FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Ontology-based Matchmaking to Provide Personalized Offers

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Sozial- und Wirtschaftswissenschaften

eingereicht von

Mag.rer.soc.oec. Christoph Grün
Matrikelnummer 9855025

an der Fakultät für Informatik
der Technischen Universität Wien

Betreuung: Univ.-Prof. Dipl.-Ing. Dr. Hannes Werthner

Diese Dissertation haben begutachtet:

Prof. Dr. Markus Zanker

Prof. Dr. Matthias Fuchs

Wien, 21. November 2016

Christoph Grün

FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Ontology-based Matchmaking to Provide Personalized Offers

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doktor der Sozial- und Wirtschaftswissenschaften

by

Mag.rer.soc.oec. Christoph Grün
Registration Number 9855025

to the Faculty of Informatics
at the TU Wien
Advisor: Univ.-Prof. Dipl.-Ing. Dr. Hannes Werthner

The dissertation has been reviewed by:

Prof. Dr. Markus Zanker	Prof. Dr. Matthias Fuchs

Vienna, 21st November, 2016

Christoph Grün

Erklärung zur Verfassung der Arbeit

Mag.rer.soc.oec. Christoph Grün
Pasettistrasse 75/1, 1200 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 21. November 2016

Christoph Grün

Acknowledgements

During the process of writing this thesis, I have received inspirations and support from numerous people.

First of all, I would like to thank Prof. Hannes Werthner for being my supervisor of the thesis. Thank you for giving me the opportunity to work in the academic field and especially to support me to do research in the exciting field of e-tourism. You supported me throughout the last years and encouraged me multiple times to finish the PhD thesis - without your encouragement, I would not have reached this goal. Your door was always open to discuss any problems or issues in the course of the PhD work which I appreciate a lot. Moreover, I would like to thank you for providing me the possibility to visit different conferences, summer and winter schools in order to discuss and exchange ideas with the research community - which I think is a critical aspect when doing the PhD.

During my studies, I had the pleasure to share the office with Birgit Dippelreiter, Michael Pöttler and Marco Zapletal. Thank you for all the discussions and feedback you provided me.

I further would like to thank all members of the EC group for their valuable comments and hints on many topics: Prof. Dieter Merkl and Prof. Jürgen Dorn, Amra Delic, Laszlo Grad-Gyenge, Birgit Hofreiter, Thanh Tran Thi Kim, Rozeia Mustafa, Dimitris Sacharidis, Gudrun Salomon, and Rainer Schuster. Special thanks go to Julia Neidhardt and Martin Hochmeister, with whom I had endless discussions on ontologies, semantic matchmaking and PhD issues in general.

I want to acknowledge all colleagues who were willing to act as co-author in writing research proposals and papers. Good ideas, improvements and new insights always stem from discussing things together and it was fantastic to see how each one's contribution helped to form a collective work.

Thanks to all the anonymous testers who were kind to participate in the evaluation of the prototypical application and fill in the questionnaire.

Last but not least, I would like to thank my family, relatives and friends for all the support and patience, especially to my parents Helga and Helmut, my brother and sisters and to my girl-friend Barbara who have stood with me during the PhD process.

Abstract

This thesis addresses the challenge to provide personalized offers to users. The level of personalization directly depends on the quality of the matchmaking process, i.e. finding those offers that are most attractive to a particular user based on his or her interests. We study the matchmaking process in the context of e-tourism. The goal is to support tourists in their decision-making during the pre-trip phase and to facilitate the process of identifying those tourism objects of a specific destination that best fit the tourists' preferences.

To achieve this goal, we propose an iterative matchmaking process that matches tourist profiles with the characteristics of tourism objects in order to obtain a ranked list of appropriate tourism objects for a particular tourist. The matchmaking process is composed by two main steps.

For the first step, we devise a stereotype approach based on tourist types to model a basic user profile, reflecting tourists' preferences according to travel-related categories (e.g., culture, sightseeing, sports or nature). This profile is then related against the generic characteristics of the tourism objects in order to recommend a top-N list of tourism objects. The generic characteristics of the tourism objects are modeled based on the same typology. To continuously enhance the quality of the matchmaking process, we exploit tourist feedback to dynamically adapt and refine tourist profiles (e.g., a tourist may be a cultural type but may express a dislike of museums).

This is achieved by the second step. Its task is to consider positive/negative tourist feedback in order to derive specific interests and to re-adapt the matches between one particular tourist and the set of relevant tourism objects by taking into account his/her specific interests. We develop a tourism ontology and use it as basis in order to model the specific interests of the tourists and the specific attributes of the tourism objects.

The first and second matchmaking steps can be combined. Tourists can criticize the proposed items by stating positive/negative feedback, which will be used to refine their profiles and to deliver a new set of tourism objects. As long as they are not satisfied with the recommendations, they can repeat this process.

Our approach is tested through a prototypical recommendation system that recommends tourists visiting Vienna appropriate tourism attractions that are tailored to their personal needs. We conduct a user study by asking users to interact with the system and fill in a questionnaire afterwards.

Kurzfassung

Diese Arbeit befasst sich mit der Thematik Nutzern personalisierte Angebote bereitzustellen. Zentrales Element dafür ist das Zusammenführen von Nutzerpräferenzen mit Angeboten (Matching), um aus dem Gesamtangebot geeignete Objekte auszuwählen. Als Anwendungsbereich dient der Tourismussektor. Ziel ist, Touristen in der Phase der Entscheidungsfindung zu unterstützen, indem ausgewählte Sehenswürdigkeiten einer Tourismusregion vorgeschlagen werden.

Um dieses Ziel zu erreichen stellen wir einen iterativen Matching-Prozess vor, der das Interessenprofil eines Touristen mit dem Profil der Sehenswürdigkeiten einer Region vergleicht und eine Rangliste an geeigneten Sehenswürdigkeiten generiert. Der Prozess wird in zwei Schritte unterteilt.

Im ersten Schritt werden die generischen Präferenzen eines Touristen anhand von Stereotypen ermittelt und in einem Benutzerprofil erfasst. Die Merkmale der Stereotypen basieren auf einer bestehenden Typologie von Touristen und bilden die Präferenzen anhand von Kategorien wie Sightseeing, Kultur, Sport, Natur, etc. ab. Dieses Profil wird einem generischen Profil der Sehenswürdigkeiten gegenübergestellt, um eine Liste mit Top-N Sehenswürdigkeiten zu erstellen. Das generische Profil der Sehenswürdigkeiten wird anhand derselben Typologie erstellt. Touristen können eine Bewertung zu den vorgeschlagenen Sehenswürdigkeiten abgeben, die in die Verbesserung ihres Profils einfließt (z.B. kann ein Tourist ein Kulturliebhaber sein, aber keine Museen mögen) und somit die Qualität des Matchings kontinuierlich verbessert.

Die Berücksichtigung dieser Bewertungen erfolgt im zweiten Schritt. Anhand des positiven/negativen Nutzer-Feedbacks werden spezifische Interessen abgeleitet und das Benutzerprofil dynamisch erweitert. Für die Erstellung der spezifischen Profile des Touristen als auch der Sehenswürdigkeiten verwenden wir eine Tourismus-Ontologie, deren Konzepte für die Modellierung der spezifischen Interessen aber auch der qualitativen Eigenschaften von Sehenswürdigkeiten dienen.

Die beiden Schritte sind kombinierbar. Die initial vorgeschlagenen Sehenswürdigkeiten können bewertet werden. Die Bewertungen fließen in das Profil des Touristen ein und bewirken eine gezielte Verfeinerung der Vorschläge. Dieser Prozess kann beliebig oft durchgeführt werden, bis der jeweilige Tourist mit dem Ergebnis zufrieden ist.

Der Matching-Prozess wird prototypisch in einem Empfehlungssystem implementiert. Dieses empfiehlt potentiellen Wien-Touristen geeignete Sehenswürdigkeiten, die auf ihre persönlichen Bedürfnisse zugeschnitten sind. Die Evaluierung erfolgt im Rahmen einer Nutzerstudie, bei der die Teilnehmer aufgefordert werden, das System zu testen und im Anschluss einen Fragebogen auszufüllen.

Contents

1 Introduction

Formerly being regarded as a pure information search & presentation channel, the Internet is nowadays used to communicate with people (email or instant messaging), buy physical goods (books or clothing), digital goods (music or games) as well as purchase tickets for events or travel related items (flights or hotels) [42]. According to the Eurostat report titled “E-commerce statistics for individuals“ [1], 53% of all consumers aged 16-74 in the EU28 have purchased goods or services via the Internet in 2015. E-commerce is therefore widely used in the EU.

Information and Communication Technology (ICT) as a driver for travel and tourism.

Tourism is one of the most important domains in the World Wide Web [152]. This is confirmed by a survey by Universal McCann (2008), which reveals that the type of products/services Internet users worldwide most searched for is holidays/destinations (61.9%), followed by consumer electronics and travel items such as flights/trains (New Media Trend Watch). Hence, the tourism sector has gained enormously from the use of the Internet. Information technology has created an online travel market where tourism businesses are able to sell their travel products and communicate with their customers through electronic media [42]. On the other hand, the richness of information that is available online has empowered tourists to exploit the World Wide Web to search travel-related information and even partially book objects for their trip online. This way, both stakeholders (suppliers as well as consumers) benefit from the use of the Internet for information research and as (additional) selling channel. The provision and consumption of online travel services have become for both nearly a 'daily' business.

1.1 The tourist life cycle

In the ideal case, travel services should support tourists with travel-related information in the different stages of their travel process [64, 51]:

- travel decicision-making and anticipation
- travel to a tourism destination or attraction
- the on-site experience
- return travel and
- recollection of the experience and influcence on future decision-making

These stages can be mapped to the tourist life cycle [152], which is shown in (Figure 1.1) and consists of the pre-trip, on-trip and post-trip phases.

Pre-trip phase.

In the *pre-trip phase*, tourists need information for planning purposes and decision-making. During decision-making, only an abstract model of the product is avaible [153]. Therefore, gathering sufficient information

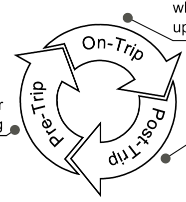


Figure 1.1 *The tourist life cycle and its phases [152].*

from a variety of information sources is a vital task for tourists to facilitate *travel decision making.*

Travel decision making is a complex process.

Trip planning is a process that includes several decision steps. According to a study by Zins [164], mode of transportation (84%) and travel companions (82%) is given highest attention, followed by accommodation and the destination. Decisions on accommodation, travel duration and destination require the most input from external information sources, comprising Internet sources (e.g., tour operator or travel agency Web sites) and traditional sources (brochures, friends or travel guides). Besides internal information (personal experience), information from friends and relatives are the most consulted sources.

Online travel agencies are used to find low prices.

As stated by [80], tourists heavily utilize the Internet to compare the prices of travel-related products from different online travel agencies, as they might offer similar products but with different prices. In fact, finding *low prices*, followed by *security* and *ease of use* are the most critical attributes when choosing an online travel agency.

User generated content is getting important.

In addition, online reviews are gaining more importance and influence tourists in their decision making. Yoo and Gretzel investigated in [58] the importance of reviews in the different stages of planning and point out that travel reviews are mostly used to *narrow down choices* in the middle of the planning process, but are also used for *idea generation* at the beginning of the process as well as to *confirm decisions* in later stages. At the same time, tourists increasingly participate in online travel communities (e.g., VirtualTourist.com, CouchSurfing.com, Lonelyplanet.com) as well as blogging communities (e.g., Travelpod.com, Travelblog.com). The main reasons for their participation are, besides social-psychological benefits (seeking identity, forming relationships) and hedonic benefits (enjoyment), information acquisition benefits as they can obtain up-to-date, freely available and trustworthy tourism-related content [157].

On-trip phase.

With the growth and evolution of mobile Internet-based devices, tourists increasingly use mobile information services in the *on-trip phase.* Such services are used for getting routing support while traveling to the destination but also at the destination itself as tourists act in unknown environments. There, they need personalized, up-to-date on-trip assistance in the form of information about accommodation, points of interest (POIs) (e.g., environmental and landscape attractions or gastronomy), flight delays, events weather forecasts, news or safety issues [60]. Mobile tourism services that can be used independently of temporal and spatial constraints and that are accessed through a mobile handset, may address these issues. Mobile information systems comprising services for the tourism domain are often referred to as *mobile tourist guides.* These

Mobile tourist guides address needs of tourists in the on-trip phase.

systems usually constitute a number of mobile services for supporting tourists during their destination stay [61]. Gretzel et al. state in [144] that the provision of such ubiquitous services will have a big impact on travel planning behavior and travel patterns. Due to mobile technologies, more travel decisions will be made while being on the move.

Post-trip phase.

In the *post-trip phase*, focus is on reminiscing about the journey and sharing the gained impressions and experiences with friends. Here, online travel communities and travel blogs [41] enable members to collect, view and exchange travel items such as blog entries or pictures or to add own content and reviews.

Tourists are confronted with overwhelming information.

