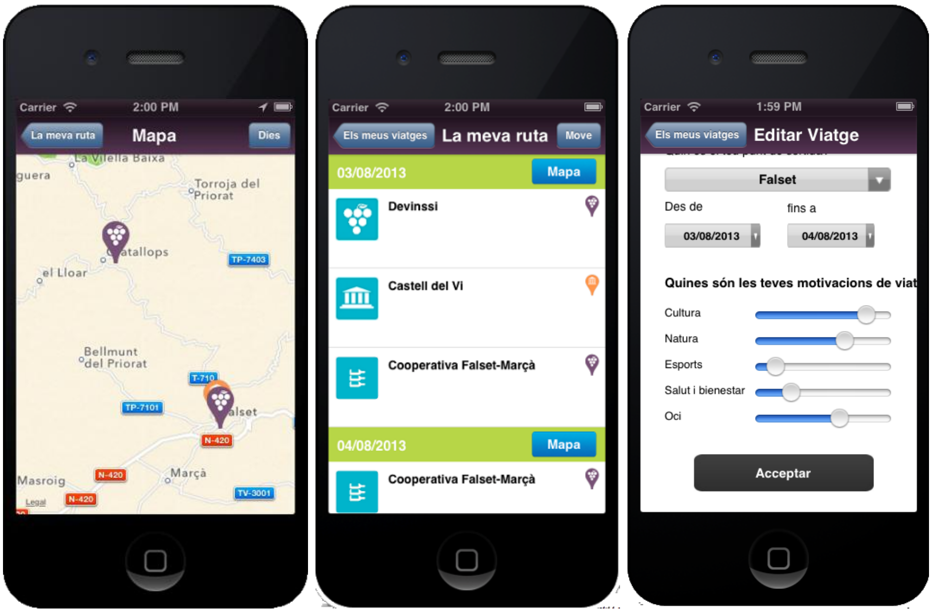


Figure 76. Eno-SigTur app screens of the iOS version

The system has not only been adapted at the front-end, but also at the back-end. First of all, the ontology has been increased with new concepts specialized in enotourism, such as *WineComercial*, *WineTherapy*, *WineFairs*, *ModernistCellars*, *EcoMaking*, etc. Figure 77 shows part of the new ontology focused on wine concepts. The recommender engine uses the same source code as the generic tourism recommender system explained in previous sections. However, it is able to use this new enotouristic ontology or the more generic ontology, depending on the way in which the user has accessed the system.

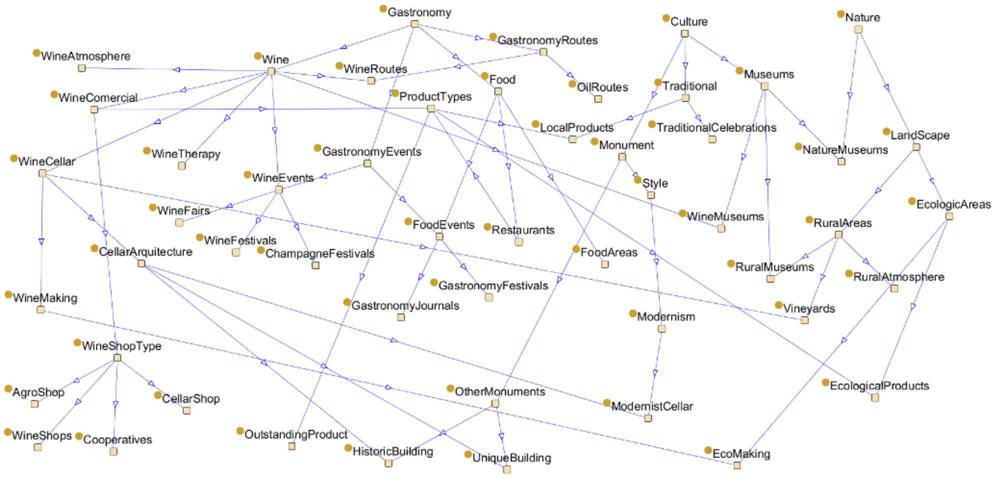


Figure 77. Part of the extended enotourism ontology

Some recommendation methods had to be adapted to the changes in the user profile. For instance, when measuring similarities between users (section 5.3.2.1), the accommodation type criterion was suppressed. The system adapts the similarity measure depending on the domain to include the corresponding parameters.

On the contextual part of the user profile, EnoSigTur does not consider the travel budget in the multi-criteria decision process. However, it includes as a new parameter the size of the group. This factor is used in pre-filtering and post-filtering processes. Pre-filtering is applied for big groups of tourists to filter out those places that do not accept such big groups. On the other hand, in the post-filtering process the activities that only accept small groups are ranked first for this kind of visitors. To do so, the group size factor has been included as a new factor in the MCDA process (section 5.4.3). In order to obtain the score of the group size factor for each item, we compute the difference of the group size and the maximum acceptable size of the item. This value is then normalized between 0 and 1 among all items. The parameters of this criterion for the ELECTRE method (see Table 16 for the parameters of other criteria) were set to *Indifference* $q = 0.1$, *Preference* $p = 0.3$ and *Weight* $= 0.2$.

The system has been adapted to enotourism, but it could also be adapted to regions with winter sports like skiing, or more nature-focused activities like trekking. The recommendation engine core is the same for any domain and it may adapt both to the user profile (by adding or supressing some characteristics of the user) and to the new types of items, using specific ontologies. Moreover, the front-end layer can be adapted to the needs of the domain with specific functionalities and an interface that accurately represents a region or domain using related pictures, logos and corporate colours.

6.2. Visit Costa Daurada & Terres de l’Ebre

Another adaptation of the system has been made to fulfil the requirements of the official tourism destination management organization of Costa Daurada[50] and Terres de l’Ebre[51] (henceforth “Visit CD & TTE”), which includes the coastal area of the province of Tarragona. The front-end layer has been adapted with the corporate colours and logos, and the graphical design has been improved to increase the attractiveness of the site. Their main requirement is the offer of personalized information on cultural activities on the area based on the user profiles, as SigTur/E-Destination does. However, the user profile has slightly different characteristics, as explained below.

As shown in Figure 78 the user profile to be filled includes the traveller group type (family, couple, friends, alone or business), average age of the group, trip dates, transportation means and 7 motivations (*beach*, *leisure and entertainment*, *nature*, *culture*, *sports*, *enotourism*, and *health and care*). The ontology has also been simplified to include only those concepts associated to activities available in this region.

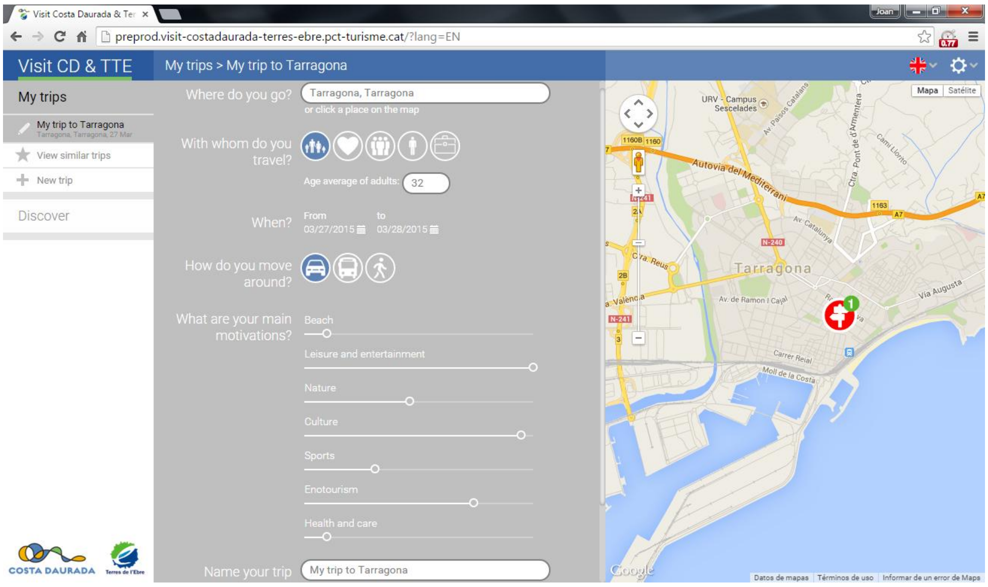


Figure 78. Visit CD & TTE: form to build the user profile with travel motivations and characteristics

Since the system has also been enhanced with more languages (Catalan, English, Spanish, French, Italian, Deutsch and Russian), we use the information about the language of the user (the default language set in the browser) as a proxy for the country of origin. Most of the visitors of Costa

50 http://www.costadaurada.info/ (last access March 2015)
51 http://www.terresdelebre.travel/ (last access March 2015)

Daurada and Terres de l'Ebre have the main nationalities associated to these languages, so the loss of accuracy is very low and we reduce the volume of the initial form.

Another parameter added to the similarity between users (section 5.3.2.1) is the average age of the group of tourists. This new characteristic adds new values (*x* and *w*) to the equation (16). *x* is set to 1 if the users have the same age and to 0 if the difference is over 20 years; otherwise, the value is the age difference divided by 20. The weight *w* of this parameter was empirically set to 0.3, and then all the weights were normalized to add 1.

Another characteristic added to the user profile is the transportation means used by the travellers (walking, car driving and public transport). The benefits of the transport factor are twofold. On the one hand, the times and routes on the map can be customized for any kind of transportation (even public transport can be managed with the Google directions API). On the other hand, we can get rid of the bar to choose the maximum distance the user is willing to move, since we can estimate an approximate value given the chosen transport. Therefore, the system automatically sets the MAX_DISTANCE value (from section 5.3.3.1) depending on the option selected (car: 50 Kms., public transport: 20 Kms., walking: 5 Kms.).

Once the user fills up his/her profile, the system uses the recommendation engine explained in the previous sections to suggest a personalized and diversified list of items. Figure 79 shows the page where the user receives such suggestions. The list of suggestions, which keeps continuously adding new items as the user scrolls down, is shown in the centre. At the right hand side the map shows items that are on the list, and on the left there is a menu that allows the user to switch between his/her trip (at the top) or to focus on a particular type of activities.

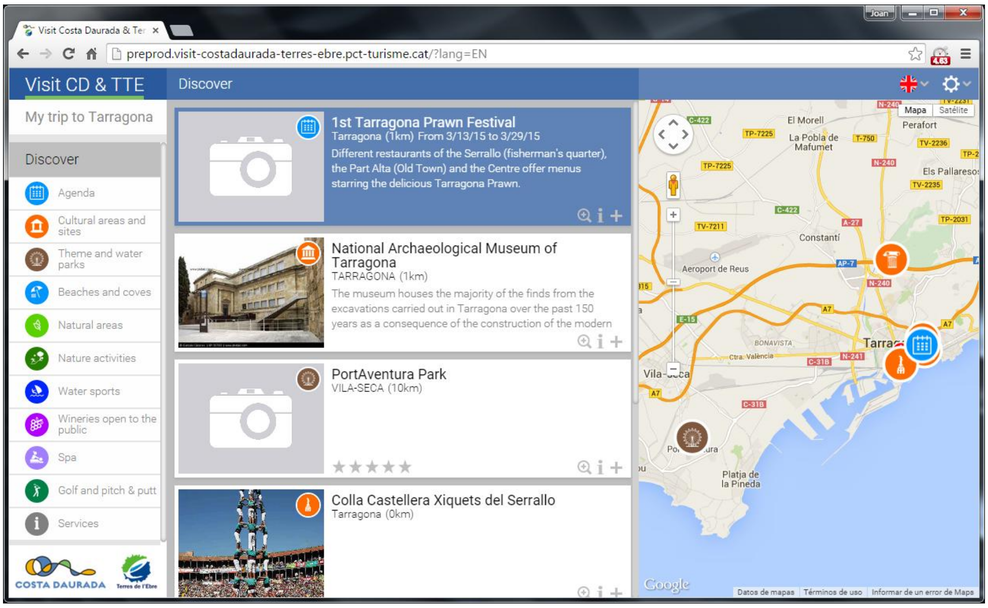


Figure 79. Visit CD & TTE: list and map of suggested activities

The system monitors continuously the navigation of the user through the map. When the selected geographic area does not contain any item from the current list, the system adds to the map the items better ranked within such region. The system tends to avoid overcrowding the list and the map with a large number of items, but if the user zooms in the map to a particular area to see one item, the system will push new items on the map. If it is possible, the system will always show at least 6 items on the map, taking into account the ranked list of items within the map region.

The user can select an item to see more information about it or to add it to the travel plan. Note that in this interface the list of activities in the travel plan is not shown at the same time than the suggestions. Nevertheless, the planned items are shown on the map. This has been done to reduce the amount of information shown on the web page, where sometimes the user may feel overwhelmed. The planned route is shown when the user clicks on the link of the name of the trip (top-left side of Figure 79). Then the list of suggestions switches with the list of the planned items, as shown in Figure 80. The order of the items to be visited can be arranged manually (drag and drop action) or automatically (with the most efficient path). If an automatic route is requested, the system distributes the activities in the available days depending on their location and their visiting time. Items are printed with numbers on the list and on the map so that the user may follow the planned route easily. The weather forecast application OpenWeatherMap API[52] has been used to print the weather prediction for each day of the trip, thus helping the user to decide if an item should be scheduled or not on a particular day. The system does not take into account the weather forecast in order to rank items; however, we keep it as a future action, since we consider it a relevant factor that may heavily affect the decision making process.

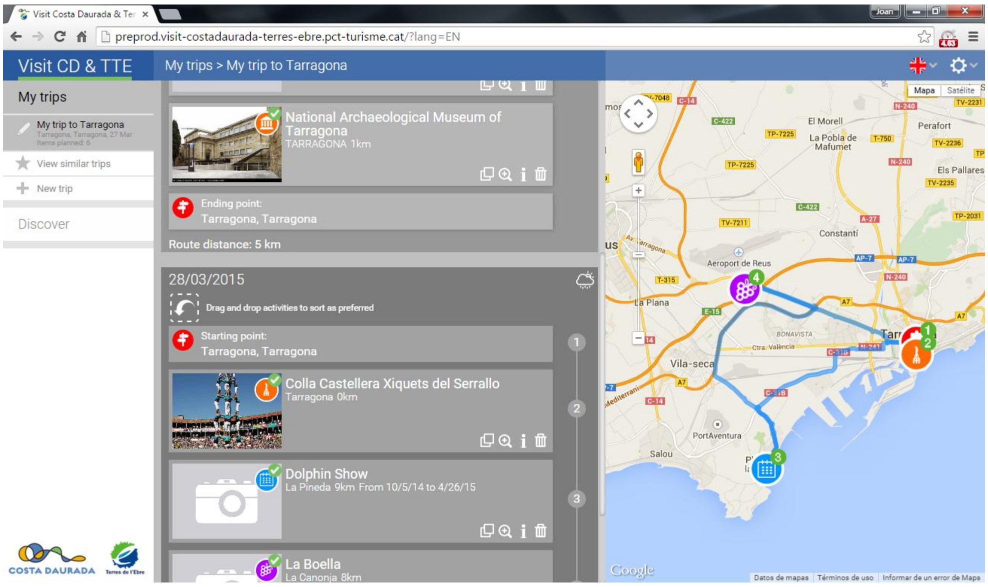


Figure 80. Visit CD & TTE: list and map of planned routes for each day of the trip

52 http://openweathermap.org/api (last access March 2015)

Another option available to the users permits them to get inspiration from trips created by other travellers. Therefore, we have included a new social aspect to the system, where users can explore trips of other users and copy the one that fits better with his/her preferences. Since the number of trips created by other users may be high in the future, the system will also show a ranked list of trips (with similar characteristics to the current user trip). We use the same process of similarity between users shown in section 5.3.2.1 to rank the trips shown to the user. However, since the main function is to copy an entire trip from other user, the similarities between trips using only demographic characteristics are not enough. For example, it is not useful for a user to copy an entire trip from another user that is similar in terms of the travel group and the motivations if such trip is done by car whereas the current user goes walking. Therefore, we added other parameters to this similarity measure, such as the location of the trip, the transport means, and the number of days of the trip. In equation (16) we added these three new parameters with a weight 0.5, since we consider them the most relevant factors (even more than the group composition). The x value for the location is measured with the function $dist(u,i)$ (21), where u is the current user location, i the location of the compared trip, and MAX_DISTANCE is given by the transport means selected by the current user. The x value for the transport means is set to 1 if they are the same and 0 otherwise. Finally, the value x of the parameter number of days is set to 1 if they are equal, to 0 if the difference is more than 5, and to the difference divided by 5 otherwise.

Figure 81 shows the page in which the user can navigate through similar trips. At the top of the centre panel there is a ranked list with similar trips. In this example we can see that the first trips are located in Tarragona with car driving directions and two or three days of trips as indeed was the trip created by the current user. Language, age and travel group are also taken into account to rank the trips but they are not as restrictive as the other parameters. Below the list of trips there are the details of the route of the selected trip selected, which the user may explore for each route day and the map. If he/she is satisfied with the trip, it can be fully copied to the current trip and thereafter be modified as necessary.

Another social functionality added to the system permits sharing the trip with friends. There are two options to share. The *private* sharing allows sending a link by mail to other participants in the trip, so that they can acess and modify the trip as they wish. The *public* option is to share the trip on social networks, such as Facebook or Twitter, in public mode. In this case all the user's friends will be able to see the trip, but they will not be able to modify it; however, they can copy it into another user session and then modify it. Figure 82 shows the panel that allows the user to share by social networks or by mail his/her trips.

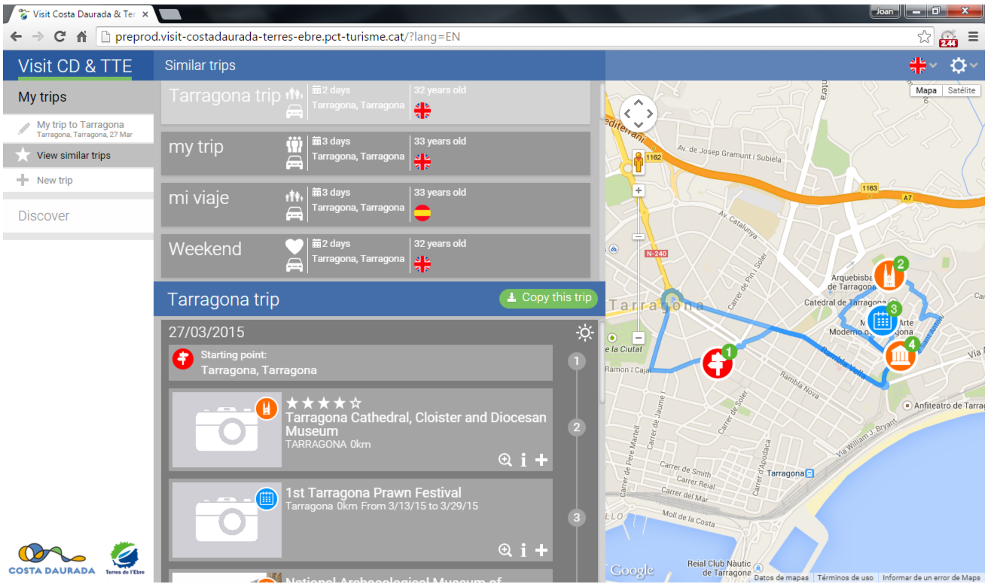


Figure 81. Visit CD & TTE: list of similar trips to be inspired or copy

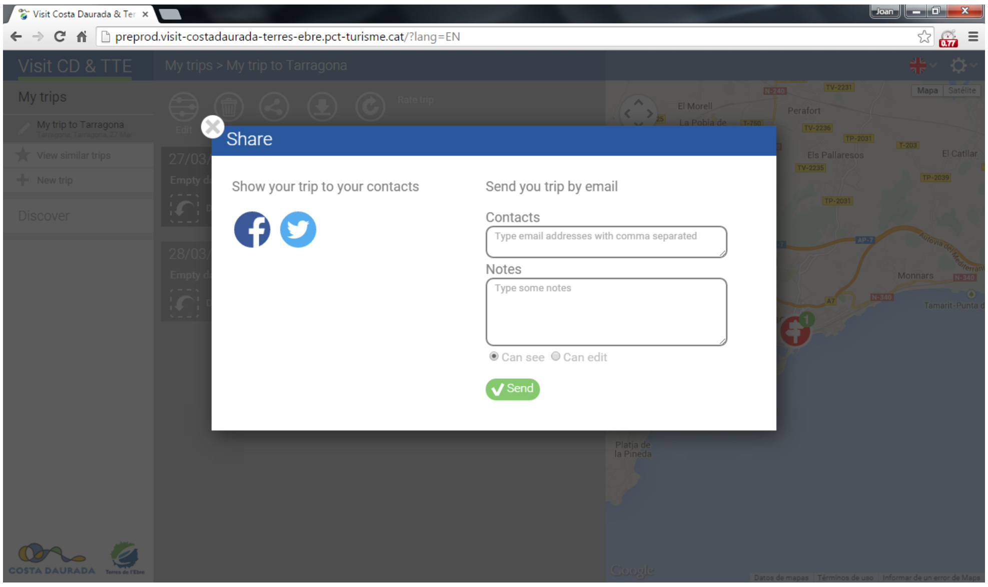


Figure 82. Sharing options of the trip

6.3. Validation

The validation of the SigTur/E-Destination system has been made from two different perspectives. The first one, of a quantitative nature, analyzed the quality of the recommendations made by the system to different tourist stereotypes. The second one, more qualitative, takes into account the whole system as a recommender product and analyzes the feedback received from several users during a public presentation at a FITUR meeting (the main Tourism fair in Spain).

6.3.1. 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.

Table 28 represents the demographic and travel data related to the profile of each different user stereotype: country of origin, travel group composition, accommodation, city destination, maximum distance allowed to move, travel dates, budget level and the segment prototype assigned as the most similar from the one shown in Table 15. Table 29 shows the degree of interest in each of the nine available motivations, with a percentage between 0 and 100.

Table 28. Demographic and travel data of stereotypes

User	Origin	Travel group	Accommodation	Destination	Max distance	Trip period	Budget	Prototype
1	ES	Friends < 25 yrs.	3 Stars hotel	Tarragona	15Km	15/6/15-17/6/15	23%	83
2	FR	Family > 36	Own home	L'Ampolla	45Km	15/7/15-31/7/15	63%	74
3	ES	Senior	3 Stars hotel	Falset	5Km	2/5/15-3/5/15	30%	4
4	FR	With children more than 12 years old	Camping	Cambrils	24Km	1/6/15-10/6/15	50%	31

Table 29. Interest value of the motivations selected by the stereotypes

User	Beach	Shopping	Relaxation	Leisure	Culture	Nature	Gastronomy	Sports	Shows/events
1	87%	17%	5%	45%	18%	9%	43%	7%	40%
2	92%	17%	36%	10%	18%	82%	28%	23%	20%
3	1%	17%	18%	10%	74%	1%	56%	0%	46%
4	81%	54%	31%	7%	5%	34%	20%	52%	26%

Tables 30-33 show the specific activities recommended to each tourist group. For each activity, the table shows its name, the type of the activity, the score (*S*) and confidence level (*CL*) obtained by the preference aggregation of both *CB* and *CF* recommendation methods, tag concepts of the ontology to which it is associated, the distance to the user destination and the *NFS* value measured with ELECTRE (we avoid other context variables for simplicity). To evaluate the quality of the recommendations, we have calculated the *Precision*, *Diversity* and F_{PD} (measures explained in section 4.4.1), and average distance of all recommended activities to the destination, which is given in Table 34.