As conclusion, ICT-based services are becoming more and more popular among tourists to satisfy their varying requirements in the different trip phases. Although such services help to access reliable and accurate information as well as to undertake reservations in a fraction of time, cost and inconvenience required by conventional methods [105], the huge amount and variety of information available on the Internet entails high cognitive costs and might lead to daunting information overload.

Hence, researching and booking a trip might be a time-consuming process and a frustrating experience for tourists when they cannot locate the information they are looking for in the pre-trip phase nor the right mobile service to access during the destination stay [101].

1.2 Recommendation systems in the travel & tourism domain

To counteract the risk of information overload, intelligent systems are needed that assist users in searching through the vast amount of information that is available for them. However, users typically have different needs and would like to receive suggestions that are tailored to their individual situation. Therefore, personalization plays a crucial role. However, to address the needs of thousands of different users, *mass customization* is necessary. This can be achieved by recommender systems.

Recommender systems are defined as applications that e-commerce sites exploit to suggest products and provide users with information to support them in their decision-making [76]. Recommender systems have been successfully deployed in the domain of travel & tourism. Their task is to mimic the traditional interaction process with travel agents when users would like to receive personal trip suggestions and advice on planning their trip [133].

Travel planning is a complex process, involving several stages.

Building a personalized travel plan is a complex process and involves several *stages*, consisting of one or more *destinations* to visit, *tourism objects* (e.g., museums, palaces or theatres), *accommodation* and *means of transportation* [113]. Systems that provide personalized support in the vacation planning process are based on three main functionalities [7]:

a) content selection
b) user model adaptation, and
c) presentation of results

Content selection refers to delivering useful information with respect to the different process stages, ranging from information about the destination, its attractions, to accommodation offers and route suggestions. Systems that provide support in all stages of the tourist's decision-making process are rare (e.g., trip@dvice). Most systems rather focus on delivering support according to a specific stage. *User model adaptation* refers to the process of creating user profiles reflecting their needs and constraints, and keep them up-to-date by incorporating feedback obtained from users over time. Recommendations are then generated by matching user profiles against travel-related items and presented to the users. Hereby, the usability of the system is a crucial issue as the interaction and interface design affect the user's decision-making process. If the recommendations are useful, but the *presentation* is poor, users might refuse to use the system [113].

1.3 Challenges

The development of travel systems that provide such adaptive, personalized travel plans poses several issues:

- **Integration of travel-related information.** This issue deals with the *integration of travel-related information* from heterogeneous touristic data sources. In each of the three trip phases, tourists have varying information needs that have to be satisfied by tourist services. Such services provide travel-related information about hotels, flights, tourist attractions, events & activities, public transport, car rentals, weather forecasts or geospatial information in form of maps or routing advices. These services should cover the whole value chain of tourism, which consists of the phases information/-booking, transport, accommodation and destination/information [42].
- **User modeling.** Key for the provision of customized, user-tailored information is the user profile. A system that has no a priori knowledge about the tourist has little chance to offer *personalized information.* It would need too much interaction to extrapolate the tourist's interests in real time. The main goal is to build a model of the interests and preferences of the user in the pre-trip phase and keep it up-to-date while the tourist is on the move.
- **Matchmaking task.** The third issue is the *matching* of tourist profiles against tourist attractions in order to come up with a ranked list of attractions. If the interest profile of the tourist matches the characteristics of a certain attraction, the attraction contributes to the tourist's satisfaction and thus should be recommended to the tourist. Thereby, both the tourist profile and the touristic resources offered by service agents have to be intersected by the matchmaking algorithm to examine whether they share similar structures. The top-N ranked matches can then be proposed to the tourist.

Next generation tourism information systems have to integrate these issues in order to help tourists in their decision-making process and to better satisfy their information needs in the respective trip phases.

1.4 Research questions

A study by Xiang & Gretzel [155] investigates the 'semantic' representation of the tourism domain by analyzing information provided on tourism Web sites as well as tourists' information needs expressed through search engine queries. Their study shows that there is a gap between the domain ontology derived from tourism Web sites and the ontology that results from search engine queries. The results of this study imply that tourists and tourism providers do not use the same 'language' and do not share the same view on the objects (cf. Figure 1.2). Thus, the major problem is to combine the user's view (tourist's personality and preferences) with the supplier's perspective (tourism objects).

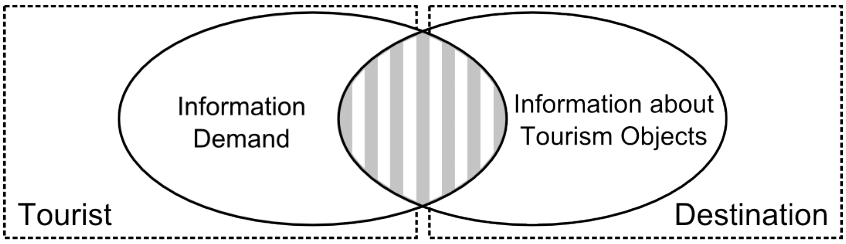


Figure 1.2 *Matching the tourist's view with the travel suppliers' perspective.*

This thesis addresses the needs of the two above-mentioned two user groups, tourists and tourism organizations by matching their respective views. In this way, tourism organization can provide more personalized services and tourists can better satisfy their information needs. The work presented within this thesis targets to support tourists in their decision-making and to facilitate the process of identifying those tourism objects of a specific destination that best fit the tourists' preferences and personality. This goal directly depends on the quality of the matchmaking process (i.e., finding those tourism objects that are most attractive to the tourist).

We propose a matchmaking process that matches tourist profiles with the characteristics of tourism objects in order to obtain a ranked list of appropriate tourism objects for a particular tourist.

Specifically, the thesis addresses three main research questions:

- Q 1: *Can tourist types existing in scientific tourism literature be used to obtain a high-level user profile?*
  Tourist typologies have been introduced in order to explain the motivations of people to go on vacations and their different travel styles. In this work, tourist types are used in order to generate high-level profiles of users' interests.
- Q 2: *Can user feedback be exploited to improve the matchmaking process?*
  User feedback in form of positive or negative ratings of the proposed set of tourism objects can be a valid source to refine the user's specific interests and deliver a new set of objects that better fit the user's interests.

- Q 3: *Can we exploit the semantic relations within a tourism ontology to infer the user's interest in objects not having been rated yet?*
  Given that the tourism objects are semantically annotated with concepts from a tourism ontology, these semantic relations within the ontological graph may help to infer a user's interest in concepts that have not been rated by the user yet. The semantic relations can be exploited to propagate user interests between parent and child concepts and thus allow a fine-granular tracking of users' interests.

1.5 Main contributions

In order to provide highly personalized trip recommendations for tourists, we propose an *iterative matchmaking process* (see Figure 1.3) that matches tourist profiles with the profiles of tourism objects in order to propose personalized recommendations (i.e., a list of appropriate tourism objects) for a particular tourist. The process consists of two sub-processes (cf. Fig. 1.3, no. 1a and 2a), which are described in the following.

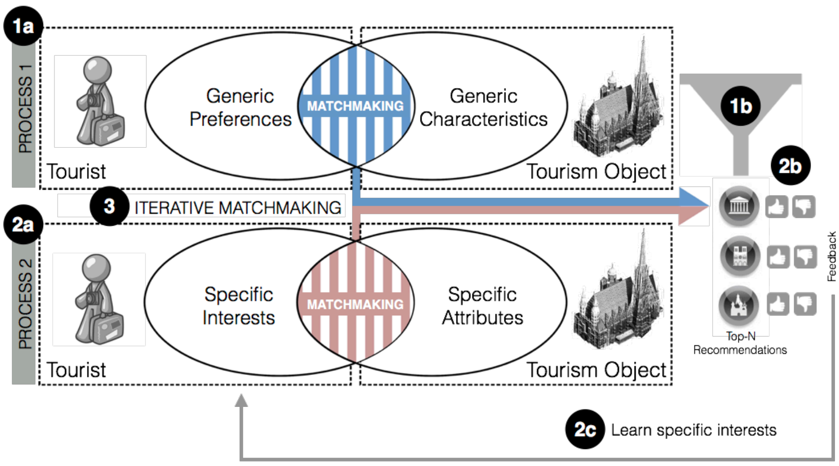


Figure 1.3
Overview of the matchmaking process.

- For the *first* matchmaking process (cf. Fig. 1.3, no. 1a), a *stereotype* approach is devised to model tourists' *generic preferences* and to establish a basic user profile. Tourists are typically not able to exactly specify all their interests during trip-planning but rather describe their predispositions through statements such as "I am more interested in culture than in history". In order to model such statements, the concept of tourist types (e.g. Cultural Visitor, Sight Seeker, Nature Lover, etc.) is leveraged.
- This profile is then related against the *generic characteristics* of the tourism objects in order to *recommend* a top-N list of tourism objects (cf. Fig. 1.3, no. 1b) to a tourist, by showing him or her the N closest, i.e. most similar, tourism objects with respect to his or her profile. For that, we propose a semi-automatic way to model the generic characteristics of the tourism objects. In a first step, domain experts mark manually for each of the prototypical tourist

factors (e.g., Action Seeker or Cultural Visitor) a small sample of typical tourism objects that are closely related to these types. In a second step, the ratings of the domain experts can be propagated to other tourism objects that are similar to the ones rated by the domain expert by using a semantic similarity measure.

- The focus of the *second* matchmaking process (cf. Fig. 1.3, no. 2a) is to refine the generic tourist profile and to enrich the generic preferences of a tourist through more specific interests (e.g., a tourist may be a Sight Seeker but may dislike churches). This is achieved by exploiting user feedback on the proposed top-N list of objects (cf. Fig. 1.3, no. 2b) and by using this information to derive a more specific profile that is capable to model statements such as "dislike of churches" (cf. Fig. 1.3, no. 2c). In order to generate these specific profiles, the main ideas of *spreading activation over ontologies* [128] are applied. In our case, a specific profile is represented as an overlay of an ontological domain model describing the tourism objects.
- The *first* and *second* matchmaking processes are combined and executed iteratively (cf. Fig. 1.3, no. 3), thus resulting in consecutive recommendation cycles. At the beginning, tourists state their generic preferences and obtain a first top-N list of recommendations. As long as they are not satisfied with the recommendations, they can criticize the proposed tourism objects by stating positive/negative feedback, which will be used to refine their profile and to deliver a new set of top-N objects. The combination of the two matchmaking processes is done with the help of a weighting factor that controls the influence of the two processes on the resulting similarity value between a tourist and a certain tourism object.
- A tourism ontology is used by both sub-processes as pivotal element to drive the matchmaking. The tourism ontology consists of concepts (e.g. museum, church, palace, historical architecture, city highlight, etc.) to describe the tourism space and contains tourism objects as instances.

Our approach is tested through a prototypical recommendation system that recommends tourists visiting Vienna appropriate tourism attractions that are tailored to their personal needs.

1.6 Methodological approach

The methodological approach of this thesis has its theoretical foundations in Information Systems design theory. It is based on the design-science paradigm presented in Hevner et al. [68]. The design science paradigm seeks to create knowledge and understanding of a problem domain and its solution through the building and application of innovative design artifacts. Thereby, one can distinguish between design processes (e.g., build and evaluate) and design artifacts. Artifacts are defined not only as the resulting instantiations (working prototype), but also comprise constructs (vocabulary), models (abstractions & representations) and methods (algorithms & practices) applied in the development as well.

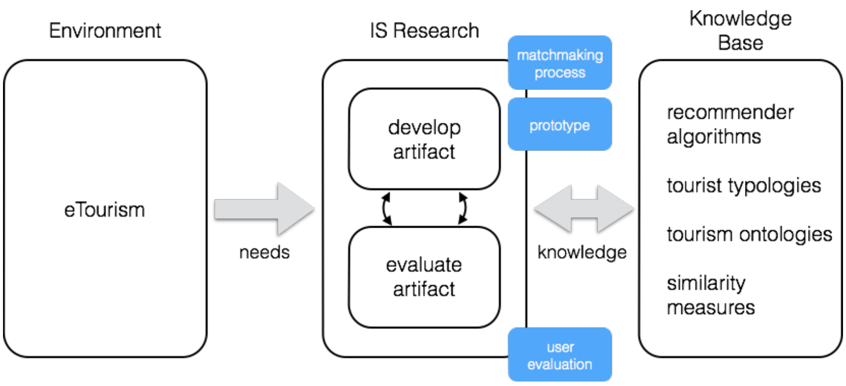


Figure 1.4 *Methodology framework.*

The problem domain that is tackled within this thesis refers to help tourists in selecting appropriate tourism objects and activities during their destination stay, which is an information-intensive task. Much research has been conducted in this area, targeting information search behavior, travel destination choice models [47] as well as recommender systems [45]. Thereby, the gap between the mental model of tourists and the model of the tourism space (i.e., the destination) is still a crucial issue [101, 155]. Obviously, this has a negative impact on the planning experience of tourists. They might not be satisfied with the proposed objects and can get frustrated when they cannot find the objects they are looking for.