Let us now analyze the recommendations obtained for each stereotype. The first profile is a group of young Spanish friends that visit Tarragona in summer, staying in a 3 star hotel. Their main motivations are going to the beach, leisure, gastronomy and attending events. They also have a small interest in culture or shopping. The first suggestions, as it is reasonable, are beaches. Then, a night life activity and a visit to a cellar are also suggested due to their interest in leisure and gastronomy. Note that the cellar has a lower CB score, since the concept *WineCellar* has ontology descendants in both *Gastronomy* and *Culture*, and the latter concept has lowered down its score. Although the user is not highly interested in culture, a museum is also recommended. The reason is that Tarragona has a large set of culture activities and therefore some items close to the ones that are more suitable to the user may be recommended so that he/she may discover new types of activities. The museum has a relatively good *NFS* because other items interesting for the user are penalised because they are further away from the city. The suggestion of the local market is caused by a high preference obtained with the collaborative filtering method, which produces the similarity with the segment prototype 83. Finally, a traditional event is suggested since the user is interest in events and this one takes place from the 5th to the 25th of June, within the trip period. All the recommendations are within 5 Km of distance; in fact, the one that is further away is the last beach, that is shown due to the high motivation value of this category and because the user may be interested in trying more than one during the trip. The precision of the suggestions is 1, because all *NFS* values are higher than 0.7 (NFS average is 0.83), and the diversity offered to the user is good enough (almost 0.5). Finally, the average distance is 1.55 Km.

Table 30. Recommended activities for stereotype 1

Name	Type	CB		CF		Tags	Distance (Km)	NFS
		S	CL	S	CL			
Platja del Miracle	Beaches	0.9	0.85	0.8	0.73	NormalBeaches, UrbanBeaches	1.71	1
Platja de l'Arrabassada	Beaches	0.9	0.85	0.8	0.73	NormalBeaches, UrbanBeaches	3.46	0.94
Platja Sabinosa	Beaches	0.9	0.85	0.8	0.73	NudismBeaches	4.43	0.9
Casino Tarragona (Rambla Nova)	NightLife	0.5	0.7	0.5	0.73	GameRoom	0.84	0.82
Agrícola Fuster	Cellars	0.3	0.62	0.5	0.43	WineCellar	0.04	0.77
Pedrera romana del mèdol	Museums	0.2	0.73	0.0	0.0	HistoryMuseums, HumanHeritage, Museums, Roman	0.04	0.73
Mercat Central de Tarragona	Shopping	0.2	0.85	0.7	0.73	LocalMarket	0.51	0.73
Tarraco Viva	Traditional Celebrations	0.3	0.77	0.0	0.0	TraditionalCelebrations	1.39	0.71

The second profile corresponds to a French family of middle-aged people that has a second home in a coastal destination at the south of Tarragona province (L'Ampolla). For their holidays they plan to stay the second fortnight of July and they basically want to go to the beach to relax and do some sports at the nature. They are willing to move 45 Kms in their excursions. First of all, the system suggests the two beaches that are within the village. Then, due to the high interest in nature and the relative interest in sports, the system suggests exploring some biking routes at the natural park that is near the city. A spa resort is also recommended since they set a high budget level and the associated prototype segment (74) produces a high score in *Relaxation* (ancestor concept of *SpaResorts*). Near the village there are no more beaches or nature activities, so the system suggests some sports that can be done close to the destination, increasing the diversity of the trip. The average *NFS* of the recommendations is 0.85 with a *Precision* of 1 and a good *Diversity* (0.61) that lead to a 0.76 value in F_{PD}. The average distance of the suggestions is less than 2 Kms., with a maximum of 3.57 Kms. If this profile is willing to move longer distances, he/she can discover other activities further down the list of recommendations. An alternative to suggest activities that are further away from the centre of the destination would be to decrease the weight of the location variable, but this is a measure that we want to evaluate in the future with an analysis of the use of the system.

Table 31. Recommended activities for stereotype 2

Name	Type	CB		CF		Tags	Distance (Km)	NFS
		S	CL	S	CL			
Platja de les Avellanes	Beaches	0.9	0.85	0.9	0.73	NormalBeaches	0.58	1
Cala Maria	Beaches	0.9	0.85	0.9	0.73	Coves	1.77	1
De l'Ampolla a la bassa de les Olles i lo Goleró	Biking	0.8	0.58	0.5	0.43	Biking, CoastalRoutes, InlandWatersRoutes, NaturalPark	2.93	0.9
L'entorn rural de l'Ampolla	Biking	0.8	0.58	0.5	0.43	Biking, CoastalRoutes, RuralRoutes	2.52	0.89
Spa & Wellness Les Oliveres Beach Resort	Spa	0.4	0.7	0.9	0.58	SpaResorts	3.57	0.76
Kitesurf. Badia l'Ampolla	Surfing	0.2	0.55	0.5	0.43	Kitesurfing	0.21	0.75
Tir amb arc. L'Ampolla	Archery	0.2	0.55	0.5	0.43	Archery	0.33	0.75
Excursió amb barca	Sailing	0.2	0.55	0.5	0.43	Boating	0.21	0.75

The next profile is a Spanish senior group that is mainly interested in culture, gastronomy and events. They are staying in a 3 star hotel on the first weekend of May in Falset, a region well known by its wine-related activity. They are not willing to move much distance since they do not have means of transport, that is why the suggested activities are very close to the village (an average of 0.14 Kms. in this case). The suggested activities, given their interest in gastronomy and events, are mainly visiting wine cellars and a wine fair that is scheduled every year on the first weekend of May. In relation to cultural activities the system also suggests the visit to a museum and a castle. Finally, even though they have not expressed a high interest in shopping, the associated segment prototype has full interest in it. Hence, the system suggests visiting the local market of the village. Since the village offers mainly activities related to the profile motivations, the results match quite accurately their interests and high values are obtained: average *NFS*=0.95 and *Precision*=1. However, due to the limitations of the distance, the diversity cannot be increased as in other profiles. In any case, the results seem very reasonable for this kind of travellers that do not want to move further away.

Table 32. Recommended activities for stereotype 3

Name	Type	CB		CF		Tags	Distance (Km)	NFS
		S	CL	S	CL			
Terra Personas	Cellars	0.6	0.64	0.3	0.39	WineTasting, WineCellar	0	1
William David Garsed	Cellars	0.6	0.64	0.3	0.39	WineTasting, WineCellar	0	1
Baronia d'Entença	Cellars	0.6	0.64	0.3	0.39	WineTasting, WineCellar	0	1
Don Carles Vins	Cellars	0.6	0.64	0.3	0.39	WineTasting, WineCellar	0	1
Fira del vi de Falset	Gastrono myEvents	0.5	0.7	0.3	0.39	FoodEvents, WineEvents	0	0.98
Castell del Vi	Museums	0.8	0.81	0.3	0.39	Museums, WineMuseums	0.42	0.9
Castell de Falset	Culture	0.7	0.55	0.0	0.0	Castle	0.63	0.86
Mercadet de Falset	Shopping	0.2	0.85	1.0	0.69	LocalMarket	0.14	0.83

The last simulated stereotype is a French family with children more than 12 years old, staying ten days of July in a camping in Cambrils (a village known for its coastal area). Their main motivations are going to the beach, shopping and sports. The associated prototype segment (31) gives also a high preference in beaches but also in relaxation. That is why the system suggests visiting a spa for relaxation and beaches. A local market and several sport activities are also suggested, profiting from the large offer of sport activities that can be done in Cambrils. The precision of the recommendations (0.88) is lower than those of the other profiles due to the recommendation of an activity with *NFS* lower than 0.7. This happened because it was not possible to find in Cambrils other relevant (and diversified) activities and hence the best one was a bit further away (6.82 Kms.) and such distance has decreased the *NFS*. Although the *Precision* of the recommendation in this profile has decreased, the *Diversity* has improved with a value of 0.72 and the final F_{PD} is 0.79 (the best value among all profiles). The average distance among all the recommendations is 1.82 Kms.

Table 33. Recommended activities for stereotype 4

Name	Type	CB		CF		Tags	Distance (Km)	NFS
		S	CL	S	CL			
Platja de la Riera	Beaches	0.8	0.85	0.9	0.77	NormalBeaches, UrbanBeaches	0.37	1
Mercadet de Cambrils	Shopping	0.5	0.85	0.5	0.77	LocalMarket	0.84	0.84
Creuers pel litoral	Sailing	0.5	0.6	0.4	0.47	Boating	0.45	0.81
Busseig. Cambrils	Under Water	0.5	0.55	0.4	0.47	Snorkelling	0.82	0.81
Belles Aigües Spa	Spa	0.3	0.7	0.7	0.62	SpaResorts	0.23	0.78
Windsurf. Cambrils	Surfing	0.5	0.55	0.4	0.47	Windsurfing	1.48	0.78
Pitch & Putt GOLF CAMBRILS	Golf	0.5	0.7	0.4	0.62	PitchAndPutt	3.58	0.7
De Montbrió del Camp al parc de Samà	Biking	0.4	0.63	0.4	0.47	Biking, CultureRoutes, RuralRoutes	6.82	0.53

Table 34. Precision, diversity, F_{PD} and average distance of the recommendations given for each user

User	Precision	Diversity	F_{PD}	Average NFS	Average distance (Km)
1	1	0.49	0.66	0.83	1.55
2	1	0.61	0.76	0.85	1.51
3	1	0.34	0.5	0.95	0.14
4	0.88	0.72	0.79	0.78	1.82

6.3.2. Feedback from real users

During the FITUR'11 conference (International Tourism Fair) held in Madrid, the SigTur/E-destination web site 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 (you can see some picture of the event in Figure 83). 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.

Figure 83. Pictures from FITUR'11. On the left, a user testing the system. On the right, two groups of users answering the questionnaire with interviewers.

Table A. 1 and Table A. 2 in the annex show the whole questionnaire (questions and allowed values) used in the evaluation. At the end of 5 days, 78 forms were collected: 28 from Tourism professionals and 50 from end users. Figure 84 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 was decided to implement the mobile version to be run on smart phones.

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).

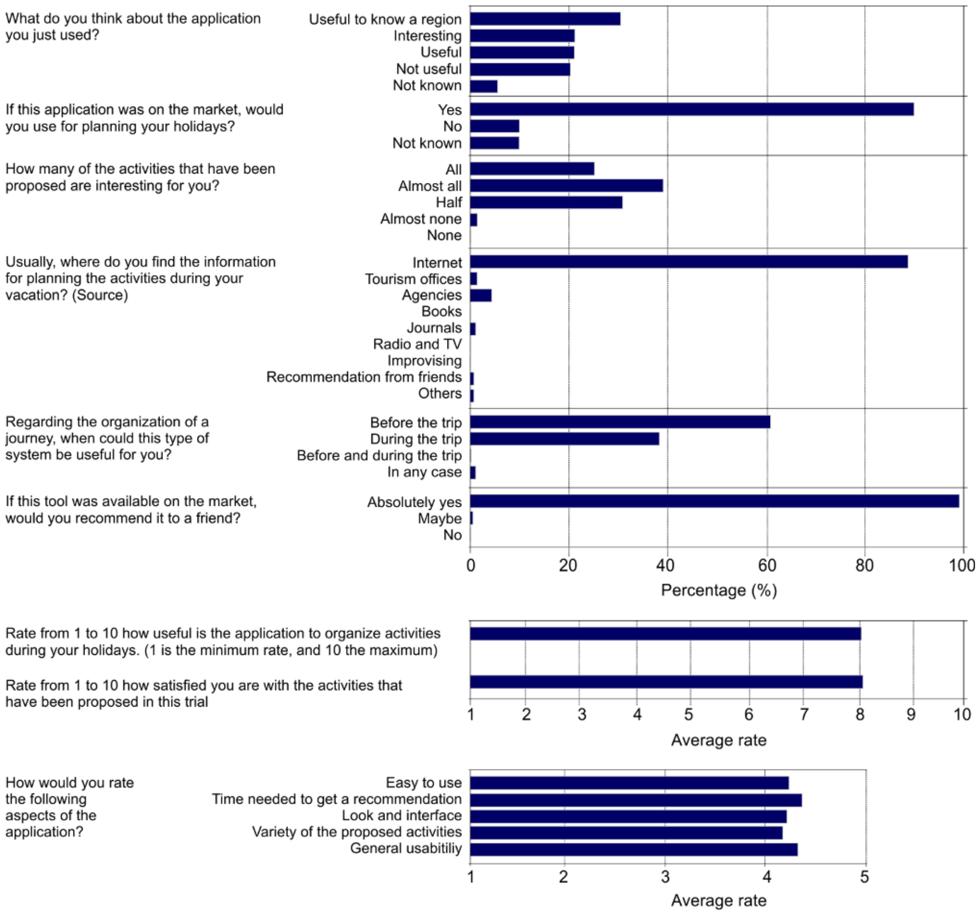


Figure 84. Results of the questionnaire.