The ontology-driven matchmaking process which is presented in this thesis aims to make a contribution to this area (cf. Figure 1.4). The matchmaking process is able to generate customized trip proposals by (a) modeling and integrating the customers' and suppliers' perspective, (b) leveraging the concept of tourist types, (c) exploiting ontologies, and (d) integrating tourist feedback to revise the current recommendation of tourism objects.

To evaluate our work, we conduct a user study by asking users, mainly students, to interact with the prototypical recommendation system and fill in a questionnaire afterwards.

1.7 Structure of this thesis

The remainder of this thesis is split into the following chapters.

In Chapter 2 we discuss *related work* related to the topic of this thesis, comprising recommendation techniques, user modeling, tourist typologies and Semantic Web technologies.

In Chapter 3, we outline the *first matchmaking process*, which exploits the notion of *tourist types* to generate both a high-level profile of the user and of the tourism objects. We show that these profiles can be represented in form of vectors and use vector-based matchmaking between the user profile and the profiles of the tourism objects in order to come up with a ranked list of top-N objects. The tourism ontology which is a key component of the matchmaking process is described as well.

Chapter 4 then describes the *second matchmaking process*. User feed-back on the proposed set of tourism objects is used in order to derive a more fine-granular profile of the user. An ontology-based approach is used to exploit the user feedback given for certain objects and propagate the user interests within the ontological graph to predict the interest in other tourism objects, that have not been rated yet by the user.

In Chapter 5, we show how the *two matchmaking processes* can be combined based on a weighting function that controls the influence of each of the two matchmaking processes.

Chapter 6 presents the *implementation* of the matchmaking process in form of a Web-based prototype. The user interface and functionalities are described, including a basic routing service that calculates a route between the proposed tourism objects. In addition, the different data sources used to describe the tourism objects as well as the system architecture and technologies are stated.

Chapter 7 presents the *evaluation* of the matchmaking process, which has been carried out in form of a user study. It comprises the set up of the user study, the dataset used and the results.

Finally, Chapter 8 *concludes* this thesis by pointing out the main contributions and giving an outlook on further research issues.

2 State of the art

The design of a matchmaking process in the domain of e-tourism covers different research areas. In detail, this thesis draws on previous work in the area of recommendation systems, user models, tourist typologies and Semantic Web technologies. This chapter first gives an overview of recommendation systems in the tourism domain and outlines the most important recommendation techniques (Section 2.1). Recommendation systems exploit user information in order to provide personalized content. Thus, we analyze different forms of user models and methods to acquire such user models in (Section 2.2). For modeling tourist profiles, we leverage a novel approach based on tourist types. Therefore, we review existing tourist typologies in (Section 2.3) as well. Finally, we describe in (Section 2.4) how Semantic Web technologies have been applied in the domain of e-tourism and review classification schemes of tourism attractions and ontologies that have been developed in this area.

2.1 Recommendation systems in tourism

In the following, projects, systems and techniques are mentioned that provide travel-related recommendations in the specific stages of the decision-making process before going on vacation and during the destination stay. These recommendation systems support tourists in selecting the right destination, accommodation, a set of tourism objects, appropriate means of transportation as well as mobile services during their stay at the destination that satisfy ad-hoc information needs.

Tourism recommendation systems use a variety of approaches to deliver personalized travel items or products to users [47]. Burke [25] classifies these approaches based on the used knowledge source and proposes a classification in *demographic-based*, *collaborative-based*, *content-based*, and *knowledge-based* recommendation approaches. An overview of these approaches is depicted in (Figure 2.1). Within this section, their basic principles as well as related problems are outlined. Often, a combination of such approaches is used in order to improve the accuracy of recommendations and to overcome the shortcomings of the individual approaches. Therefore, *hybrid approaches* [160] have been introduced which are discussed as well.

Destination recommendations.

The Destination Recommender presented in [36] follows a hybrid architecture that leverages content-based filtering and case-based reasoning to generate an initial list of destinations that might be attractive for a given user. Social attributes of destinations (e.g., crime rate, volume of traffic, noise level), which are captured in a destination context ontology, are used to revise the initial recommendations based on the social attribute preferences of the user.

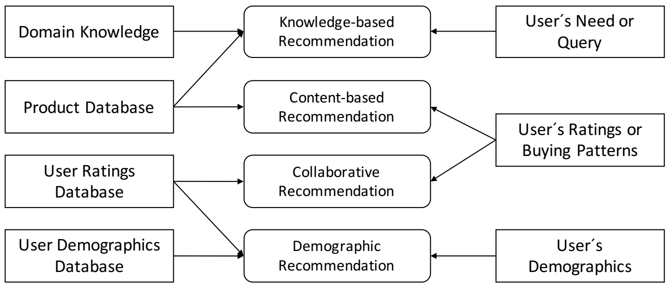


Figure 2.1 *A classification of recommendation techniques based on their knowledge source [25].*

Accommodation recommendations.

In the project Reisewissen [111], a hotel recommendation engine has been developed that exploits Semantic Web technology to enhance the quality of the hotel search process. User requirements are semantically matched against hotel resources, resulting in a ranked list of suitable hotels. TrustYou [141] is a semantic hotel search engine, which recommends hotels based on user reviews, which are aggregated from different platforms such as Trip Advisor, Expedia or Qype. The user reviews are examined through linguistic analysis (negative, neutral and positive comments can be distinguished) and annotated semantically. In this way, a search query for *Barcelona hotels at the beach for good value* thus also finds hotels described with *cheap rate*, *incredible rate* as well as *beachside location* or *great beach*. For a specific hotel result, all positive and negative comments gathered from the different reviews are presented in aggregated form to the user. In [159] an approach is presented that derives semantic annotations of tourism products based on the proximity to certain accommodations and recommends the accommodation with the highest utility score for the user.

Tourism objects recommendations.

Systems that focus on the second stage of planning a trip propose tourism objects at the destination which was selected beforehand during the destination selection stage. Tourism objects, including attractions and activities, are often the reason for tourists to visit a particular destination as they expect their needs will be satisfied during their stay [118]. Huang & Bian [72] present an intelligent system that provides personalized recommendations of city attractions. The recommendation process utilizes Bayesian network techniques to generate personalized suggestions by taking account different factors, comprising tour motivation, traveler type, occupation and personality. The system provides an interactive geographic interface for displaying the recommendation results.

Route & transportation recommendations.

Heracles II [4] is a constraint-based framework for interactive planning and allows tourists to select a suitable means of transportation to reach the destination, including flying, taking a train, renting a car or taking a taxi.

The penetration of high-end mobile devices together with the decrease in mobile data prices have resulted in tourists asking for personalized advices not only in the trip planning process, but also during their destination stay [62]. In this (mobile) context, the information overload problem even becomes more evident because of the intrinsic features of mobile phone usage (e.g., small screen, no keyboard, low data transfer rates). However, mobile recommender systems can exploit two peculiarities which come up with mobility [114]. The first one is related to *context-awareness*, i.e., the ability to exploit context factors (e.g., user

mobile recommender systems

location, time, weather situation) to adapt user interaction and application behavior to the current situation of the user. For example, the etPlanner system [70] delivers personalized information regarding events, sights, restaurants or accommodations and provides push messages with information about changing weather conditions. The second refers to *ubiquity*, which was first stressed by Mark Weiser [151], envisioning a scenario in which computers will be available throughout our physical environment while making them effectively invisible to the user. Almost every object in our everyday environment will be equipped with embedded processors and wireless communications to facilitate interaction with users and to perform and control a multitude of tasks and functions. In the area of mobile tourism services, ubiquity is not seen in this highly pervasive sense but rather as a challenge for tourists to be anywhere at anytime and to consume travel-related information. Research with respect to mobile, context-aware services has resulted in a wide range of mobile tourist guides [61]. Since one of the first famous prototypes [10], the sophistication of mobile guides has increased, and research in this field now specializes on features such as personalization, recommendation, context-awareness together with new forms of user interaction, collaborative usage and social integration.

2.1.1 Demographic-based recommendations

One of the simplest approaches is to propose recommendations based on the *demographic profile* of a user. The demographic profile may consist of user characteristics such as *age*, *gender*, *language* or *area code*. These characteristics can be used to find similar users and then, recommend items that are preferred by those similar users. Schiaffino & Amandi present in [124] an expert software agent named Traveller that (besides other recommender techniques) exploits demographic information of tourists in order to propose travel tours. In detail, the agent compares attributes (e.g., age, marital status, travel party composition) in the user profile against corresponding attributes in a sightseeing tour to filter appropriate tours.

Advantages & problems

The advantage of demographic-based recommendation systems is that a new user can receive recommendations without being required to provide any preference information. However, this strategy comes along with certain problems. As only limited user information is available, it is not possible to propose highly personalized recommendations.

2.1.2 Collaborative-based recommendations

Collaborative filtering (CF) is a technique that produces recommendations for a given user and a set of items by exploiting the opinions of other people on these items [67]. Usually, the opinions of other people are expressed in form of ratings. In this way, it does not need to exploit the features or characteristics of the items that should be recommended. Basically, a rating is a value on a set of items from a specific user and represents to which extent the user likes these items. As shown in Table

2.1, ratings can be visualized in form of a user-item preferences matrix whereby the fields represent individual rating values. If a field of the matrix is empty, it indicates that the user has not yet rated this item. A rating can be either given in an *explicit form* (e.g., the user can

Ratings	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	5	8		7	8
User 2	10		1		
User 3	2		10	9	9
User 4		2	9	9	10
User 5	1	5		1	
User 6	2		9	10	

Table 2.1 *A matrix depicting the ratings of a set of items by users.*

be asked explicitly about his/her opinion on an item after purchasing it) or *implicitly* (e.g., through observing the user's interaction with the recommendation system and monitor his/her clicks or based on his/her transaction data). In Table 2.1, the ratings are given exemplary in a scale between 1 and 10. Such *scalar ratings* are heavily used within collaborative filtering approaches, but other rating schemes can be used as well, including *binary models* (e.g., agree/disagree) or *unary models* (e.g., clicked/purchased/bookmarked items) [23].

The two main functionalities of collaborative filtering approaches are *recommending* a set of suitable items for a specific user or for a given item, calculate its *predicted rating*. For doing so, collaborative filtering needs ratings for these items in order to make any predictions. However, it does not need to analyze the content of the items as required by *content-based approaches*. According to Breese et al. [21], collaborative filtering algorithms can be classified in either memory- or model-based algorithms.

- *Memory-based algorithms* exploit the entire database to make predictions. Basically, these algorithms use statistical techniques (e.g., Pearson correlation) to find highly correlated users known as *neighbours*. Therefore, they are also called *nearest-neighbour algorithms*. After identifying the set of neighbours, their ratings are used to create a prediction or recommendation for the current user [121].
- *Model-based algorithms* propose recommendations by *learning a model* of the user ratings. To build such models, a variety of *machine learning algorithms* can be used, such as Bayesian networks, clustering and rule-based approaches. *Bayesian network models* form a probabilistic model of the user's predicted ratings. *Clustering approaches* (e.g., k-means [84]) classify users into groups and then calculate the probability of a specific user to belong to a certain group, which is finally used to estimate the probability of his/her ratings. *Association rules* identify the sets of items that are typically purchased together. Those items that have been not yet purchased are then proposed to the user [122].

As model-based algorithms pre-compute the user model *offline* and only need to access this model in the runtime-phase, such algorithms scale

better if the recommender system has to manage a huge number of items or users. Nevertheless, memory-based approaches are quite popular for mid-sized user bases due to their simple algorithms.
Basically, two forms of collaborative-filtering algorithms are distinguished in literature, comprising *user-based* and *item-based* collaborative filtering (CF). User-based CF techniques look at the similarities between users (based on their ratings) while item-based CF techniques are based on the similarities between items (based on the items' ratings) [123].

2.1.2.1 User-based collaborative filtering

User-based collaborative filtering techniques work by first finding highly correlated users and then, recommend items that are preferred by those users [67]. The basic idea behind this technique is that if like-minded users prefer an item, then there is a high probability that the target user will also like this item. Their similarity can be calculated by comparing the other users' ratings with the ratings of the target user.

Ratings	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	5	8		7	8
User 2	10		1		
User 3	2		10	9	9
User 4		2	9	9	10
User 5	1	5		1	
User 6	2		9	10	

Table 2.2
An example for user-based collaborative filtering, which exploits the similarity between users (rows).

As depicted in Table 2.2, *User 3* is highly correlated with *User 6*. At the same time, *User 3* prefers *Item 5*, which has not yet been purchased by *User 6*. Therefore, *Item 5* might be recommended to *User 6*. In order to identify highly correlated users, different similarity measures can be used such as *Cosine similarity* or *Pearson correlation* similarity. These similarity measures require a vector representation of the user information. In general, Pearson correlation computes the correlation of two vectors in a scale between the values 1 and -1. A value near to 1 indicates that the values from the two vectors are rather similar whereas a value towards -1 indicates dissimilarity. A value around 0 states that the two vectors are rather independent from each other. Using Pearson correlation, the similarity between the *User u* and a *Neighbour n* is calculated by comparing the ratings of all these items which are rated by both persons [123]. I denotes this set of co-rated items, while $\bar{R}$ depicts the average rating of the respective person.

$$sim(u,n) = \frac{\sum_{i \subset I}(R_{u,i} - \bar{R_u})(R_{n,i} - \bar{R_n})}{\sqrt{\sum_{i \subset I}(R_{u,i} - \bar{R_u})^2}\sqrt{\sum_{i \subset I}(R_{n,i} - \bar{R_n})^2}} \tag{2.1}$$

Once the set of neighbors is identified, a recommendation can be proposed by looking at the items that are preferred by those neighbors. Then, a ranked list of recommendations can be generated by taking, for example, the weighted average of their ratings.

2.1.2.2 Item-based collaborative filtering

Item-based collaborative filtering techniques target at finding items that are highly correlated with the known preferred items [123]. The idea is that a user is likely to have the same preference for similar items. However, in contrast to content-based filtering methods which exploit item features to obtain a similarity value, item-based CF techniques calculate the similarity between items by exploring how other users have rated these items. As shown in Table 2.3, *User 6* gave a high score to *Item 4*. At the same time, this item is similar to *Item 5* as users who preferred *Item 4* also rated *Item 5* with a high value. Thus, *Item 5* might be recommended to *User 6*. In order to find items that are similar to the known preferred items, Pearson correlation can be used as well [121]. The following equation computes the similarity between *Item i* and *Item j*. *U* depicts the set of users who rated both *Item i* and *Item j*:

Ratings	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	5	8		7	8
User 2	10		1		
User 3	2		10	9	9
User 4		2	9	9	10
User 5	1	5		1	
User 6	2		9	10	

Table 2.3
An example for item-based collaborative filtering, which exploits the similarity between items (columns).

$$sim(i,j) = \frac{\sum\limits_{u \subset U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum\limits_{u \subset U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum\limits_{u \subset U} (R_{u,j} - \bar{R}_j)^2}} \tag{2.2}$$

The main difference between the two Pearson correlation formulas above is that in the former Equation (2.1) the *average rating of a user* is used whereas in the latter Equation (2.2) the *average rating of an item* is exploited.

Advantages & problems

The advantage of collaborative filtering techniques is that they do not require to analyze the characteristics of the individual items. However, a critical issue is the *cold-start problem* [121]. In case that there are too few ratings available, the quality of the proposed recommendations is rather low. This is certainly a big problem when a recommender system is introduced the first time and there are no ratings by the community yet. But also well-established recommenders may have to fight against *sparsity problems*. This is because in general only a few members of a community tend to be very active. Most of the items are only rated by a few people. To tackle such problems, different strategies can be used such as utilizing hybrid recommenders (together with content-based approaches) [25] or giving incentives for users in order to boost the users' willingness to make ratings. Such strategies can also be used if a *new* *item* is inserted in the database. As long as there are no ratings for this item, it will not be recommended to any user. In case that a *new user* is registered to the system, *stereotypes* may be used to initialize a basic profile, or the user might be asked to rate a set of items during the registration process to elicit some basic information about him/her.

2.1.3 Content-based recommendations

Content-based recommendation systems are based on similarities between item descriptions and the preferences of users [102]. The idea is to find items that have similar features compared to the known preferred items or the user model. Thereupon, the items with the highest similarity are recommended to the user. In case of unstructured data (e.g., news articles), information extraction techniques might be used in order to convert the unstructured data to a structured representation. For example, texts are usually represented in a *vector space model* based on TF-IDF (term frequency - inverse document frequency) values [11]. Besides describing the items in a structured form, it is also important to look at what the user has liked in the past and use this kind of information to learn a *user model*. This can be accomplished either through *implicit user feedback*, which monitors the interaction behavior of the user with the system (e.g., by tracking items clicked or purchased) or *explicit user feedback*, in which the user rates items through an interface (e.g., clicking on thumbs up and thumbs down buttons to express like/dislike). Obviously, implicit user feedback methods have a higher level of uncertainty than explicit user feedback methods. To put it in other words, the latter shows a higher level of confidence as the user expresses his/her interests in a direct way. A number of algorithms have been proposed in order to estimate the user's interest in a new item based on his/her known preferred items [102]. For example, in case of vector space models, a simple method is to use Cosine similarity between keyword vectors in order to find similar items the user may like.

Advantages & problems

Content-based recommendation systems avoid problems related to sparsity as they do not depend on the ratings of other users. Pure content-based recommendation systems perform well if the items can be properly represented in form of features. In some situations, however, this might be rather challenging. For example, in case of multimedia content such as images or video, extracting such a feature set might be difficult. Moreover, content-based recommenders might suffer from over-specialization. Due to the applied similarity metrics, users are restricted in viewing rather similar items and are not capable of exploring completely new items that are very different in terms of features with respect to the previously rated item set.

2.1.3.1 Case-based recommenders

Case-based recommenders are a special kind of content-based recommendation systems and avoid some of the previous mentioned problems. Basically, they have their origin in case-based reasoning techniques [148]. In contrast to knowledge-based recommenders which rely on a strong domain model and pure content-based approaches that deliver recommendations based on what the user has liked in the past, case-based recommenders exploit a case base that stores past problem solving experiences in form of cases [130]. Such a case usually consists of two parts, namely a specification and a solution part. The specification part describes the problem to be solved, which represents some needs of the user and is specified by him/her through a query language. The solution part details the solution that is used to solve the user needs. When a query with a new problem description is received by the case-based recommender, it retrieves a set of cases with similar problem descriptions, whose solutions might be valid for the current problem. In case that they do not fit the current problem's specification, they can be adapted and finally, the problem specification and proposed solution is inserted as new case in the case base. Case-based recommendation systems are quite popular in the domain of tourism. One reason for this is that travel experiences of tourists can be adequately modeled in form of cases, another reason that travel items can be represented in a structured way based on a set of features, which are used to define a similarity measure between cases. For example, a specific destination might be represented in form of its price, duration, accommodation, location or proposed activities [130]. Examples for a case-based recommender in tourism are the Intelligent Travel Recommender System (ITR) [115] and the travel advisory system Dietorecs [14] that guide travelers through their travel decision making process in form of a conversation. In detail, a case consists of following components (cf. Figure 2.2), namely *travel wishes* (e.g., biking in Trento), *user information* (e.g., John, aged 42, German), *travel bag* that depicts all the travel items the user selects during the recommendation process (e.g., accommodation), *navigation history* of the user as well as *reward* which is a rating given by the user for the items in the travel bag.

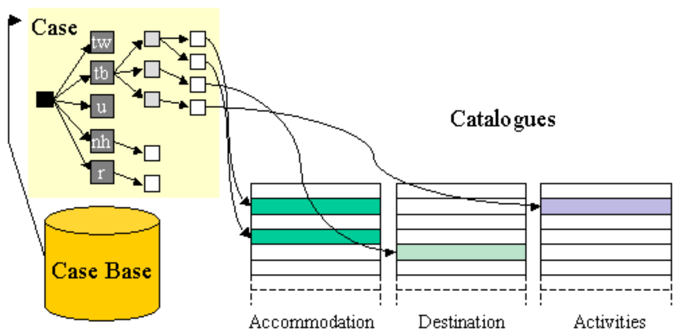


Figure 2.2
A case base depicting the components of a particular case and associated items in a travel catalogue [14].

As soon as the user issues a query which includes a set of travel wishes (e.g., travel party 'family', accommodation 'hotel', period 'July'), a set of similar cases has to be retrieved from the case base. Measuring similarity is a complex issue as different features (e.g., numeric features such as price and nominal features such as travel party composition) have to be compared against each other. Hence, different similarity metrics are required to measure similarity for the individual features (e.g., for the feature price, an asymmetric similarity measure may be used as a lower price should be rated higher than a price that exceeds the target price). Other features such as vacation types may be even more difficult to compare. An example is given in [130], arguing whether a *skiing holiday* is more similar to a *walking holiday* than it is to a *countryside* or a *beach holiday.* If the holiday types and their relations are modeled in an ontology, ontological similarity measures can be used to estimate the similarity/distance between the different concepts as shown in (Figure 2.3). After a set of similar cases is identified, they are used as candidates in order to retrieve up-to-date items from travel catalogues. The reason is that the items recommended in the past might be already out-of-date and not appropriate for the tourists anymore. Hence, another similarity metric is applied to retrieve items from the catalogues that are similar to those in the case set. In case that no results can be retrieved that satisfy the requirements of the user, query relaxation is used to relax some constraints (e.g., choose another accommodation category or accommodation location). Consequently, the number of items returned by the query is increased.

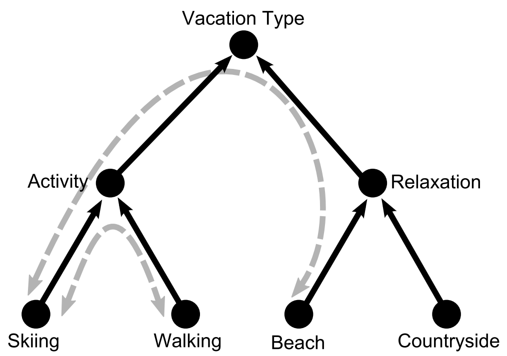


Figure 2.3
Ontological fragment depicting different vacation types and their relations, adapted from [130]. The similarities between the concepts can be measured according to the length of the shortest path between them within the hierarchical structure.

Advantages & problems

The advantage of case-based reasoners is that they can be used also in cases of unprecise queries and problem specifications. However, prerequisite for this is that some cases must already exist in the case base that can be adapted for such problems, otherwise no solution might be found [87]. In this way, case-based reasoning suffers from a kind of cold-start problem as well as a minimum set of reference cases covering different problem areas is needed in order to work properly. Such cases might be defined by domain experts at the beginning.

2.1.4 Knowledge-based recommendations

Knowledge-based recommendation systems propose recommendations based on a user's needs and requirements [24]. A knowledge base stores all the domain knowledge that is needed to associate certain user needs with products or items. Hence, a knowledge-based recommender does not need any kinds of ratings and it also does not need to obtain any information about a particular user before its usage. They are typically used in application domains that require deep knowledge about the product domain. For simple products such as books conventional recommendation approaches (content-based or collaborative-based filtering) are sufficient. In contrast, when selling financial services, recommendations must adhere to legal regulations, suit the customers' financial restrictions and are in line with a company's sales strategy. For such complex products, more intelligent interaction mechanisms are required that engage the user in form of a dialogue in order to find those services that fit the user's needs and pre-defined constraints. The conversation is used to elicit the requirements of the user which serve as input in order to find appropriate solutions. If no solution can be found, different approaches such as relaxation of filter constraints can be used to come up with at least one suggestion [46].

Advantages & problems

Knowledge-based recommendation systems do not have a cold-start problem as their recommendations are not based on user ratings. They also do not require a pre-defined user model as they follow a conversational approach that elicits all users' needs and requirements within an interaction process. In this way, they react to short-term user requirements and are not capable of raising the accuracy of recommendations by learning the long-term user interests. Moreover, establishing a knowledge base is quite an extensive engineering task, which has to be done manually by domain experts.

2.1.5 Hybrid recommenders

Hybrid recommendation systems combine different techniques such as content-based and collaborative-based ones in order to eliminate their individual shortcomings and thus, to improve the quality of the recommendations. For example, Schiaffino and Amandi [124] propose for their expert software agent named Traveller an approach that combines demographic information about users with collaborative filtering and content-based approaches in order to deliver tour recommendations to tourists. Another approach in the domain of tourism is presented by Huang & Biang in [72] and [73], which utilizes Bayesian network techniques in order to recommend tourist attractions to a specific user. A Bayesian network is a directed acyclic graph that consists of a set of nodes and arcs. The nodes represent variables and the arcs encode the probabilistic relationships between those variables. The Bayes' theorem to calculate the conditional probability of Y given X is stated as:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

P(Y) is called the prior proability as no information about X is included. P(Y|X) is the conditional probability of Y, also expressed as the posterior probability as its probability depends on X.

In order to estimate the tourist's preferred activities, Huang & Biang utilize the travel behaviors of similar users as well as the tourist's characteristics within a Bayesian network. The characteristics are modeled through the variables *age*, *occupation* and *personality*. These three variables influence the variable *traveler type*. And finally, both this variable and the variable *tour motivation* influence the *preferred activities*, comprising *natural-based sightseeing, cultural-based sightseeing, outdoor activities, entertainment activities* and *events*. (Figure 2.4) depicts the probability distributions of the variables. The conditional probability tables for the variables have been defined based on a review of existing research in the travel domain. After obtaining feedback from the tourists based on the recommendations, this data can be used to adjust the probability tables. An example to calculate the posterior probabilities is given

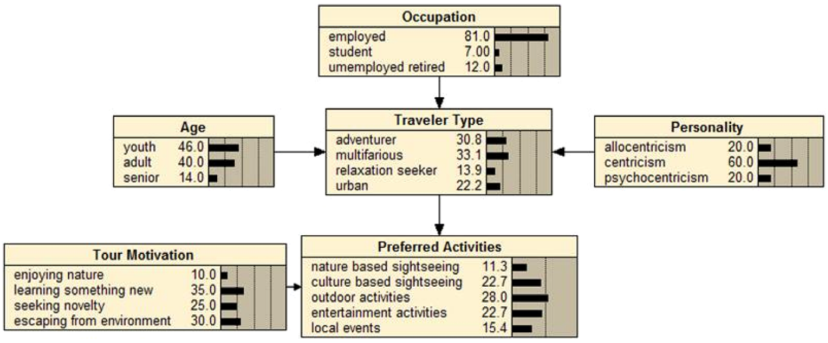


Figure 2.4
Bayesian network to infer the probabilities of preferred activities [72].

in [72]. As destinations differ from each other in the variety of offered activities and tourist types, the authors suggest to adapt the probabilistic influences for each destination based on destination-specific surveys.