Thanks to this feedback, the SigTur/E-Destination system has been heavily improved during the last years. 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.

6.4. Conclusions

This chapter has described two adaptations of the SigTur/E-Destination system, showing how its core can be reused for different domains or areas adapting their functionalities, interfaces and touristic requirements.

The first adaptation focused on enotourism, which is a domain with a great potential in the province of Tarragona. These kinds of tools may also increase the discovery of related places for tourists visiting Tarragona, allowing a diversification of the tourism supply. Although there are enotourism information systems in some destinations, they do not provide personalised recommendations to satisfy different user preferences, as EnoSigTur does. The system is focused on suggesting activities related (or complementary) to an enotourism trip, such as visiting cellars or going for a walk through vineyard landscapes. Hence, the ontology has been extended with new concepts that represent these specific kinds of activities. The management of the user profile has also been adapted to represent eno-tourists with different variables, such as the motivations or the size of the group. The interface of both the Web and the mobile platforms has been changed with more wine-related colours, logos and a new presentation page with a better design.

Another adaptation of the system has been made for the official tourism destination management of the *Costa Daurada & Terres de l'Ebre*. The interface of the system has been designed to follow the corporate representation of their current Web site. The official tourism destination managers have some different requirements on the user profile representation: they considered interesting to ask the age of the tourist and the transport means, and they suggested some slightly different motivations and allowed values for the tourist group's type. The language of the user accessing the Web site has been used instead of the country of origin. In addition, the system has been enhanced with two new social aspects. First, a new section where users may get inspired with trips from other tourists with similar preferences. And second, a new functionality that allows to share a trip in public and private mode to collaborate with the other members of the travel group. For the near future, we plan to develop this adaptation of the system to mobile platforms including these new functionalities.

Finally, the core of the system has been validated using the generic SigTur/E-Destination recommender. First, an analysis of the recommendations given to some stereotyped (but realistic) tourist profiles has proved that the results are reasonable, taking into account the preferences and needs of each kind of user. This chapter has also shown how the satisfaction of real users was evaluated. The feedback of 78 people was obtained in a specialised Tourism fair through a questionnaire in which they expressed their overall positive level of satisfaction with the interface of the system and the recommendations.

Chapter 7 – Conclusions and future work

The advent of the Social Web has lead to an overwhelming amount of information available on Internet, which hampers the task of finding the data that are more relevant for a specific user faced with a decision problem. Web search engines return the most popular pages associated to a certain textual query, but they do not take into account the specific needs and interests of the user and they are certainly not suited to the capture and analysis of large quantities of options. In this context, *Recommender Systems* appeared as intelligent tools able to retrieve, analyze, filter and rank different options depending on the user's preferences.

The Tourism sector has been deeply affected by the increasing use of Communication and Information Technologies, and currently 74% of tourists use the Internet to search for information on the Web to plan their trips (Google, 2014). In this planning stage they tend to spend quite a long time looking on the Web for precise, up-to-date and trustable information; in fact, a study by Expedia affirms that they visit around 38 sites before booking a vacation (Expedia, 2013). *Travel Recommender Systems* may provide to potential visitors a set of tools that may inspire and help them when planning their next trips, offering personalised information that fits their needs and preferences.

Our first contribution has been a ***thorough analysis of the state of the art in intelligent Travel Recommender Systems*** described in the main Artificial Intelligence journals and conferences during the last years (reported in chapter 2). The main aspects analysed have been the type of interface they offer, their main functionalities and the recommendation techniques and AI methods they apply. Some points of improvement and some general guidelines to be considered in the development of this kind of systems were given in (Borràs et al., 2014).

We detected that most of the existing Tourism recommender systems do not provide native apps for the currently most popular mobile platforms (Android and iOS). This aspect is very relevant, taking into account the huge increase of search of information done by tourists from their mobile devices when they are already at their destination. *We have developed apps for both Android and iOS* that not only offer the same functionalities than the Web application, but also provide additional services such as the automatic detection (and consideration in the recommendation process) of the user's location with the GPS and the ability to use the app in an offline mode for those foreign tourists without data roaming capabilities. In the future we intend to improve the current Web site, making a more responsive design that may adjust the visualization to the available screen size, making the planning site more useful for those users accessing the Web application through the Web browser of mobile devices.

Other approaches take profit of social recommendations to provide more accurate and trusted suggestions, based on the opinions of the closest friends. Although we have not included social opinions directly in the recommendation process, we make use of social networks to *allow users to share content publicly or privately*, so that a group of people can cooperate in the design of the route. Moreover, in the system we have included *functionalities that permit users exploring, copying and rating trips* from similar users, paving the way towards a future explicit management and exploitation of these (implicit) communities.

Travel Recommender Systems commonly use content-based and collaborative filters to provide suggestions based on the tastes of the user and those of similar users. The combination of both methods has been widely applied to try to overcome their individual drawbacks. In this dissertation we have widely integrated several recommendation methods: content-based, collaborative filters and demographic-based techniques. The main contribution in this regard is the definition of a ***novel method that allows to measure different kinds of similarities between users***, taking into account several viewpoints (preferences, demographic data, similarity to predefined stereotypes, etc). Hence, collaborative filtering methods may be applied in different ways, depending on the way in which the resemblance between users is defined. The new similarity measure, which combines some complex aggregation operators, allows measuring the level of the similarity between two users considering a set of attributes.

The mobility of visitors is one of the key points to consider when making recommendations in Tourism. Context-aware techniques have already been used to customize the suggestions taking into account aspects such as the location of the tourist or the time of the day. We have also integrated the tourist context to improve the quality of the recommendations, modelling information about the location, the travel dates, the budget or the group composition (e.g. the presence of children, the size of the group or the spoken language). In order to combine all those variables we have defined ***a new multi-criteria sorting approach that permits to rank the options*** for each individual user (Del Vasto-Terrientes et al., 2015; Borràs et al., 2012c). This kind of MCDA methods is starting to be applied to improve the accuracy of recommender systems; however, those based on *outranking* have not been exploited in recommender systems yet. Such a varied combination of methodologies gives rise to a high number of configuration parameters, which allow giving specific weights to each attribute. In the near future we would like to study the possibility of dynamically setting the values of these parameters depending on the user profile. Another important improvement of the system would be to increase the contextual inputs with aspects such as the weather forecast, the time of the day, the season of the year and the management of unexpected events. In this way the system could suggest indoor or outdoor activities depending on the weather, send a notification to the user if he/she is close to a restaurant that offers his/her favourite food at lunch time, or even re-organize dynamically the trip plan if the user can not visit some places due to an unforeseen event.

An important current line of research is the use of semantic domain knowledge to improve the quality of recommendations. A *semantic recommender system* can leverage this knowledge to represent user profiles and domain items. The hierarchical structure of ontologies allows an analysis of preferences at different abstraction levels and provides reasoning capabilities to recommender systems. ***We have analysed recent semantic recommender systems*** (Valls, et al 2013) focusing on how they represent, update and infer the information about the users' preferences. Another contribution of this dissertation is the development of a ***new framework that exploits any ontology hierarchy to make both bottom-up and top-down dynamic inferences not only about the preferences of the users but also about their certainty*** (Borràs et al., 2012b). We have tested this framework with a Tourism ontology designed ad-hoc for this work. In chapter 3 we detailed the study of the semantic approaches and the new framework.

Preferential knowledge is usually acquired directly from the user (explicitly or implicitly) or through the analysis of similar users. However, the integration of the context through the acquisition of such preferences has been seldom considered. For instance, a tourist having interests in sports in summer might not be necessarily interested in sports in winter, or a tourist who enjoys going to the beach whenever he/she visits a coastal area may have more cultural interests when he/she is located in a historical city far from the coast. Therefore, we consider an interesting research line to integrate this contextual information when reasoning about the user preferences.

Recommender systems aim to provide personalised information to users in order to satisfy their needs. The evaluation of their results is normally done with mathematical metrics that measure the accuracy of the correspondence between the inferred user preferences and the characteristics of the suggested items. However, accuracy is not the only factor that satisfies users. For instance, even though a tourist is very interested in going to the beach, it is probably not very useful to suggest only beaches, but it could also be interesting to suggest some activities or markets that are near them. It has been argued that *smart recommenders* should provide *diversified* recommendations to increase the satisfaction of the user. Moreover, such diversification may produce *serendipitious* results, which allow users to discover unexpected (but relevant) items. Last but not least, applying diversification on the tourist offer may be also beneficial for retailers, through the increase of the visibility and the sales of less popular items, and for destinations, by improving the flow of tourists through different areas. In this regard ***we have analysed the state of the art of the methods that apply individual diversification*** on lists of recommendations. We have also proposed, as explained in chapter 4, ***a new diversification method based on semantic clustering*** (Borràs et al., submitted) that has been tested against the current methods, showing good results in terms of precision and diversity. This method also makes the recommendation process more scalable to large datasets, due to their clusterization. Moreover, the method dynamically adapts the diversity level based on the preferences of the user.

Although diversity may increase user satisfaction, it is still an open problem to measure this satisfaction so that the system may auto-configure its parameters if necessary. This could be done either *explicitly*, asking users about their agreement with each recommendation, or *implicitly*, analysing the user's behaviour (e.g. checking if the user selects diversified items to be added to his/her routes, or to analyze the position of the selected items). From the retailers and DMOs point of view, another way to improve the recommendation would be to apply *aggregate* diversity mechanisms, thereby increasing the probability to recommend less popular activities. It has been said that diversity may produce serendipity; hence, an interesting research line would be to measure the serendipity of each item (with respect to a given user) and implicitly include it in the recommendation process.

Finally, from a technical perspective, ***a fully functional Travel Recommender System has been designed and developed*** (Moreno et al., 2013a; Borràs et al., 2011; De la Flor et al., 2012). The system, ready to be used on real tourists, applies the scientific methods proposed in this dissertation. SigTur/e-Destination is a user-friendly Web and mobile application that interacts with the recommender engine to provide dynamically suggestions based on the actions of the user. A functionality that might be included in the system is the ability to produce natural language explanations of the rationale behind the recommendation of each item. For instance, the system could show messages like "it is good for kids", "users from your country have enjoyed it recently", "it is near your accommodation" or "it is complementary to your main interests but it can be interesting". In this way, the user may understand better why the system suggests one item and not another. Moreover, we could enhance the user's feedback by allowing different actions for each suggested item, like "I do not want to receive more items for kids", "I do not want to receive more beaches", "I want more items similar to this" or "I want items near this one".

The system has been tested with real users to detect its main limitations and their level of satisfaction with the recommendations. Although these tests have been limited to a few users, we intend to make the system publicly available in short to get as many users as possible and get their feedback. Moreover, the core of the recommendation engine has been used with different interfaces and user profiles. Concretely, ***a thematic specialisation on Enotourism*** (Borràs et al., 2013; Borràs et al., 2012a) ***and a geographical specialisation on Costa Daurada and Terres de l'Ebre have been successfully completed***. These results show that this kind of approach is generic enough to be potentially reusable in other areas with few modifications, opening the door to possible marketing possibilities through the Scientific and Technological Park of Tourism and Leisure.

In summary, the main future research lines are the following:

- To make a more advanced study of the influence of the parameters in the recommendation process, so that the system may automatically adapt their values to the characteristics and needs of the user.

- To enhance the contextual information used by the recommender, including aspects such as the weather forecast, the season of the year, the time of the day or even the management of unforeseen events (e.g. to reconsider the activities in a route and re-plan it if there has been a traffic accident and the road the tourist had to take is blocked).

- To develop new ways in which information about the preferences of the user may be implicitly inferred from the context surrounding the recommendation.

- To study new ways in which aggregate diversity may be incorporated in the recommendation process. Regarding the issue of diversity, we could also think of new ways of producing serendipitous results and measuring their effect on the user.

- To continue developing ways of receiving explicit and implicit feedback information from the users and analyzing it to improve the way in which recommendations are selected and presented.