2.2 User modeling & personalization

As tourists have individual preferences, user modeling plays an essential role in the provision of personalized travel information. As stated by [23], a user model is a representation of information about an individual user. This information is needed to provide customization, i.e., adapting an application towards the particular user's characteristics, which are often labelled as *features*. The most important features that are typically represented in a user model, are [23]:

- *the user's knowledge*, which is typically modeled in adaptive educational systems and represent the level of knowledge of certain items/topics within a particular domain.
- *the user's interests*, which are vital for recommendation systems in order to propose items that are preferred by the user.
- *the user's background*, which comprises features that influence the adaptation process, including the user's profession, language abilities or previous experience with the system.
- *the user's individual traits*, which describe the individuality of a user. Similar to the features describing the background of a certain user, these features are rather stable and do not change over a short period of time. They comprise features reflecting the personality (e.g., introvert/extrovert) or referring to cognitive factors (e.g., working memory capacity). Typically, such features are collected based on specially-designed psychological tests. For example, in [56], effects of the Big Five personality factors regarding adaptivity of mobile museum guides were explored.

With the rise of ubiquitous systems, features that describe the context of the user are getting important and have to be represented in the user model as well. This is especially the case for mobile, context-aware recommender systems that leverage such factors to pre-filter the list of recommendations based on the situation of the mobile user. According to Dey et al. [40], *context is any information that can be used to characterize the situation of an entity, where entity means a person, place, or object, which is relevant to the interaction between a user and an application, including the user and applications themselves.* Hence, a system is context-aware, if it utilizes contextual information to provide tailored information or services to the user. Chen & Kotz [29] distinguish between two classes of context-awareness, namely *active* and *passive context-awareness*. The former refers to applications that automatically adapt to (newly) discovered context. Such systems are also called *adaptive* as they perform the adaptation task without explicitly asking the user for permission. *Passive context-awareness* means that applications present new context information to the user who can then make the decision. Such systems are called *adaptable* systems as the user is in control of the customization step. Context factors can be classified into following categories, whereby the first three categories stem from Schilit et al. [125], while the last one was proposed by Chen & Kotz [29]:

- computing context, referring to network connectivity, communication costs and communication bandwidth

- user context, including the user's current location, as well as social context such as people nearby
- physical context, such as temperature, weather conditions, lightning or noise levels
- time context, reflecting the current time or the opening hours of buildings

2.2.1 User model representation

User models have to be represented in a machine-processable form in order to allow personalization systems to access user-related information. In the most simple form, user models can be represented in form of *key-value pairs*. Strang & Linnhoff-Popien [134] list additional representation approaches, including *object-oriented*, *logic-based* and finally, *ontology-based models*.

2.2.1.1 Key-value models

Key-value models represent the simplest data structure for representing user models. Typically, the values represent the level of relevance/knowledge of the key concepts either on a quantitative (e.g., a number ranging from 0 to 6) or qualitative (e.g., expert, intermediate, and novice) scale. The INTRIGUE recommendation system [12] that proposes tourism attractions is a good example for such a model. It uses the key-value structure to model both the properties of attractions as well as the preferences of users. For example, the property 'artistic current' can take a value from the set Baroque, Gothic, Romanic. For each property, an importance value is given that scales from 0 (no interest) to 1 (high interest).

2.2.1.2 Object-oriented models

Object-oriented models leverage the benefits of any object oriented approach, including encapsulation and reusability. The details of user profiles is encapsulated on an object level and hidden to other components. An example of an object oriented model is the GUIDE system [30], which provides a context-sensitive tourist guide for visitors to the city of Lancaster. Its information model represents three types of information, comprising geographic information, hypertext information and active components that can react to events. As soon as a request for information about a specific attraction is received by the system, the corresponding information items are dynamically composed and displayed on the user's device.

2.2.1.3 Logic-based models

Logic-based models represent user information in form of facts, expressions and rules. Hence, logic based models show a high degree of formality. Based on asserted facts and predefined rules, new information items can be inferred and added to the knowledge base. One approach in the domain of tourism is the IDUM system [74], which exploits Answer Set Programming (ASP) in order to select holiday packages that best fit the customer needs. Hereby, information about touristic offers is automatically extracted from touristic leaflets and used as input to a logic program which matches the offers to the information about tourists.

2.2.1.4 Ontology-based models

Ontological information can be used by reasoner engines to infer additional higher-level elements from the given ones based on automatic classifications and rules. Thereby, specialization and generalization features based on the hierarchical structure of an ontology as well as different kinds of semantic links allow for interest propagation within the ontological network. Furthermore, ontology-based models are especially suited for representing contextual information due to their high expressiveness and reasoning capabilities [89]. For example, different low-level contextual information (e.g., location and social context) can be combined to classify the situation of an individual as business-related (ontology fragment from [89]):

$$\begin{aligned} Business := Situation \sqcap (\exists location.Business\_place \sqcup \\ (\exists location.Public\_place \sqcap \exists time.Office\_hour)) \end{aligned}$$

Ontology-based models are heavily used by applications in the tourism domain. For example, Tomai et al. [140] explored the use of ontologies to assist tourists in planning their trip. A *user ontology* is proposed that includes the concepts *kind of trip*, *time*, *money*, *accompanying persons*, *interests* as well as *preferred leisure activities*, while touristic offers are described via a separate *tourism domain ontology*. The user ontology is populated with instances via an interface (cf. Figure 2.5) that poses ontology-driven queries to elicit the required information form the user. In this way, any information given by the user can be

Figure 2.5 *Ontology-driven user interface [140].*

easily added as instances within the ontology. However, if the ontology is modified the user interface has to be adapted as well. Therefore, it is vital that the ontology is designed thoroughly in order to prevent frequent changes. As soon as a user profile is created, it is inputted to a semantic matchmaking algorithm that filters out the offers that do not fit the information in the user profile. For that it is necessary to match the concepts of both ontologies.

Ontology-based overlay models

The purpose of an overlay model is to represent the user model based on the concepts of the domain ontology as well [128]. The advantage of this approach is that a separate user ontology is not required and that the matchmaking process between different concepts of both ontologies can be omitted. In fact, the user model is represented as a weighted (sub) set of the domain model, while the weights indicate the level of user interest in the particular domain concepts. (Figure 2.6) represents the

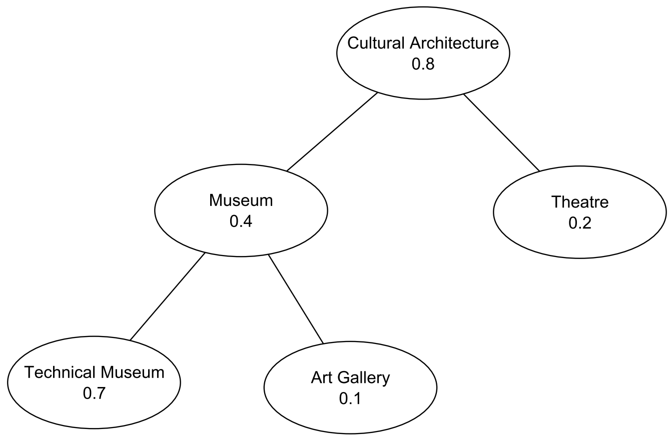


Figure 2.6 *Overlay model of user's interests on domain concepts.*

ontological fragment of a tourism domain model. The individual weights reflect the degree of interest in the corresponding concepts. Overlay models are highly effective when they utilize the structure of the semantic network for interest propagation. The concepts are linked through different kinds of semantic links (e.g., is-a, part-of) which may result in different propagation strategies. Overlay models clearly outperform conventional vector-based models, which model user features as sets of unrelated concepts. By leveraging the relations between concepts, a change of the interest weight of a particular concept can be propagated to related concepts that are semantically similar. For example, if the user shows disinterest in the concept `cultural architecture`, this information can be automatically propagated to the subclasses `museum` and `theatre`, and thus, the interest weights of both these values will be decreased. In addition, automatic interest propagation tackles the sparsity problem [23], which is a critical issue as soon as new users with unknown preferences are added to the system. By propagating interest weights within the domain network, a few elicited user preferences are sufficient in order to infer the weights of other (yet) unknown concepts.

2.2.2 User model acquisition methods

In general, user-related information is acquired in different ways, which vary with respect to the degree of user involvement. Firstly, such information can be inserted explicitly by the user. Secondly, information can be inferred by the system based on monitoring user behavior. Thirdly, information about the user can be inferred based on stereotype approaches.

2.2.2.1 Explicit user information

First-time visitors are often required to fill in questionnaires in order to elicit personal information before they can receive customized information. Usually, such kind of data is rather static and does not change frequently, such as the age, language spoken, skills or food habits. However, inputting personal data into the system is a tedious task for users and involves privacy concerns as well. Experiments have shown that users faced to more than four prompts for information from a query system tend to give up using the system [158]. Besides, users may not be able (or may not want) to describe themselves, to state their preferences and personal motives. For example, some basic questions may not be enough to derive the personality profile of users, which can only be revealed by comprehensive psychological tests. Besides, some preferences might be just described at a very high level, which may result in imprecise (travel) suggestions. For example, tourists may state that they are interested in museums without referring to the category of the museum. In this situation, the system cannot infer whether it should favor a technical to a cultural museum.

In general, the accurateness of information given explicitly by the user can be assumed to be very high (unless the intention of the user is to betray the system). However, if not being updated on a regular basis, the information could get inaccurate after some time. Therefore, some applications combine such kind of information with a kind of forgetting factor, which decreases the accuracy level over time.

2.2.2.2 Implicit user information

Another approach is to update the user model implicitly by the system through observing the behavior of the tourist, his/her interaction and situation. Typical indicators for inferring interests of users are time spent reading a Web page, bookmarks of Web pages, clicks on specific links or repeated visits of Web pages [31]. A common drawback of implicit feedback methods is that they mainly rely on capturing *positive* interest. When a user spends quite much time reading a page, it is reasonable to assume that the user is interested in its topic. However, not clicking on a link or not reading a page is no reliable indicator for disinterest [53]. Therefore, implicit methods are rather used to obtain positive feedback as users mainly pursue information they find interesting. However, Lee & Brusilovsky [82] tested negative implicit user feedback in a job recommendation system that delivers job descriptions to users. If users save such a description, positive feedback is indicated. In contrast, if a user closes a job description, this action is regarded as implicit negative feedback. The results of this study showed that negative implicit feedback reinforced the distinction between good and bad jobs and improved the quality of the recommendation process.

A method to derive positive interest in certain topics based on usage-data is stated by Kobsa [48], which has been utilized in the context of a personalized city tour guide. It is based on the assumption that the occurrence of object features (in this case topics) in the navigation history of users is normally distributed. If a feature in a specific user's navigation history appears less frequently than in a random sample, it is assumed that the user is not interested in this feature. In contrast, if a feature appears more frequently in his/her navigation history, this confirms that he/she is interested in this feature. The probability of a user's interest

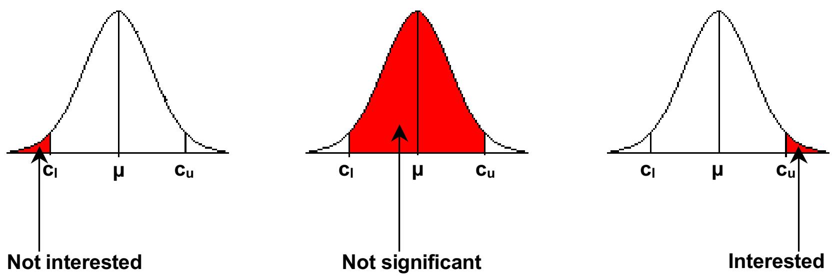


Figure 2.7
Normal distribution of users' interest in a certain feature [48].

in a certain feature is calculated using a sigmoid function based on the feature occurrences in his/her navigation history in relation to the occurrences distribution of all users. μ is the means of the distribution and c_l and c_u are the lower/upper confidence limits that determine whether a user's interest/disinterest is statistically significant. This means if the number of occurrences of a certain feature is greater or lower the confidence limits, the user has interest/disinterest in this feature.