- To add more explanations to the user about the reasons that have motivated a certain recommendation, and to allow the users to provide more precise feedback not only on the recommendations but also on the reasoning behind them.

- To develop a new version of the Web-based interface, making a more responsive design. After that we would like to make the SigTur/e-Destination system openly available, so that it can be used by any tourist wishing to visit the Tarragona province. From a more technical perspective, we would also like to continue exploring the possibility of adapting the main recommendation interface and engine to other kinds of Tourist destinations.

The main scientific publications derived from the work done in this Ph.D. Thesis are the following:

Journals:

- Moreno, A., Valls, A., Isern, D., Marin, L., & Borràs, J. (2013). *Sigtur/e-destination: ontology-based personalized recommendation of tourism and leisure activities. Engineering Applications of Artificial Intelligence, 26(1), 633-651.* ISI-JCR Impact Factor: 1.962, Engineering, Multidisciplinary, (Q1, 15/87)

- Borràs, J., Moreno, A., Valls, A. (2014) *"Intelligent tourism recommender systems: a survey". Expert Systems with Applications 41.16 (2014): 7370-7389.* ISI-JCR Impact Factor: 1.965, Computer Science, Artificial Intelligence, (Q1, 30/121)

    - According to data provided by Springer in April 2015, this paper had been already downloaded or viewed more than 3,400 times since its publication (November 2014 issue). On April 20th, 2015, it appears in the 18th position in the list of most downloaded articles from the *Expert Systems with Applications* journal in the last 90 days[53].

- Del Vasto-Terrientes, L., Valls, A., Zielniewicz, P., Borràs, J. (2015) "*A hierarchical multi-criteria sorting approach for recommender systems*". Submitted to *Journal of Intelligent Information Systems* (in Press). ISI-JCR Impact Factor: 0.632, Computer Science, Information Systems, (Q3, 97/135)

- Borràs, J., Moreno, A., Valls, A. *"Diversification of recommendations through semantic clustering"*. Submitted to *IEEE Transactions on Knowledge and Data Engineering.* ISI-JCR Impact Factor: 1.815, Computer Science, Information Systems, (Q1, 25/135)

Book chapters:

- Borràs, J., de la Flor, J. Pérez, Y., Moreno, A., Valls, A., Isern, D., Orellana, A., Russo, A., Anton-Clavé, S. (2011) *SigTur/E-destination: A system for the management of complex tourist regions. In: Information and Communication Technologies in Tourism Conference.* R. Law, M. Fuchs, F. Ricci, Eds., pp. 39-50, Springer Verlag. This book contains the proceedings of the 18th International Conference on Information Technology and Travel and Tourism (ENTER-2011), held in Innsbruck (Austria) in January 2011.

[53] http://www.journals.elsevier.com/expert-systems-with-applications/most-downloaded-articles/

- De la Flor, J., Borràs, J., Isern, D., Valls, A., Moreno, A., Russo, A., Pérez, Y., Anton-Clavé, S. (2012). *Semantic Enrichment for Geospatial Information in a Tourism Recommender System. Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications, 134-156.* Eds: L.Díaz, C.Granell, J.Huerta. IGI-Global.

- Borràs, J., Valls, A., Moreno, A., Isern, D. (2012). *"Ontology-based management of uncertain preferences in user profiles". Information Processing and Management of Uncertainty in Knowledge-Based Systems, Part II.* Eds: S.Greco, B.Bouchon-Menier, G.Colletti, M. Fedrizzi, B.Matarazzo, R.Yager. Communications in Computer and Information Science 298, pp. 127-136, Springer Berlin Heidelberg. This book contains the proceedings of the 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-2012), held in Catania (Italy) in July 2012.

- Valls, A., Moreno, A., Borràs, J. (2013). *"Preference Representation with Ontologies". Multicriteria Decision Aid and Artificial Intelligence: Links, Theory and Applications, pp. 77-99.* Eds: M.Doumpos, E.Grigoroudis. John Wiley and Sons.

Other conferences, workshops and scientific meetings:

- Borràs, J., Moreno, A., Valls, A., Ferré, M., Ciurana, E., Salvat, J., Russo, A., Anton-Clavé, S. (2012). *Uso de técnicas de Inteligencia Artificial para hacer recomendaciones enoturísticas personalizadas en la Provincia de Tarragona*. *IX Congreso Turismo y Tecnologías de la Información y las Comunicaciones, TURITEC-2012.* Málaga, Spain, 2012, 217-230.

- Borràs, J., Valls, A., Moreno, A., Isern, D. *Using MCDA techniques to build personalized and contextualized tourist trip plans*. Book of abstracts of the 75th Meeting of the European Working Group on Multiple Criteria Decision Aiding, pp. 9, Tarragona, Spain, April 2012.

- Borràs, J., Valls, A., Moreno, A., Isern, D. Ontology-based management of uncertain preferences in user profiles. Proceedings of the workshop DA2PL-From Multiple Criteria Decision Aid to Preference Learning, pp. 83-89. Mons, Belgium, November 2012.

- Borràs, J., Salvat, J., Valls, A., Moreno, A., Russo, P., Anton-Clavé, S. EnoSigTurMòbil, un sistema intel·ligent de recomanació personalitzat d'activitats enoturístiques. Technical Meeting on Research on Enology and Wine in Catalonia. URV, Tarragona, June 2013.

- Borràs J. GO-ENOTUR: Una app para descubrir territorios enoturísticos. Presentation on Tourism and Technologies Forum TuristTIC, April 2014. http://www.forumturistic.com/

- Borràs J. SigTur/E-Destination. Presentation on International Tourism Fair in Spain (FITUR), February, 2011.

Appendix A – Full questionnaire employed in FITUR-2011

Table A. 1. Questions and allowed answers of the questionnaire

Id.	Question	Allowed answers
1	What do you think about the application you just used?	Free answer
2	If this application was on the market, would you use for planning your holidays?	Yes / No / I don't know
2a	If the answer to question 2 was 'Yes', Which would be the use? What would be the advantages for you?	Free answer
2b	If the answer to question 2 was 'No', Why wouldn't you use it? What disadvantages would it have?	Free answer
2c	If the answer to question 2 was 'I don't know', What aspects make you hesitate?	Free answer
3	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).	1..10
4	Rate from 1 to 10 how satisfied you are with the activities that have been proposed in this trial	1..10
5	How many of the activities that have been proposed are interesting for you?	One of these values: all / almost all / half of them / almost none / none
6	Concerning the obtained recommendations, have you missed any activity that could be interesting for you?	Yes / No
6a	If the answer to question 6 was 'Yes', give examples.	Free answer
7	Referring to the information about the activities, have you missed any particular piece of data that could be interesting for you?	Yes / No
7a	If the answer to question 7 was 'Yes', give examples.	Free answer

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

Id.	Question	Allowed answers
8	How would you rate the following aspects of the application? - Easy to use - Time needed to get recommendations - Look and interface - Variety of the proposed activities - General usability	1..6 (1 means unsatisfied, 5 means satisfied, 6: Don't know)
9	Usually, where do you find the information for planning the activities during your vacation? (Source)	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
10	Regarding the organization of a journey, when could this type of system be useful for you?	One of the following items: before the trip, during the trip to get particular activities, both, never because I like to improvise
11	Would you use a recommender like this to plan your vacations?	One of these values: Sure / I would use it with some doubts / No
11a	If the answer to question 11 was 'I would use it with some doubts', Why?	Free answer
11b	If the answer to question 11 was 'No', Why not?	Free answer
12	If this tool was available on the market, would you recommend it to a friend?	One of these values: Absolutely yes, maybe, No
12a	If the answer to question 12 was 'Maybe', explain the reason	Free answer
12b	If the answer to question 12 was 'No?, explain the reason	Free answer

References

Adomavicius G, Kwon Y (2010) Improving Aggregate Recommendation Diversity Using Rank-Based Techniques. IEEE Transactions on Knowledge and Data Engineering. 24(5), pp. 896-911.

Adomavicius G, Manouselis N, Kwon Y (2011) Multi-criteria recommender systems. In Recommender systems handbook, Springer US, pp. 769-803.

Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17, pp. 734-749.

Akiyama T, Obara K, Tanizaki M (2010) Proposal and Evaluation of Serendipitous Recommendation Method Using General Unexpectedness. In Proc. Workshop on the Practical Use of Recommender Systems, Algorithms and Technologies, CEUR, pp. 3-10.

Albadvi A, Shahbazi M (2009) A hybrid recommendation technique based on product category attributes. Expert Systems with Applications, 36, pp. 11480-11488.

Al-Hassan M, Lu H, Lu J (2015) A Semantic Enhanced Hybrid Recommendation Approach: a Case Study of E-government Tourism Service Recommendation System. Decision Support Systems, 72, 97-109.

Alonso K, Zorrilla M, Confalonieri R, Vázquez-Salceda J, Iñana H, Palau M, Calle J, Castro E (2012) Ontology-based tourism for all: recommender and information retrieval system for interactive community displays. Procs. of 8th International Conference on Information Science and Digital Content Technology, pp. 650-655.

Amato F, Mazzeo A, Moscato V, Picariello A (2013a) A Recommendation System for Browsing of Multimedia Collections in the Internet of Things. In Internet of Things and Inter-cooperative Computational Technologies for Collective Intelligence, Springer Berlin Heidelberg, pp. 391-411.

Amato F, Mazzeo A, Moscato V, Picariello A (2013b) Exploiting Cloud Technologies and Context Information for Recommending Touristic Paths. Intelligent Distributed Computing VII - Proceedings of the 7th International Symposium on Intelligent Distributed Computing, IDC 2013, Prague, Czech Republic, pp. 281-287.

Anton-Clavé S (2009) Experience and emotion consumption. Perspectives for Innovation in Tourism and Leisure in Catalonia. Paradigmes, 3, pp. 93-100.

Anton Clavé, S. (2010). Leisure parks and destination redevelopment: the role of PortAventura, Catalonia. Journal of Policy Research in Tourism, Leisure and Events 2 (1), pp. 67-79.

Aseervatham S, Bennani Y (2009) Semi-structured document categorization with a semantic kernel. Pattern Recognition 2009, 42(9), pp. 2067-2076.

Aytekin T, Karakaya M (2014) Clustering-based diversity improvement in top-N recommendation. Journal of Intelligent Information Systems, pp. 1-18.

Barta R, Feilmayr C, Pröll B, Grün C, Werthner H (2009) Covering the semantic space of tourism: An approach based on modularized ontologies. In: 1st Workshop on Context, Information and Ontologies, CIAO 2009. Heraklion, Greece; 2009. p. 1.

Batet M, Moreno A, Sánchez D, Isern D, Valls A (2012) Turist@: agent-based personalised recommendation of touristic activities. Expert Systems with Applications, 39(8), pp. 7319-7329.

Batet M, Valls A, Gibert K, Sánchez D (2010) Semantic clustering using multiple ontologies. 13th International Conference of the Catalan Association for Artificial Intelligence, IOS Press, 220, pp. 207-216, Spain, 2010.

Baumann P (1994) Management of multidimensional discrete data. Very Large Databases Journal 3(4), pp. 401-444.

Berka T, Plößnig M (2004) Designing recommender systems for tourism. Proceedings of the 11th international conference on information technology in travel & tourism (ENTER'2004).

Berners-Lee T, Hendler J, Lassila O (2001) The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, Scientific American 2001, 284, pp. 34-43.

Bhatt M, Wenny J, Prakash S, Wouters C (2009) Ontology driven semantic profiling and retrieval in medical information systems. J. Web Sem. 7(4), pp. 317-331.

Blanco-Fernández Y, López-Nores M, Gil-Solla A, Ramos-Cabrer M, Pazos-Arias J (2011a) Exploring synergies between content-based filtering and spreading activation techniques in knowledge-based recommender systems. Information Sciences 181(21), pp. 4823-4846.

Blanco-Fernández Y, López-Nores M, Pazos-Arias JJ, García-Duque J (2011b) An improvement for semantics-based recommender systems grounded on attaching temporal information to ontologies and user profiles, Engineering Applications of Artificial Intelligence 2011, 24, pp. 1385-1397.

Borràs J, De la Flor J, Pérez Y, Moreno A, Valls A, Isern D, Orellana A, Russo A, Anton-Clavé S (2011) SigTur/E-destination: A system for the management of complex tourist regions. In: Information and Communication Technologies in Tourism Conference. R. Law, M. Fuchs, F. Ricci, Eds., pp. 39-50, Springer Verlag.

Borràs J, Moreno A, Valls A (2014) Intelligent tourism recommender systems: A survey. Expert Systems with Applications, 41(16), pp. 7370-7389.

Borràs J, Moreno A, Valls A, Ferré M, Ciurana E, Salvat J, Russo A, Anton-Clavé S (2012a) Uso de técnicas de Inteligencia Artificial para hacer recomendaciones enoturísticas personalizadas en la Provincia de Tarragona. IX Congreso Turismo y Tecnologías de la Información y las Comunicaciones, TURITEC-2012. Málaga, Spain, 2012, pp. 217-230.

Borràs J, Moreno A, Valls A. Diversification of recommendations through semantic clustering. Submitted to IEEE Transactions on Knowledge and Data Engineering.

Borràs J, Salvat J, Valls A, Moreno A, Russo P, Anton-Clavé S (2013) EnoSigTurMòbil, un sistema intel·ligent de recomanació personalitzat d'activitats enoturístiques. Technical Meeting on Research on Enology and Wine in Catalonia. URV, Tarragona, June 2013.

Borràs J, Valls A, Moreno A, Isern D (2012b) Ontology-based management of uncertain preferences in user profiles. Information Processing and Management of Uncertainty in Knowledge-Based Systems, Part II. Eds: S.Greco, B.Bouchon-Menier, G.Colletti, M. Fedrizzi, B.Matarazzo, R.Yager. Communications in Computer and Information Science 298, pp. 127-136, Springer Berlin Heidelberg.

Borràs J, Valls A, Moreno A, Isern D (2012c) Using MCDA techniques to build personalized and contextualized tourist trip plans. Book of abstracts of the 75th Meeting of the European Working Group on Multiple Criteria Decision Aiding, pp. 9, Tarragona, Spain, April 2012.

Bouneffouf D (2013) Situation-aware approach to improve context-based recommender system. arXiv preprint arXiv:1303.0481.

Braunhofer M, Elahi M, Ricci F (2015) User Personality and the New User Problem in a Context-Aware Point of Interest Recommender System. In Information and Communication Technologies in Tourism 2015, pp. 537-549, Springer International Publishing.

Braunhofer M, Elahi M, Ricci F, Schievenin T (2013) Context-Aware Points of Interest Suggestion with Dynamic Weather Data Management. Information and Communication Technologies in Tourism 2014. Springer International Publishing, pp. 87-100.

Breese JS, Heckerman D, Kadie C (1998) Empirical Analysis of Predictive Algorithms for Collaborative Filtering. 1998

Brewer JD (1984) Tourism and ethnic stereotypes: Variations in a Mexican town. Annals of Tourism Research, 11(3), pp. 487-510.

Buján D, Martín D, Torices O, López-de-Ipiña D, Lamsfus C, Abaitua J, Alzua-Sorzabal A (2013) Context management platform for tourism applications. Sensors, 13(7), pp. 8060-8078.

Burke R (2002) Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction 12(4), pp. 331-370.

Candillier L, Chevalier M, Dudognon D, Mothe J (2011) Diversity in Recommender Systems: bridging the gap between users and systems. In CENTRIC 2011, The Fourth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services, pp. 48-53.

Candillier L, Chevalier M, Dudognon D, Mothe J (2012) Multiple Similarities for Diversity in Recommender Systems. International Journal On Advances in Intelligent Systems, 5(3 and 4), pp. 234-246.

Cantador I (2008) Exploiting the Conceptual Space in Hybrid Recommender Systems: a Semantic-based Approach. PhD thesis, Universidad Autónoma de Madrid, Madrid, Spain.

Cantador I, Castells P (2011) Extracting multilayered Communities of Interest from semantic user profiles: Application to group modeling and hybrid recommendations. Computers in Human Behavior, 27(4), pp. 1321-1336

Castillo L, Armengol E, Onaindía E, Sebastiá L, González-Boticario J, Rodríguez A, Fernández S, Arias J D, Borrajo D (2008) Samap: An user-oriented adaptive system for planning tourist visits. Expert Systems with Applications, 34, pp. 1318-1332.

Ceccaroni L, Codina V, Palau M, Pous M (2009) PaTac: Urban, Ubiquitous, Personalized Services for Citizens and Tourists. In Third International Conference on the Digital Society (ICDS 2009), February 1-7, 2009, Cancun, México, pp. 7-12.

Celma O, Serra X (2008) FOAFing the music: Bridging the semantic gap in music recommendation. Web Semantics: Science, Services and Agents on the World Wide Web 6(4), pp. 250-256.

Cena F, Likavec S, Osborne F (2013) Anisotropic propagation of user interests in ontology-based user models. Information Sciences, 250, pp. 40-60.

Cha KJ (2014) Experimental Evaluation of an Expert System for Travel Recommender Systems. International Journal of Software Engineering and Its Applications. 8(4), pp. 115-128.

CISCO (2014) The Zettabyte Era—Trends and Analysis. [Online] Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html [Last accessed: 15-Feb-2015].

Codina V, Ceccaroni L (2010) Taking Advantage of Semantics in Recommendation Systems. R. Alquézar, A. Moreno, J. Aguilar (Eds.), IOS Press, Amsterdam, The Netherlands, The Netherlands, pp. 163-172.

Coelho B, Martins C, Almeida A (2009) Adaptive Tourism Modeling and Socialization System. In CSE '09 Proceedings of the 2009 International Conference on Computational Science and Engineering, 4, pp. 645-652.

Collins T, Mulholland P, Zdrahal Z (2005) Semantic browsing of digital collections, The semantic Web, ISWC, Lecture Notes in Computer Science, 3729, pp. 127-141.

Dasarathy B (1991) Nearest neighbour (NN) norms: NN pattern classification techniques. IEEE Computer Society Press.

De la Flor J, Borràs J, Isern D, Valls A, Moreno A, Russo A, Pérez Y, Anton-Clavé S (2012) Semantic Enrichment for Geospatial Information in a Tourism Recommender System. Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications, pp. 134-156. Eds: L.Díaz, C.Granell, J.Huerta. IGI-Global.

Debattista J, Scerri S, Rivera I, Handschuh S (2012) Ontology-based Rules for Recommender Systems. In SeRSy, pp. 49-60.

Del Vasto-Terrientes L, Valls A, Zielniewicz P, Borràs J. (2015) A hierarchical multi-criteria sorting approach for recommender systems. Submitted to Journal of Intelligent Information Systems (in Press).

Desrosiers C, Karypis G (2011) A Comprehensive Survey of Neighborhood-based Recommendation Methods. In: Ricci F, Rokach L, Shapira B, Kantor PB, editors. Recommender Systems Handbook: Springer; 2011, pp. 107-144.

Dey A, Abowd G (1999) Towards a Better Understanding of Context and Context-Awareness. In HUC '99: Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing, pp. 304-307, Springer Berlin Heidelberg.

Di Noia T, Mirizzi R, Ostuni VC, Romito D (2012) Exploiting the web of data in model-based recommender systems. InProceedings of the sixth ACM conference on Recommender systems, pp. 253-256, ACM.

Ding C, He X (2004) K-means Clustering via Principal Component Analysis. In: 21st International Conference on Machine Learning, ICML 2004. Banff, Canada, 2004. pp. 29-38.

Dongxing C, Zhanjun L, Karthik R (2011) Ontology-based customer preference modeling for concept generation. Advanced Engineering Informatics 25(2), pp. 162-176.

Dujmović JJ, Nagashima H (2006) LSP method and its use for evaluation of Java IDEs. International Journal of Approximate Reasoning, 41, pp. 3-22.

Expedia (2013) The Traveler's Path to Purchase. [Online]. Available: http://info.advertising.expedia.com/path-to-purchase [Last accessed: 22-Jan-2015].

Felfernig A, Gordea S, Jannach D, Teppan E, Zanker M (2007) A short survey of recommendation technologies in travel and tourism. OEGAI Journal, 25(7), pp. 17-22.

Fenza G, Fischetti E, Furno D, Loia V (2011) A hybrid context aware system for tourist guidance based on collaborative filtering. Proceedings of IEEE Int. Conference on Fuzzy Systems, pp. 131-138.

Figueira J, Mousseau V, Roy B (2005) Electre Methods. In J. Figueira, S. Greco & M. Ehrogott, Multicriteria decision analysis: State of the Art Surveys. New York: Springer.

Figueira JR, Greco S, Roy B, Slowinski R (2013) An Overview of ELECTRE Methods and their Recent Extensions. Journal of Multi-Criteria Decision Analysis, 20, pp. 61-85, 2013.

Gan M, Jiang R (2013) Improving accuracy and diversity of personalized recommendation through power law adjustments of user similarities. Decision Support Systems, 55(3), pp. 811-821.

Gao M, Liu K, Wu Z (2010) Personalisation in web computing and informatics: theories, techniques, applications, and future research. Information Systems Frontiers, 12, pp. 607-629.

Garcia A, Arbelaitz O, Linaza MT, Vansteenwegen P, Souffriau W (2010) Personalized Tourist Route Generation. In F. Daniel and F.M. Facca (Eds.), 10th International Conference on Web Engineering 2010, ICWE2010, LNCS 6385, pp. 486-497.

Garcia A, Torre I, Linaza, MT (2013a) Mobile Social Travel Recommender System. Information and Communication Technologies in Tourism 2014. Springer International Publishing, pp. 3-16.

Garcia A, Vansteenwegen P, Arbelaitz O, Souffriau W, Linaza MT (2013b) Integrating Public Transportation in Personalised Electronic Tourist Guides. Computers & Operations Research, Special issue: Transport Scheduling, 3, pp. 758-774.

Garcia I, Sebastià L, Onaindia E (2011) On the design of individual and group recommender systems for Tourism. Expert Systems with Applications, 38, pp. 7683-7692.

García-Crespo A, Chamizo J, Rivera I, Mencke M, Colomo-Palacios R, Gómez-Berbís JM (2009) SPETA: Social pervasive e-Tourism advisor. Telematics and Informatics, 26, pp. 306-315.

García-Crespo A, López-Cuadrado JL, Colomo-Palacios R, González-Carrasco I, Ruiz-Mezcua B (2011) Sem-Fit: a semantic based expert system to provide recommendations in the tourism domain. Expert Systems with Applications, 38(10) pp. 13310-13319.

García-Lapresta J.L, Marley AAJ, Martínez-Panero M (2010) Characterizing best-worst voting systems in the scoring context. Social Choice and Welfare 34, pp. 487-496.

Gavalas D, Kenteris M (2011) A web-based pervasive recommendation system for mobile tourist guides. Personal and Ubiquituous Computing 15, pp. 759-770.

Ge M, Delgado-Battenfeld C, Jannach D (2010) Beyond accuracy: evaluating recommender systems by coverage and serendipity. In Proceedings of the fourth ACM conference on Recommender systems, pp. 257-260, ACM.

Gibert K, Cortés U (1997) Weighting Quantitative and Qualitative Variables in Clustering Methods, Mathware and Soft Computing 1997, 4, pp. 251-266.

Google (2014) The 2014 Traveler's Road to Decision. [Online]. Available: https://think.storage.googleapis.com/docs/2014-travelers-road-to-decision_research_studies.pdf. [Last accessed: 22-Jan-2015].

Gretzel U (2011) Intelligent systems in tourism: A social science perspective. Annals of Tourism Research, 38(3), pp. 757-779.

Gretzel U, Fesenmaier DR, O"Leary JT (2006) The transformation of consumer behaviour. In Tourism Business Frontiers: Consumers, Products and Industry edited by D. Buhalis & C. Costa. Burlington, MA: Elsevier, pp 9-18.

Gursoy D, McCleary KW (2004) AN INTEGRATIVE MODEL OF TOURISTS'INFORMATION SEARCH BEHAVIOR. Annals of tourism research, 31(2), pp. 353-373.