2.2.2.3 Stereotypes

The usage of stereotype user modeling in computer systems was introduced by Rich in 1979 [116]. These models attempt to classify users into several groups and afterwards, to make predictions about them based on a stereotype that is associated with each group [3]. This procedure allows to build models of individuals quickly, as only a small amount of information is needed to assign them to stereotypical descriptions. Systems that apply stereotypes for user modeling are required to have following two kinds of information:

- the set of stereotypes, i.e., the clusters of characteristics that define the different groups. The particular set of characteristics are determined by the domain and purpose of the envisioned application.
- the set of triggers, which are events that assign a user to a particular stereotype as soon as they occur. Typically, more than one stereotype will be active for a certain user. After assigning a user to a stereotype, all the assumptions it makes about different characteristics need to be inserted in the user model.

An example for stereotype modeling is given in [26], which uses this approach to initialize a user model. It utilizes two stereotypes, namely *the specific cultural tourist* and *the general cultural tourist*, which differ in characteristics such as *user profession*, *gender*, *age* and *education level*.

As shown in the upper part of (Figure 2.8), the weights of the characteristics belonging to the same type are normalized. In the lower part, the set of predicted interest weights is depicted that users attached to this stereotype are assumed to have. The stereotypes are activated based on the information about the characteristics (e.g., profession) that is given by the user as he/she first registers to the system. In a follow-up step,

SPECIFIC CULTURAL TOURIST										
Profile characteristics										
Profession Set			**Gender**		**Age Set**			**Education Level**		
UpperClass	MiddleClass	Worker	Male	Female	< 23	23-40	>40	Higher	Secundary	Primary
0.55	0.40	0.05	0.5	0.5	0.05	0.6	0.35	0.80	0.15	0.05

Predictions on general interests *(FunctionalType)*					
Architectural	**Cultural**	**Natural**	**Recreational**	**Traditional**	**Sportive**
HIGH	HIGH	MEDIUM	LOW	MEDIUM	LOW

Figure 2.8 *Specific cultural tourist type [26].*

the interest of a user in a certain category (e.g., architecture) is calculated based on the weighted average of the interests' predictions for each category according to the degree of match with the respective stereotypes.

The usefulness of stereotype modeling depends on the number and quality of stereotypes, the accuracy of matching users against the stereotypes as well as on the quality of assumptions made from the assigned stereotypes [3].

One drawback of stereotype user modeling is the fact that users belonging to a certain stereotype are treated the same way, although they might differ with respect to various features. However, these features may not be covered by the stereotypes and thus might be ignored. In this way, stereotypes are a promising way to initialize a user model. However, during application usage, new (more specific) features might be added and some might change. Therefore, it is vital to track those changes and keep the user model consistent and up-to-date, so that a high quality of personalization can be ensured.

2.3 Tourist typologies

Tourists are not always able to exactly specify all their interests before a trip but rather describe their predispositions through statements such as '*I am more interested in culture than in history*'. Such statements can be modeled based on tourist types found in scientific tourism literature, which distinguishes tourist types according to their generic interests and emotional attitudes. For example, a thrill seeker can be described as type of person '*interested in risky, exhilarating activities which provide emotional highs for the participant*'.

Before giving an overview of existing *tourist typologies*, it is interesting to have a look at how the term *tourist* is defined.

Basically, tourism is a rather fuzzy concept. According to the World Tourism Organisation (UNWTO), *tourism* represents the *activity* of visitors, while *visitors* are travelers who take a trip to a destination outside their normal environment for less than a year and for any main purpose [145]. The UNWTO classifies visitors as *tourists*, if their trip includes an overnight stay.

A traveller is someone who moves between different geographic locations for any purpose and any duration. (UNWTO [145]).

Instead of merely classifying a tourist being a subcategory of visitors, Cohen introduces 6 dimensions to describe the characteristics of a tourist. As stated in [33], a tourist is

1. a *temporary traveller*: a tourist has a fixed place of abode, even during his trip. This distinguishes him/her from a nomad, who is a permanent traveller.
2. a *voluntary traveller*: a tourist can, whenever he wants, abandon his trip and return home.
3. a traveller on a *round-trip*: a tourist leaves his home to go on a trip but returns to the point of departure after his trip, opposed to an emigrant, who rather moves on a one-way-trip and does not return to his former abode.
4. on a relatively *long journey*: while a day-tripper or a visitor is on a relatively short trip or excursion, the journey of a tourist lasts longer, but Cohen does not specify a minimum time frame, as given by the UNWTO specification.
5. on a *non-recurrent trip*: this means, that he does not undertake a specific trip at regular time intervals, as done by owners of week-end houses.
6. on a trip, whose purpose is *non-instrumental*: a trip is not a means to achieve another goal (e.g., having a business meeting) but is a goal itself. Cohen argues that the central non-instrumental purpose, which distinguishes a tourist from other traveler roles, is the expectation of *pleasure* from the trip. A tourism trip offers elements of novelty and strangeness which tourists cannot experience in their daily routines within their home environment.

2.3.1 Cohen's classification of tourist types

According to Cohen, the concept of a *tourist* is based on these 6 dimensions and defined as '*a voluntary, temporary traveller, traveling in* *the expectation of pleasure from the novelty and change experienced on a relatively long and non-recurrent round-trip*' [32].

However, as the dimensions leave a wide scope, the distinction between a *tourist role* and other forms of traveling remains fuzzy. In his opinion, the classification between a tourist and a traveller should not merely be based on the objective characteristics of a trip (e.g., length of trip, travel party composition), but rather on the *expectation of pleasure* derived from *novelty* or *change* offered by the trip. In this way, novelty and strangeness are central elements in the tourist experience.

Based on the combination of these values, Cohen distinguishes between four kinds of tourist types. (Figure 2.9) depicts the four tourist types on an axis, which reflects their degree of *novelty* and *familiarity*.

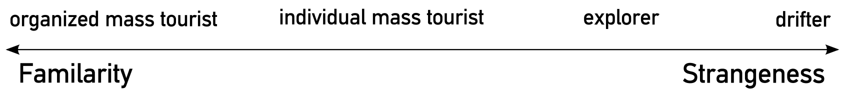


Figure 2.9 *Cohen's classification of tourist types [32].*

- *the organized mass tourists* are the least adventurous kind of tourists. Their trip is organized in advance, they do not have to make any decisions for themselves and contact with the host community is minimal. These tourists tend to become aware of their host environment only when they reach important attractions.
- *the individual mass tourists* are similar to the organized ones, but like to visit sights that are included in the trip package. However, the major parts of the trip are still organized by the travel agency.
- *the explorers* organize the trip alone. They like to visits places that are not crowded by other tourists and try to experience the social lifestyle of the local people although they do not identify with the natives emotionally. They rather look for comfortable accommodations and reliable means of transportation.
- *the drifters* have no fixed itinerary or timetable and completely immerse in their host environment (e.g., by living together with the local people).

To conclude, the interaction of the *mass tourists* with the host community is limited. They need a host environment that is similar to their home environment. The *explorers* like to experience new things and interact with the host community, but do not get involved in the life of the locals as much as the *drifters* do.

2.3.2 The travel career patterns (TCP) by Pearce & Lee

Pearce developed a travel motivation theory which is based on Maslow's needs hierarchy theory of motivation. Compared to Maslow's theory, the needs of travelers can be organized in a hierarchy or ladder [103]. Therefore, the model is called the *Travel Career Ladder (TCL)*. Its core idea is that with the increase of travel experience, travelers' motives change and motives organized more on the top of the ladder (e.g., self-development and self-actualization needs) become important. Later, Pearce & Lee de-emphasized the importance of the hierarchical structure of the model and changed the model to the *Travel Career Pattern (TCP)* model [104], in which they investigated different travel motivation patterns and how they are influenced by different levels of travel experience and age of travelers. They identified a set of 13 motivation patterns (cf. Table 2.4). *Novelty* (having fun, experience something different), *escape/relax* (re-

Table 2.4
List of travel factors by Pearce et al. [104].

Travel factors		
Novelty	Self-development (Host-site involvement)	Isolation
Escape/relax	Stimulation	Nostalgia
Relationship (strengthen)	Relationship (Security)	Romance
Autonomy	Self-actualize	Recognition
Nature		

laxing, being away from daily routine) and *relationship* (being together with friends/family) motivations are the most important factors in forming travel reasons. Pearce et al. found out that these factors are not influenced by different levels of travel experience. They thus declared these three dimensions as being the *core factors* in all travel motivation patters and constitute the main motivations for people traveling regardless of their travel experience. Travel motives concerning *self-development* through *host-site involvement* (e.g., learning new things and experiencing different cultures) are reported to be more important for people with more travel experience. In contrast, people with lower levels of travel experience emphasize more motives such as *personal development* (e.g., developing my personal interests, knowing what I am capable of) and *relationship* (security) (e.g., feeling personally safe and secure). Finally, factors such as *nostalgia*, *romance* and *recognition* are considered to be the least important travel motives for people regardless of their level of travel experience.

2.3.3 Plog's model of personality types

Plog developed a psychocentric/allocentric model of personality types, which was first published in 1974 and has been applied since then to study the relationship between travel personalities and destination selection [106]. This model is based on the findings of a study by Plog and his colleagues, which was initiated in 1967 and in which they were asked to explore reasons why a certain amount of the population refuses to travel by airplanes. They claimed that non-flying people share three personality characteristics:

- *generalized anxieties*, these individuals suffer from a constantly anxiety
- *sense of powerlessness*, they believe that they have no influence on what is happening around them,
- *territory boundness*, they have not travelled a lot during their lifetime

Plog call such people *dependables* (originally, he called them psychocentrics) as they invest all their energy to make their lives predictable. According to Plog, the typical *dependable* personality can be characterized as being rather conservative and passive in his daily-life, less venturesome as well as desire to be surrounded by friends and family. In contrast, people showing the opposite behavior are called *venturers* (originally, allocentrics). The typical *venturer* makes decisions quickly, likes to choose new products, is curious about what is happening around and faces everyday life full of personal energy. In national samples the dependable and venturer dimensions are distributed on a normal curve (cf. Figure 2.10). The largest group falls into the segment of the *mid-centrics*, who exhibit a mixture of personality characteristics.

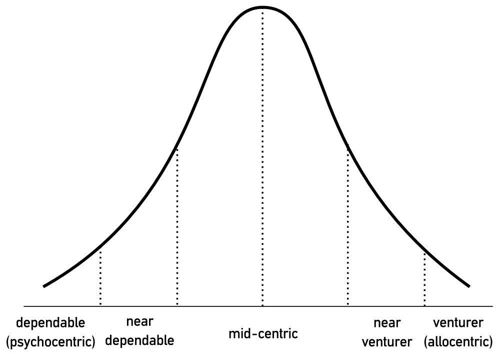


Figure 2.10
Plog's psychographic personality types [106].

These personality types are also useful in the domain of tourism as they determine travel patterns and preferences of tourists. Much research has been done in this field. (Table 2.5) outlines some differences in terms of travel patterns between the two personality types. Based

Table 2.5
Differences between dependables and ventures with respect to travel patterns (cf. Plog [106]).

Dependables	Venturers
• travel less frequently	• travel more frequently
• select well-defined, escorted tours	• prefer to be on their own, participate in local customs and habits
• prefer highly developed touristic spots	• prefer not crowded destinations
• are likely to return to a destination again	• tend to seek new destinations

on these travel personalities' characteristics, Plog explores in [106] *why destinations rise and fall in popularity*. According to his model, destinations follow a predictable but uncontrolled development pattern that comprises different stages, from *birth* to *maturity* and finally, to *decline*. He argues that at each stage, the destination appeals to a different segment of travelers. At first, a strange and rather unknown destination offers only a small range of tourism services (such as small hotels or restaurants). At this stage, the destination is only attractive to a few *venturers*, who like to experience the destination on their own. Over time, based on word-of-mouth recommendation, the destination becomes attractive to the group of *near-venturers* and more people are visiting the destination. As demand increases, the destination invests more money to provide more tourism services. As the destination becomes more touristic, *mid-centric tourists*, who comprise the largest group according to (Figure 2.10) become visitors of this destination. Gradually, the destination begins to look like many other overdeveloped destinations. Venturers refuse to visit this destination as it looses its distinctive character. On the other hand, *dependables* are likely to visit this destination, but they travel less and are a smaller group than the mid-centrics. Following this development, the destination's popularity declines.

Plog argues that that the ideal positioning for most destinations lies in the middle of the near-venturer segment as at this stage, a destination attracts the largest portion of the psychographic curve.

2.3.4 Yiannakis & Gibson's model of tourist roles

Based on a number of quantitative studies, Yiannakis and Gibson [156] identified 14 different tourist roles (cf. Table 2.6). For example, a *thrill seeker* can be described as type of person *interested in risky, exhilarating activities which provide emotional highs for the participant.* In order to delineate the roles from each other and to emphasize their individual characteristics, Yiannakis & Gibson proposed a three-dimensional framework to structure these types along three bipolar dimensions (cf. Figure 2.11).