Gyorodi R, Gyorodi C, Dersidan M (2013) An Extended Recommendation System using Data Mining Implemented for Smart Phones. International Journal of Computers and Technology, 11(3), pp. 2360-2372.

Hagen K, Kramer R, Hermkes M, Schumann B, Mueller P (2005) Semantic matching and heuristic search for a dynamic tour guide. Information and Communication Technologies in Tourism 2005, A. Frew et al. (eds.), Springer Computer Science.

Han J, Lee H (2014) Adaptive landmark recommendations for travel planning: Personalizing and clustering landmarks using geo-tagged social media, Pervasive and Mobile Computing.

Hartigan JA, Wong MA (1979) Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society, Series C (Applied Statistics) 28, 1, pp. 100-108.

Heckmann D, Schwarzkopf E, Mori J, Dengler D, Kr A (2007) The User Model and Context Ontology GUMO revisited for future Web 2.0 Extensions. Proc. Contexts and Ontologies: Representation and Reasoning, 2007, pp. 37-46.

Hepp M, Siorpaes K, Bachlechner D (2006) Towards the Semantic Web in E-Tourism: Can Annotation Do the Trick? In: Fourteenth European Conference on Information Systems, ECIS 2006. Gothenburg, Sweden; 2006. pp. 2362-2373.

Herzog D, Wörndl W (2014) A Travel Recommender System for Combining Multiple Travel Regions to a Composite Trip. CBRecSys 2014, 42.

Heu F, Lin Y, Ho T (2012) Design and implementation of an intelligent recommendation system for tourist attractions: the integration of EBM model, Bayesian networks and Google Maps, Expert Systems With Applications 2012, 39, pp. 3257-3264.

Hsu FM, Lin YT, Ho TK (2012) Design and implementation of an intelligent recommendation system for tourist attractions: the integration of EBM model, Bayesian networks and Google Maps. Expert Systems with Applications, 39, pp. 3257-3264.

Huang Y, Bian L (2009) A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet. Expert Systems with Applications, 36(1), pp. 933-943.

Iaquinta L, Gemmis M, Lops P, Semeraro G, Molino P (2010) Can a recommender system induce serendipitous encounters? E-commerce, IN-TECH, Vienna, pp. 229–246

Jannach D, Zanker M, Jessenitschnig M (2010) Developing Knowledge-Based Travel Advisor Systems: A Case Study. In N. Sharda (Ed.), Tourism Informatics: Visual Travel Recommender Systems, Social Communities, and User Interface Design, pp. 38-53. Hershey, PA: Information Science Reference.

Jiang J, Conrath D (1997) Semantic similarity based on corpus statistics and lexical taxonomy. In: Proceedings of the International Conference on Research in Computational Linguistics (ROCLING X). Taiwan, 1997. pp. 19–33.

Jiang X, Tan A (2009) Learning and inferencing in user ontology for personalized Semantic Web search. Information Sciences, 179, pp. 2794–2808.

Jin R, Chai JY, Si L (2004) An automatic weighting scheme for collaborative filtering. In: 27th Annual International ACM SIGIR conference on Research and development in information retrieval. Sheffield, UK, 2004. pp. 337-344.

Kabassi K (2010) Personalizing recommendations for tourists. Telematics and Informatics, 27(1), pp. 51-66.

Kelly D, Teevan J (2006) Implicit feedback for inferring user preference: a bibliography, ACM SIGIR Forum 2006, 37, pp. 18-28.

Koceski S, Petrevska B (2012) Empirical Evidence of Contribution to E-Tourism by Application of Personalized Tourism Recommendation System. Annals of the Alexandru Ioan Cuza University-Economics 59(1), pp. 363-374.

Kruszyk M, Ganzha M, Gawinecki M, Paprzycki M (2007) Introducing Collaborative Filtering into an Agent-Based Travel Support System, in: Proc. of. IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops, Silicon Valley, California, USA, pp. 439-443.

Kurata Y (2011) CT-Planner2: More Flexible and Interactive Assistance for Day Tour Planning. ENTER 2011, Information and Communication Technologies in Tourism, pp. 25-37, Innsbruck, Austria, Jan, 2011.

Kurata Y, Hara T (2013) CT-Planner4: Toward a More User-Friendly Interactive Day-Tour Planner Information and Communication Technologies in Tourism 2014. Springer International Publishing, pp. 73-86.

Lamsfus C, Alzua-Sorzabal A, Martin D, Salvador Z, Usandizaga A (2009) Human-centric Ontology-based Context Modelling in Tourism. In Proceedings of the International

Conference on Knowledge Engineering and Ontology Development, Funchal, Madeira, Portugal, October 6-8, 2009, pp. 424-434.

Lamsfus C, Alzua-Sorzabal A, Martin D, Smithers T (2011) An Evaluation of a Contextual Approach to Visitor Information System. Proceeding of the ENTER conference. Eds: Rob Law, Matthias Fuchs and Francesco Ricci. Springer Verlag, pp. 191-202. Innsbruck, Austria. January 2011, pp.179-189.

Lamsfus C, Grün C, Alzua-Sorzabal A, Werthner H (2010) Context-based matchmaking to enhance tourists’ experience. Journal for the Informatics Professional, CEPIS-Upgrade Journal, nr. 203, pp. 17-23. May, 2010.

Lee CS, Chang YC, Wang MH (2009) Ontological recommendation multi-agent for Tainan city travel. Expert Systems with Applications, 36, pp. 6740-6753.

Lemos ML, Vasquez DV, Radzimski M, Lemos AL, Gómez-Berbís JM (2012) RING: A Context Ontology for Communication Channel Rule-based Recommender System. SeRSy 2012, 73.

Linaza MT, Aguirregoikoa A, García A, Torres JI, Aranburu K (2011) Image-based Travel Recommender System for small tourist destinations. Information and Communication Technologies in Tourism 2011, pp. 1-12. Innsbruck, Austria, 2011.

Linden G, Smith B, York J (2003) Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing. 7(1), pp. 76-80.

Lorenzi F, Loh S, Abel M (2011) PersonalTour: A Recommender System for Travel Packages. Proceedings of the 2011 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT 2011, Lyon, France, August 22-27. IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.

Luberg A, Tammet T, Järv P (2011) Smart City: A Rule-based Tourist Recommendation System. In Information and Communication Technologies in Tourism 2011. ENTER 11, January 26th-28th 2011, Innsbruck, Austria. (Eds.) Law, R.; Fuchs, M.; Ricci, F. Springer-Verlag.

Lucas J, Laurent A, Moreno M, Teisseire M (2009) Fuzz-CBA: Classification à base de règles d'association floues et systèmes de recommandation. Rencontres Francophones sur la Logique Floue et ses Applications, pp. 283-290.

Lucas J, Luz N, Moreno M, Anacleto R, Figueiredo A, Martins C (2013) A hybrid recommendation approach for a tourism system. Expert Systems with Applications, 40(9), pp. 3532-3550.

Manouselis N, Costopoulou C (2007) Analysis and Classification of Multi-Criteria Recommender Systems. World Wide Web, 10, 4, pp. 415-441.

Marin L, Isern D, Moreno A, Valls A (2013) On-line dynamic adaptation of fuzzy preferences. Information Sciences, 220, pp. 5-12.

Marques SS (2009) Fabrication of image-destination: the impact of stereotypes at the level of tourism. Revista Turismo & Desenvolvimento, 12.

Martin D, Alzua A, Lamsfus C (2011) A Contextual Geofencing Mobile Tourism Service. Proceeding of the ENTER conference. Eds: Rob Law, Matthias Fuchs and Francesco Ricci. Springer Verlag, pp. 191-202. Innsbruck, Austria. January 2011.

Martínez L, Rodríguez RM, Espinilla M (2009) REJA: a georeferenced hybrid recommender system for restaurants. Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 187-190.

Martínez S, Sánchez D, Valls A, Batet M (2012) Privacy protection of textual attributes through a semantic-based masking method, Information Fusion. Special Issue on Privacy and Security, 13, pp. 304-314 (October 2012).

Martínez-Santiago F, Ariza-López F, Montejo-Ráez A, Ureña-López A (2012) GeOasis: a knowledge-based geo-referenced tourist assistant. Expert Systems with Applications, 39(14), pp. 11737-11745.

Mcnee S, Riedl J, Konstan J (2006) Accurate is not always good: How Accuracy Metrics have hurt Recommender Systems, Conference on Human Factors in Computing Systems, Quebec, Canada. pp. 1-5.

Meehan K, Lunney T, Curran K, McCaughey A (2013) Context-Aware Intelligent Recommendation System for Tourism. Proceedings of the 11th IEEE International Conference on Pervasive Computing and Communications, pp. 328-331.

Middleton SE, Roure DC, Shadbolt NR (2009) Ontology-Based Recommender Systems. Handbook on Ontologies, International Handbooks on Information Systems Series, Springer, 2009, 779-796.

Mínguez I, Berrueta D, Polo L (2010) CRUZAR: An Application of Semantic Matchmaking to e-Tourism. In Y. Kalfoglou (Ed.), Cases on Semantic Interoperability for Information Systems Integration: Practices and Applications, pp. 255-271. Hershey, PA: Information Science Reference.

Mobasher B (2007) Data mining for web personalization, in: P. Brusilovsky, A. Kobsa, W. Nejdl (eds.), The Adaptive Web: Methods and Strategies of Web Personalization (Springer Verlag, 2007) 90-135.

Montaner M, López B, de la Rosa JL (2003) A Taxonomy of Recommender Agents on the Internet. Artificial Intelligence Review, 19(3), pp. 285-330.

Montejo-Ráez A, Perea-Ortega JM, García-Cumbreras MA, Martínez-Santiago F (2011) Otium: a web based planner for tourism and leisure, Expert Systems with Applications, 38, pp. 10085-10093.

Moreno A, Valls A, Isern D, Marin L, Borràs J (2013a) Sigtur/e-destination: ontology-based personalized recommendation of tourism and leisure activities. Engineering Applications of Artificial Intelligence, 26(1), pp. 633-651.

Moreno A, Valls A, Mata F, Martínez S, Marín L, Vicient C (2013b) A semantic similarity measure for objects described with multi-valued categorical attributes, Artificial Intelligence Research and Development, Frontiers in Artificial Intelligence, IOS Press, 256, pp. 263-272.

Moscato V, Picariello A, Rinaldi AM (2013) Towards a user based recommendation strategy for digital ecosystems. Knowledge-Based Systems, 37, pp. 165-175.

Niaraki AS, Kim K (2009) Ontology based personalized route planning system using a multi-criteria decision making approach. Expert Systems with Applications, 36, pp. 2250-2259.

Niemann K, Wolpers M (2013) A new collaborative filtering approach for increasing the aggregate diversity of recommender systems. In Proceedings of the 19th ACM

SIGKDD international conference on Knowledge discovery and data mining, pp. 955-963, ACM.

Noel J, Sanner S, Tan K-N, Christen P, Xie L, Bonilla EV, Abbasnejad E, Della Penna N (2012) New objective functions for social collaborative filtering. Proc. 21st Int. Conf. on World Wide Web, pp. 859-868.

Noguera JM, Barranco MJ, Segura RJ, Martínez L (2012) A mobile 3D-GIS hybrid recommender system for tourism. Information Sciences, 215, pp. 37-52.

Ou S, Pekar V, Orasan C, Spurk C, Negri M (2008) Development and Alignment of a Domain-Specific Ontology for Question Answering. In: International Conference on Language Resources and Evaluation, LREC 2008. Marrakech, Morocco, 2008, pp. 2221-2228.

Partarakis N, Doulgeraki C, Leonidis A, Antona M, Stephanidis C (2009) User Interface Adaptation of Web-Based Services on the Semantic Web. Proceedings of the 5th International on ConferenceUniversal Access in Human-Computer Interaction. Part II: Intelligent and Ubiquitous Interaction Environments, July 19-24, 2009, San Diego, CA doi:10.1007/978-3-642-02710-9_79.

Parundekar R, Oguchi K (2012) Driver Recommendations of POIs using a Semantic Content-based Approach. In SeRSy, pp. 13-24.

Pazzani MJ, Billsus D (2007) Content-Based Recommendation Systems. In: The Adaptive Web: Methods and Strategies of Web Personalization: Springer Verlag; 2007, pp. 325-341.