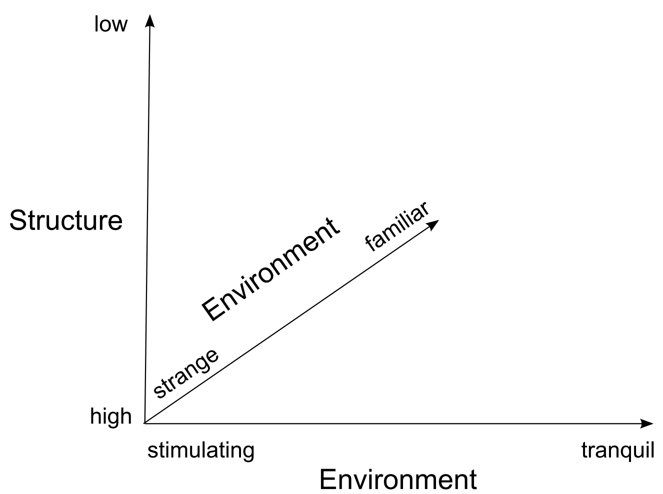


Figure 2.11 *Framework by Yiannakis & Gibson to position tourist roles [50].*

The three axes are defined as follows [50]:

- *Stimulation-Tranquillity:* Tourist roles preferring tranquil environments do not immerse in their host environment, they rather like to relax and escape from their daily routines. In contrast, certain tourist roles exhibit a high desire to interact with the host environment (e.g., by partaking in adventure or cultural activities).
- *Strangeness-Familiarity:* Those roles with a preference for familiar environments prefer traveling to destinations that are similar to their home environment, while roles scoring high on strangeness like to visit destinations that are rather dissimilar to their home environment.

- *High-Low Structure:* Tourist roles that prefer very structured travel arrangements make use of travel operators in order to fully plan their trip in advance. Roles preferring a low structure arrange their trip rather themselves and adapt their plan to peculiarities happening during their trip.

Table 2.6
A typology of leisure-based tourist roles by Yiannakis & Gibson [54].

Sun Lover	Jetsetter
Action Seeker	Seeker
Anthropologist	Independent Mass Tourist
Archaeologist	High Class Tourist
Organized Mass Tourist	Drifter
Thrill Seeker	Escapist
Explorer	Sport Lover
Educational Tourist	

Hence, *drifters* (cf. the description by Cohen [32]) seek stimulation and novelty in rather strange environments, while *thrill seekers* prefer more structured environments as their activities (e.g., bunjee jumping) are already connected with a higher amount of risk. In contrast, *organized mass tourists* are rather low-risk takers. They seek stimulation in familiar environments, while the *independent mass tourists* tend to travel to less familiar settings. They avoid package tours and are more spontaneous than the *organized mass tourists. Escapists* appear to avoid stimulation at all. As they suffer from stimulation overload during their work, they seek vacations without any form of stress and hectic life. Gibson & Yiannakis also reported some differences with respect to *gender*. For example, males associate the *organized mass tourist* with a higher setting for stimulation than females.

2.3.5 Typologies revisited

Push and pull motivations.

Understanding the travel-decision-process and the motivations why people travel is a complex problem. To explain touristic behavior, Dann already introduced in 1977 [35] the concepts of *push* and *pull factors*. *Push factors* are motivational variables internal to an individual, which predispose him/her to travel (e.g., escape, novelty). *Pull factors* relate to the characteristics of a particular destination. They govern the destination selection process and are crucial for the tourist to select his/her favourite destination.

Further research is needed.

As pointed out by Yiannakis & Gibson [50], the relationship between motivation (push) factors and preference for tourist roles needs to be further researched. In addition, Poria et al. [107] argue that most existing frameworks are classifications of reasons why people travel rather than a theory of travel motivation. Due to the interdisciplinary of tourism research, the frameworks utilize different theoretical foundations. It is still an open issue to reconcile them and to form a consolidated theory that explains the linkage between them.

(Table 2.7) depicts an overview of the different frameworks focusing on travel motives.

Table 2.7
An overview of frameworks targeting travel motives.

Authors	Purpose	Keywords	Tourist types & motivation patterns
Cohen [32]	characteristics and motivations of tourists	familiarity vs. strangeness	4 tourist types
Pearce & Lee [104]	travel motives and their relation to travel experience	travel career patterns	13 motivation factors
Plog [106]	destination's lifecycle	allocentrics vs. psychocentrics	5 personality types
Yiannakis & Gibson [156]	tourist roles	stimulation vs. tranquility, strangeness vs. familiarity, high vs. low structure	14 tourist types

Exploring the psychological needs and motives is often difficult as tourists are not able to express their real travel motives [35]. Apart from core motivation factors (cf. TCP model [104]), they might not be aware of other motivations that are a decisive factor for traveling and that push them to engage in tourist roles. However, to enact their preferred roles, they need destinations that provide an optimal balance of stimulation-tranquilty, familiarity-strangeness, and structure-independence [50]. Next to travel motives, also other variables such as socio-demographic variables (e.g., income, age, sex, education) and travel characteristics (e.g., party size, length of stay) have an influence on travel decisions [93].

Plog's model of allocentrics and psychocentrics.

Plog's model of *allocentrics* and *psychocentrics* is well known in tourism research. At the same time, it has been critized a lot. It is rather a model to explain destinations' attractiveness for certain tourist types and does not explain tourist motivations per se [131].

Critiques of Plog's model.

McKercher [91] critizes the assumption that destinations should be at (only) one definitive life cycle stage within the psychocentric-allocentric continuum. In fact, destinations attract people from different (geographic) markets. Therefore, destinations seem to exist at multiple stages simultaneously, as they might experience growth with some markets (e.g., Asian tourists) and may be in decline with others (European tourists).

Second, McKercher argues that markets can never evolve fully through the continuum. For example, New Zealand will never become a psychocentric destination for European tourists because of the large distance. In addition to physical distance, cultural distance has an impact on the markets as well. As Cohen states in [33], tourists like to experience novelty and change during their travel, but only to such an amount that is not threatening to them. In this way, adventure travels to natives in some outbacks will never attract the psychocentric market segment.

Litvin [86] tested Plog's model as well and conducted a study that revealed that Plog's model fails to represent *people's current travel patterns*. While Plog's model depicts that any population of tourists distributes approximately along a normal curve, the results of Litvin's study showed that the majority of tourists select *close-by, psychocentric destinations* rather than mid-centric destinations as predicted by Plog. However, if they were asked to describe their *ideal destination*, Plog's model proved to be rather robust. Hence, it seems that there is a gap between *intentions* and *actual travel behavior*. In this way, Plog's model rather reflects people's *generic travel attitudes*. Travel decisions are not only influenced by travelers' attitudes and needs, but assumed to be *rational* in the sense that travelers choose destinations based on other factors, such as *costs* or *length of stay*. Yet, the strength of Plog's model lies in its capability to identify the segment of tourists that might be willing to visit a specific destination [91].

Tourism activities relate travel motives to destinations.

Moscardo et al. [93] state that the critical link between *travel motives* and *destinations* are *activities*. Travel motives provide tourists with *expectations* for activities (e.g., visiting attractions, shopping) while destinations *offer* a specific set of activities to their visitors.

Tourists evaluate destinations with respect to the mix of activities they can perform there and which may satisfy their needs. The results of their study indicate that the three identified travel benefit clusters *social status*, *self-development* and *escape* had significantly different activity preferences. In detail, the *self-development group* preferred activities related to taking guided touris, visiting galleries and museums, and to go sightseeing, whereas the *escape group* tended to engage in sunbathing, beach activity and visiting entertainment places. Finally, the *social status group* preferred nightlife & shopping as well as casinos & gambling facilities, and activities such as golf and hunting.

Different travel motivations may be dominant during a decision-making process.

Andreu et al. [88] argue that motivations for travelling to a specific destination may not only vary from one person to another, but also from one decision-making process to the next. For example, a tourist may be visiting a *psychocentric destination* in summer in order to *relax*, but an *allocentric destination* in winter to seek adventure.

Gathering tourists' travel motivations based on stereotypical tourist types.

Therefore, it is essential to grasp the needs of a tourist *during the decision-making process*. The results of studies by Gretzel et al. in [57] confirmed that users can identify with particular travel types. Thereby, users preferred to choose several specific tourist types (e.g., sight seeker, culture creature) instead of a more generic one (e.g., all-arounder). Moreover, the studies showed that travel types are closely related to activities and that they can adequately be used in order to predict those activities, in which tourists wish to engage during the destination stay. However, no correspondence between travel personalities and travel destinations was found, as the destinations that were chosen for the study are rather homogenous and are able to serve the needs of different tourist types simultaneously.

In a next step, all the knowledge about the selected tourist types can be used to build a user model by inferring his/her preferences, which may be then used to propose appropriate destinations and to facilitate the choice of activities while being on vacation. In this way, tourist types facilitate the decision-making process. In addition, they are fun to use and allow users to revise their selection if the recommendations do not match their interests.

2.4 Semantic Web technologies and the tourism sector

This section starts by describing the vision of the Semantic Web and explains that ontologies are a core element to build Semantic Web-based applications. Following that, a number of tourism applications that exploit Semantic Web technologies are described. As they all make use of tourism ontologies, various classification schemes of tourism attractions are stated and existing ontologies in the tourism domain are evaluated. Finally, an overview of semantic similarity measures is given. Our matchmaking approach (see Section 3.4) uses such a measure in order to calculate the similarity between tourism resources and to facilitate profile propagation.

Applying Semantic Web technologies in the tourism domain.

The Semantic Web is the vision of a Web whereby all Web data is stored and linked in a machine-processable way [17].

Much research has been done to take advantage of Semantic Web technologies and apply them within the domain of tourism. As e-tourism is an information-intensive business, the tourism industry already utilizes Internet technologies in an extensive way. By the way, it has been regarded as a forerunner for using information technologies since the 1960s when airlines started to develop CRS/GDS (central reservation systems/global distribution systems) [153]. However, due to its large and complex structure (apart from few big players the tourism industry mainly consists of SMEs), tourism data is stored in heterogenous information systems which hamper exchange and integration of tourism resources difficult for tourism suppliers. At the same time, tourists searching the Web for travel-related services are confronted with a huge number of Web sites offering heterogenous tourism products. Often, these are described with different terms, thus making it hard to compare them. Hence, the tourists have to invest much time for their trip planning and decision making.

The goal is to counteract these problems and to facilitate the integration, linkage and finding of travel-related services and information with the help of *Semantic Web technologies*. To reach this goal it is necessary to semantically describe tourism resources based on *controlled vocabularies* so that they can be automatically accessed, processed and interpreted by machines.

Tourism resources can be semantically described by tourism ontologies in order to create a shared understanding of the domain.

One form of controlled vocabularies are *ontologies*. As such, they are a core element to realize the Semantic Web vision as they allow to annotate Web resources with information that enables machines to interpret and reuse these resources across different information systems. According to Gruber [59], an ontology is *a specification of a conceptualization*. Thereby, a conceptualization is a simplified view on the part of the world that should be represented in a formal way through a set of *concepts* and *relationships* between them. In this way, an ontology provides a shared description of the domain of interest.

The application of ontologies has the potential to cope with a number of challenging requirements related to the tourism sector. Firstly, having a common vocabulary compensates the *interoperability problem* [37] that comes along with the integration of heterogeneous data sources.

This vocabulary is important to bridge between the languages of the different information systems and helps to create a shared representation of meaning. Secondly, an ontology provides a formal basis which is the prerequisite for *formal rule statement creation* and *inferential analysis* in the tourism domain [38]. Thirdly, an ontology provides enhanced possibilities for *information search* and for *automatic discovery, negotiation* and *adaptation/personalization* of tourism services [22].

In this way, *ontology-based models* are a promising approach for modeling tourism information because of their high expressiveness and reasoning capabilities [83]. By understanding the semantics of user queries and reasoning over service descriptions, personalized recommendations can be provided based on *matching techniques* between prospect *tourist demands* and *service offerings*.

The following example explains how three semantic sources can be aligned in order to relate certain types of attractions to the preferences and interests of tourists. These interests can be aggregated to describe certain *tourist types* such as an *art lover*. Using this approach, the semantic mapping can take place on a *more abstract level*, i.e., between tourist objects and tourist types.

A geographical information knowledge base of the city of Vienna might contain the statements:

$$AustrianTheatreMuseum : GeoSite$$
$$AustrianTheatreMuseum\ hasLocation\ Vienna$$

A destination management organization might have defined the additional statement:

$$AustrianTheatreMuseum\ isAttractiveFor\ (Person \sqcap \exists interestedIn.Art)$$

The official Vienna tourism board might have developed a tourism ontology including following statements:

$$ArtLover := Person \sqcap \exists interestedIn.Art$$

$$CulturalAttraction := GeoSite \sqcap \exists isAttractiveFor.ArtLover$$

Based on these concept definitions and by leveraging the pertinent information from the other semantic sources, it is possible to infer that the `Austrian theatre museum` is a cultural attraction for tourists currently staying in Vienna, who are interested in arts. In order to facilitate *semantic matching* between tourism objects and user profile, a specific vocabulary for the (tourism) domain, user type and contextual information such as time and location is needed [16].

2.4.1 Semantic Web-based tourism applications

Tourism recommendation systems that exploit Semantic Web technologies.