Pearl J (1988) Probabilistic reasoning in intelligent systems. Networks of plausible inference. Morgan Kaufmann.

Pinho J, da silva B, Moreno M, de Almeida AM, Lopes C (2011) Applying recommender methodologies in Tourism sector. Highlights in Practical Applications of Agents and Multiagent Systems. Advances in Intelligent and Soft Computing, 89, pp. 101-108, Springer.

Pirlot M (1997) A common framework for describing some outranking methods. Journal of Multi-Criteria Decision Analysis, 6(2), 86-92.

Pisanelli DM (Ed.) (2004) Ontologies in Medicine. IOS Press, 102, of Studies in Health Technology and Informatics, ISBN: 978-1-58603-418-4

Prantner P, Ding Y, Luger M, Yan Z, Herzog C (2007) Tourism Ontology and Semantic Management System: State-of-the-arts Analysis. In: IADIS International Conference WWW/Internet 2007. Vila Real, Portugal, 2007, pp. 111-115.

Resnik P (1995) Using information content to evaluate semantic similarity in a taxonomy. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI 95). Montreal, Canada, pp. 448-453.

Rey-López M, Barragáns-Martínez AB, Peleteiro A, Mikic-Fonte FA, Burguillo JC (2011) moreTourism: mobile recommendations for Tourism. IEEE Int. Conf. on Consumer Electronics, pp. 347-348.

Ricci F (2002) Travel recommender systems. IEEE Intelligent Systems 17(6), pp. 55-57.

Ricci F, Nguyen QN, Averjanova O (2010) Exploiting a Map-Based Interface in Conversational Recommender Systems for Mobile Travelers. In N. Sharda (Ed.), Tourism Informatics: Visual Travel Recommender Systems, Social Communities, and User Interface Design, pp. 73-93. Hershey, PA: Information Science Reference.

Rojas G, Uribe C (2013) A Conceptual Framework to develop Mobile Recommender Systems of Points of Interest. International Workshop on Advanced Software Engineering. Proceedings Jornadas Chilenas de Computación, Chile, November 2013.

Rospocher M, Serafini L (2012) Ontology-centric Decision Support. In SeRSy, pp. 61-72.

Roy B (1991) The outranking approach and the foundations of ELECTRE methods. Theory and decision, 31(1), pp. 49-73.

Ruíz-Martínez JM, Miñarro JA, Castellanos D, García F, Valencia R (2011) Ontology population: an application for the e-Tourism domain. International Journal of Innovative Computing, Information and Control, 7(119), pp. 1-19.

Ruiz-Montiel M, Aldana-Montes J (2009) Semantically enhanced recommender systems. On the Move to Meaningful Internet Systems: OTM 2009 Workshops, 5872, pp. 604-609, Springer.

Ruiz-Montiel M, Molina-Castro JJ, Aldana-Montes JF (2010) TasTicWiki: a semantic Wiki with content recommendation, Proceedings of the Fifth Workshop on Semantic Wikis-Linking Data and People. 7th Extended Semantic Web Conference. pp. 31-40.

Ruotsalo T, Haav K, Stoyanov A, Rochee S, Fanid E, Deliaic R, Mäkelä E, Kauppinen T, Hyvönen E (2013) SMARTMUSEUM: A Mobile Recommender System for the Web of Data. Web Semantics: Science, Services and Agents on the World Wide Web, 20, 50-67.

Saaty TL (1980) The analytic hierarchy process. New York, NY: McGraw-Hill.

Sánchez D, Batet M, Valls A, Gibert K (2010) Ontology-driven web-based semantic similarity. Journal of Intelligent Information Systems, 35, pp. 383-413.

Sánchez D, Moreno A (2008a) Pattern-based automatic taxonomy learning from the Web. AI Communications, 21(1), pp. 27-48.

Sánchez D, Moreno A (2008b) Learning non-taxonomic relationships from web documents for domain ontology construction. Data and Knowledge Engineering Journal, 64, pp. 600-623.

Sánchez D, Moreno A, Del Vasto-Terrientes L (2012) Learning relation axioms from text: an automatic Web-based approach. Expert Systems with Applications, 39, pp. 5792-5805.

Savir A, Brafman R, Shani G (2013) Recommending improved configurations for complex objects with an application in travel planning. Proceedings of the 7th ACM conference on Recommender systems. ACM, 2013, pp. 391-394.

Schafer JB, Konstan J, Riedl J (1999) Recommender systems in e-commerce. Proceedings of the 1st ACM conference on Electronic commerce, pp. 158-166, ACM.

Sebastià L, Garcia I, Onaindia E, Guzman C (2009) e-Tourism: a tourist recommendation and planning application. International Journal on Artificial Intelligence Tools 18(5), pp. 717-738.

Sebastià L, Giret A, Garcia I (2010) A multi agent architecture for Tourism recommendation. Trends in Practical applications of Agents and Multiagent Systems. Advances in Intelligent and Soft Computing 71, pp. 547-554.

Seidel I, Gärtner M, Pöttler M, Berger H, Dittenbach M (2009) Itchy Feet: A 3D E-Tourism Environment. In Tourism Informatics: Visual Travel Recommender Systems, Social Communities, and User Interface Design, N. Sharda Ed. IGI Global, Hershey, PA.

Sendhilkumar S, Geetha TV (2008) Personalized ontology for web search personalization. In Proceedings of the 1st Bangalore annual Compute conference, Bangalore, India, 18–20 January 2008; ACM: New York, NY, USA, 2008, pp. 1–7.

Shambour Q, Lu J (2011) A Hybrid Multi-criteria Semantic-Enhanced Collaborative Filtering Approach for Personalized Recommendations. Web Intelligence 2011, pp. 71-78.

Shoval P, Maidel V, Shapira B (2008) An ontology-content-based filtering method. International Journal of Information Theories and Applications, 15(4), pp. 51–63.

Sieg A, Mobasher B, Burke R (2007) Learning Ontology-Based User Profiles: A Semantic Approach to Personalized Web Search. IEEE Intelligent Informatics Bulletin, 8(1), pp. 7-18.

Smyth B and McClave P (2001) Similarity vs. diversity. In D.W. Aha, & I. Watson (Eds.), Proceedings of the 4th international conference on case-based reasoning. Lecture Notes in Computer Science, 2080, pp. 347–361, Vancouver: Springer.

Souffriau W, Vansteenwegen P, Vanden Berghe G, Van Oudheusden D (2011) The planning of cycle trips in the province of East Flanders. Omega 39, pp. 209-213.

Sriharee G (2014) An ontology-based approach to auto-tagging articles.Vietnam Journal of Computer Science, pp. 1-10.

Staab S, Werthner H (2002) Intelligent systems for tourism. IEEE Intelligent Systems, 17(6), pp. 53-64.

Studer R, Benjamins VR, Fensel D (1998) Knowledge Engineering: Principles and Methods. Data Knowl. Eng., 25(1-2), pp. 161-197.

Swarbrooke J, Horner S (2007) Consumer behaviour in tourism. Routledge.

Torra V, Narukawa Y (2007) Modeling Decisions: Information Fusion and Aggregation Operators, 978-3-540-68789-4, Springer Berlin, Heidelberg.

TripAdvisor (2014) 24 insights to shape your TripAdvisor strategy. [Online infographic] Available: http://www.tripadvisor.com/TripAdvisorInsights/n2120/24-insights-shape-your-tripadvisor-strategy [Last accessed: 15-Feb-2015].

Tsung-Chiung W, Chyong-Ru L, Wan-Chen Y (2012) Segmenting indigenous tourists from a serious leisure perspective. Journal of Vacation Marketing, 18(1), pp. 17-29.

Umanets A, Ferreira A, Leite N (2013) GUIDEME - A Tourist Guide with a Recommender System and Social Interaction. Conference on Electronics, Telecommunications and Computer (CETC), Lisboa, December 2013.

Valls A (2003) ClusDM: a multiple criteria decision method for heterogeneous data sets. PhD thesis. AI Communications 16, pp. 129-130.

Valls A, Moreno A, Borràs J (2013) Preference representation with ontologies. Multicriteria Decision Aid and Artificial Intelligence, pp. 77-99.

Vansteenwegen P, Souffriau W (2011) Trip planning functionalities: state-of-the-art and future. Information Technology and Tourism, 12(4), pp. 305-315.

Vansteenwegen P, Souffriau W, Vanden Berghe G, Van Oudheusden D (2010) The city trip planner: an expert system for tourists. Expert Systems with Applications, 38(6), pp. 6540-6546.

Vargas S, Castells P, Vallet D (2011) Intent-oriented diversity in recommender systems. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, ACM, pp. 1211-1212.

Venkataiaha S, Shardaa N, Ponnadaa M (2008) A comparative study of continuous and discrete visualisation of tourism information. ENTER 2008 Conference, International Federation for Information Technology and Travel & Tourism (IFITT), Innsbruck, Austria, 23-25 January 2008.

Vicient C, Sánchez D, Moreno A (2013) An automatic approach for ontology-based feature extraction from heterogeneous textual resources. Engineering Applications of Artificial Intelligence, 26(3), pp. 1093-1106.

Wang RQ, Kong FS (2007) Semantic-Enhanced Personalized Recommender System. Machine Learning and Cybernetics: Proc. of the Int. Conference on Machine Learning and Cybernetics, 7(19-22), pp. 4069-4074.

Wang W, Zeng G, Tang D (2011) Bayesian intelligent semantic mashup for Tourism. Concurrency and computation: practice and experience, 23, pp. 850-862.

Wang XH, Zhang DQ, Gu T, Pung HK (2004) Ontology based context modeling and reasoning using OWL. In: Pervasive Computing and Communications Workshops in Proceedings of the Second IEEE Annual Conference, pp. 18–22.

Wei M, Ling Ma, Chang W (2014) Personalized recommendation of mobile tourism: a multidimensional user model. Pacific Asia Conference on Information Systems.

Werthner H (2003) Intelligent systems in travel and tourism. In IJCAI, 3, pp. 1620-1625.

Wooldridge M (2009) An introduction to multiagent systems. John Wiley and Sons.

Yager RR (1988) On ordered weighted averaging operators. Operators in multi-criteria decision making. IEEE Transactions on Systems, Man and Cybernetics, 18, pp. 183-190.

Yager RR (1996) Quantifier guided aggregation using OWA operators, International Journal of Intelligent Systems 1996, 11, pp. 49–73.

Yang L, Hu Z, Long J (2010) Service of searching and ranking in a semantic-based expert information system. Proc. IEEE Asia-Pacific Services Computing Conference, pp. 609-614.

Yang S (2010a) Developing an ontology-supported information integration and recommendation system for scholars. Expert Syst. Appl. 37(10), pp. 7065-7079.

Yang WS, Hwang SY (2013) iTravel: a recommender system in mobile peer-to-peer environment. The Journal of Systems and Software, 86(1), pp. 12-20.

Yang Y (2010b) Semantic User Model Inferences for Travel Recommender Systems. In N. Sharda (Ed.), Tourism Informatics: Visual Travel Recommender Systems, Social Communities, and User Interface Design, pp. 23-37. Hershey, PA: Information Science Reference.

Yu C, Chang H (2009) Personalized location-based recommendation services for tour planning in mobile tourism applications. Lecture Notes in Computer Science, 5692, Proceedings of the 10th International Conference on E-commerce and Web Technologies, Linz, Austria, September 1-4, pp. 1-49.

Zhang M, Hurley N (2008) Avoiding monotony: improving the diversity of recommendation lists. Proceedings of the 2008 ACM conference on Recommender systems, pp. 123-130.

Zheng Z (2011) A personalized product recommender system based on semantic similarity and TOPSIS method. Journal of Convergence Information Technology, 6(7), pp. 312-320.

Zhou T, Kuscsik Z, Liu JG, Medo M, Wakeling JR, Zhang YC (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. Proceedings of the National Academy of Sciences, 107(10), pp. 4511-4515.

Ziegler CN, McNee S, Konstan J, Lausen G (2005) Improving Recommendation List Through Topic Diversification, in: Proc. of. Fourteenth International World Wide Web Conference, WWW 2004, Chiba, Japan, pp. 22-32.