In the following, we describe a number of *tourism applications* that take advantage of Semantic Web technologies. [97] proposes a framework which uses Semantic Web technologies to support customers in finding suitable *hotels*. The framework provides the integration of hotel data and additional location-based information regarding *points of interest*, *restaurants* or *WLAN hotspots*. Both user requirements and aggregated hotel information are described at a conceptual level and compared based on semantic similarity, thus achieving a selection of the hotels ranked according to the customer's requirements. In this project, two main ontologies, comprising a *person* and a *hotel ontology*, together with small sub-ontologies to describe *hotel features*, *points of interest* and *means of transportation* have been developed.

A prototypical *knowledge-based e-tourism system* that leverages ontologies in order to recommend tourism information services is presented in [36]. Two ontologies were developed in order to enable the system to provide both destination and accommodation recommendation capabilities. The *Destination Context Ontology (DCO)* describes the different dimensions of a destination at a semantic level, including *security*, *population size*, *flow of traffic*, *weather temperature* or *crime rate*. The *Accommodation Ontology (AO)* is used to describe the attributes of the various types of tourism accommodation and was modeled based on the *Harmonise Ontology* [49]. The main concepts are *accommodation*, *attraction*, *facilities*, *services*, *gastro* and *state*. Both ontologies are encoded in OWL-DL. Semantic Web technologies are used to compare the content descriptions of accommodation instances with the preferences of the user based on semantic similarity in order to generate a list of recommendations.

The work done in [15] presents a partial ontological model for *cultural spaces* and describe an approach to link that model to touristic services, including *means of transport*, *hotels* and *restaurants* as well as *information services*. For describing the cultural spaces semantically, concepts from different standard vocabularies are integrated. In detail, these comprise the upper-level ontology OpenCyc 2.9, the Art & Architecture Thesaurus (AAT) vocabulary in order to describe art styles as well as the standard for learning object metadata IEEE LOM to represent basic information objects. A cultural tourism itinerary generator for the city of Alcalá de Henares in Spain has been developed that proposes a tourist route through the city. The route is customized based on the preferences of the user regarding historical figures or architectural styles which have to be selected in advance.

[75] presents an ontology-based trip planner using semantic web technologies. Users can specify their travel requirements (e.g., departure date, stay duration, number of stars of accommodation, means of transportation) in a graphical user interface. The system matches these requirements with vendor offerings in order to propose an appropriate itinerary. The AuSTO (Australian Sustainable Tourism Ontology) ontology embodies knowledge about the tourism domain and consists of concepts such as *involved party* (e.g., vendor or traveler), *traveler requirements*, *offering*, *resource* (storing information regarding tourism resources), *traveler preference*, tourism related *events* (e.g., conference vent, entertainment event), *traveler* (basic information about the traveler) and *destination* (e.g., city).

2.4.2 Classification schemes of tourism attractions

In the following, we describe different *classification schemes* found in literature that focus on tourism attractions. These comprise rather abstract typologies (cf. Prentice's typology [108]), more detailed classifications (cf.

the National Monuments Record Thesaurus of Monuments Type [2]) as well as non-hierarchical typologies (cf. Richard's typology [117]). Furthermore, we have a look at *existing tourism ontologies* (cf. Section 2.4.3) and explore whether they include concepts that are related to tourism attractions.

In 1993, Prentice [108] proposed a general typology of heritage attractions, which are visited by tourists and visitors (cf. Table 2.8). The list also contains terms for which the term 'tourism attraction' might be not well suited, such as the term 'genocide monuments' but which are visited by many people as well.

The National Monuments Record (NMR) is the public archive of English Heritage and covers millions of plans, photographs and reports concerning England's archaeology, architecture, social and local history. To classify this knowledge, different thesauri have been developed. A *thesaurus* is a collection of terms, which are related to each other over different relationships such as *broader term* (BT) and *narrower term* (NT), which reflect the hierarchical structure of the classification as well as *related term* (RT) which points to similar subjects [66]. One of the thesauri is the *Thesaurus of Monuments Types* [2] which classifies types of monuments relating to the built and buried heritage in England. It contains about 7500 terms, which are categorized under 18 main classes. (Figure 2.12) depicts a fragment of this thesaurus. It shows some information related to the term 'art gallery'. We get to know that this term is contained in the class 'recreational'. The broader term is 'art and education venue', but it is also related to the term 'commercial art gallery'. Although this

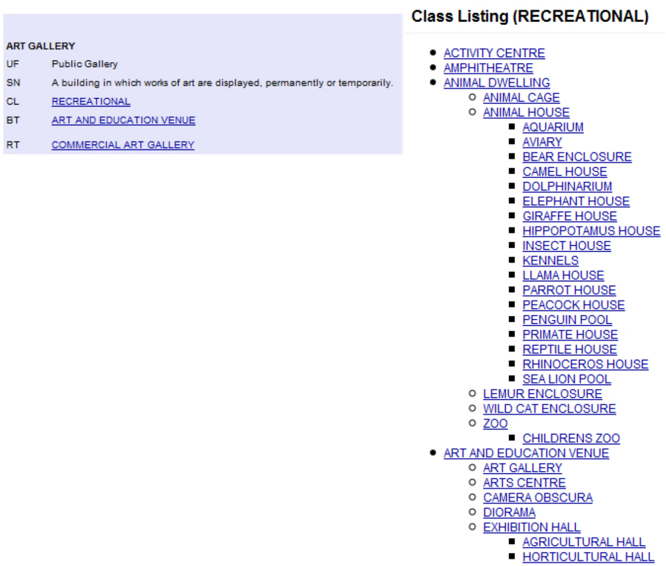


Figure 2.12
A fragment of the Thesaurus of Monuments Types, depicting the term 'art gallery' [2].

thesaurus is quite comprehensive, in [143] some issues are mentioned that arise when the thesaurus is incorporated in a tourism ontology. First, as this thesaurus is owned and maintained by English Heritage, it only can be used under a license which prevents modification of the reference data set. In addition, it contains very specific terms that might only be adequate to classify monuments in England as these terms are not used somewhere else.

Table 2.8
Prentice's typology of heritage attractions.

Term	Description
Natural history attractions	Nature reserves, wildlife parks, cliffs, etc.
Science-based attractions	Science museums, technology centres, etc.
Attractions concerned with primary production	Farming museums, vineyards, fishing, etc.
Craft centres and craft workshops	Water & windmills, Potters, Woodcarvers, Glass Makers, etc.
Attractions concerned with manufacturing industry	Pottery and porcelain factories, breweries, cider factories, etc.
Transport attractions	Preserved railways, canals, civil aviation, motor vehicles, etc.
Socio-cultural attractions	Prehistoric and historic sites (domestic houses, costume museums)
Attractions associated with historic persons	Sites and areas associated with writers and painters
Performing arts attractions	Theatres, street-based performing arts, circuses, etc.
Pleasure gardens	Ornamental gardens, period gardens, etc.
Theme parks	Nostalgia parks, adventure parks, fairytale parks, etc.
Galleries	Art galleries, etc.
Festivals and pageants	Historic fairs, festivals, etc.
Stately and ancestral homes	Palaces, country homes, manor houses
Religious attractions	Cathedrals, churches, abbeys, priories, etc.
Military attractions	Castles, battlefields, military museums
Genocide monuments	Sites associated with the extermination of other races
Towns and townscape	Historic townscape, buildings in an urban setting
Villages and hamlets	'Rural' settlements, usually of pre-twentieth century architecture
Countryside and treasured landscapes	National parks, other countryside amenity designations
Seaside resorts and 'seascapes'	Seaside towns of past eras and marine 'landscapes'
Regions	Pays, landes, counties, or other historic or geographical areas

Within the PICTURE project [143], which targets at exploring the impact of cultural tourism on the urban environment, a conceptual layer has been proposed that distinguishes *attractors* in *objects*, *events* and *places*. Based on a literature review and discussions with the project partners, a set of attractors has been identified (comprising ones such as fountain, church, theatre, garden) which are classified under these terms.

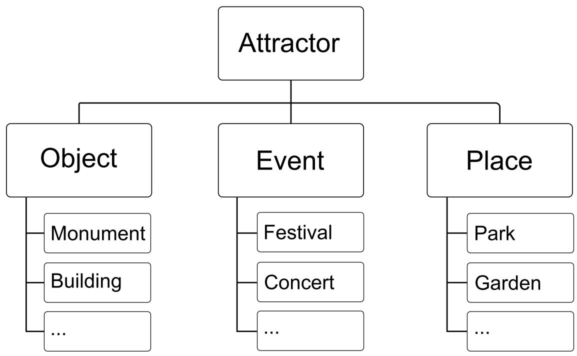


Figure 2.13 *Classification of attractor types proposed within the PICTURE project [143].*

Richards [117] presents a non-hierarchical typology, in which *cultural attractions* are positioned in a field that is spanned by two axes named *Function* and *Form*. The *Function* axis covers the dimensions *Educa-*

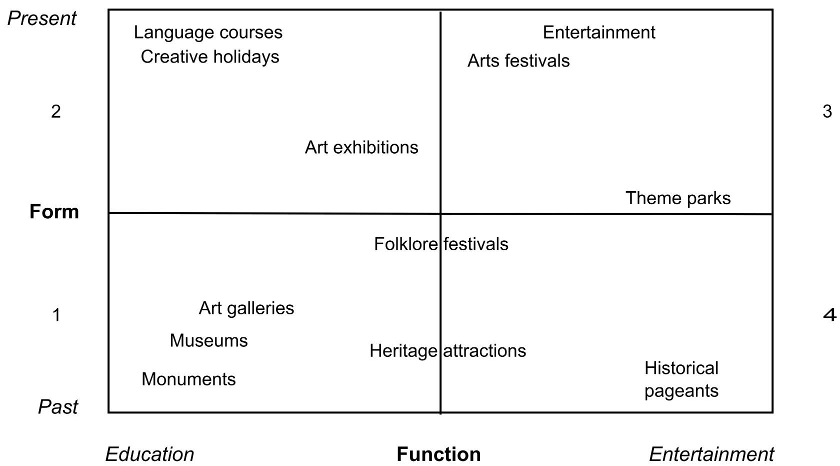


Figure 2.14 *A typology of cultural attractions along two axes labelled function and form [117].*

tion to *Entertainment* while the *Form* axis goes from *Past* to *Present*. In this way, *quadrant 1* contains rather traditional cultural attractions that are related to cultural products of the past such as *museums*, *art galleries* and *monuments*. Quadrant 2 focuses on contemporary cultural attractions that have an educational setting such as *language courses* or *art exhibitions*. Quadrants 3 and 4 reflect attractions that are related to entertainment, ranging from present *art festivals* to *historical pageants*.

In this typology, there are no distinct categories anymore. Attractions can be described based on different dimensions that reflect various characteristics of attractions. Taxonomies that show a pure hierarchical structure are well suited to describe what kind of thing the object is. An attraction is, for example, either a museum or a church. In this way, an attraction is assigned to predefined categories. Certain characteristics of tourism objects being classified are however more imprecise and uncertain and have to be modeled in a quantitative way in order to answer questions like 'How adventurous is the fun-fair? How historically afflicted is the St. Stephen's Cathedral compared to the Schönbrunn Palace in Vienna?' In order to model such knowledge, strict taxonomy- or ontology-based approaches can be relaxed through the use of *Bayesian probabilities*, *vector-based models* or *fuzzy sets* [143]. As all typologies are an approximation of reality, it depends on the focus of an application and on the user needs to determine which characteristics should be covered in the classification scheme.

2.4.3 An evaluation of existing tourism ontologies

With the increasing number of Semantic Web tourism application, there is a proliferation of ontologies (Harmonise [49], an OWL Ontology for E-Tourism [27], ebSemantics [129], DERI OnTour [39], EON Traveling ontology [43]) that have been developed in the area of e-tourism either by industry, academia or within collaborative projects. Based on a literature review, we have analyzed a number of tourism ontologies [16], thereby focusing on their purpose, set of vocabularies as well as on their advantages and shortcomings. Most of them focus on the domain of tourism in general and do not exhibit a detailed set of different tourism attraction types. To put it in other words, most of them just cover the concept *tourism attraction* but restrain from a more fine-granular classification.

In the following, those ontologies are described that contain more specific attraction types. However, they do not represent a uniform representation of tourism attractions. Each of them classifies attractions in a different form. Thus, it may be very difficult to exchange knowledge about tourism attractions across different ontologies.

(Figure 2.15) depicts the most relevant ones and compares them based on their support for modeling tourism objects, tourism activities, tourism events and user context. Furthermore, in case that they are linked with upper-level ontologies these are mentioned as well.

	Tourism Objects								
	Accommodation	Transport	Food Service Org.	Attraction	Tourism Activities	Tourism Events	Opening Hours	Location	Upper Level Ontologies
QALL-ME	✓	✓	✓	✓	✓	✓	✓	~	WordNet, SUMO
CRUZAR			✓	✓	✓	✓	✓	~	DOLCE, SKOS, FOAF, W3C Geo
SPETA	✓			✓	✓	✓	✓	~	DBpedia, YAGO, FOAF, W3C Geo
INREDIS	✓			✓		✓			SKOS, W3C Geo
Harmonise	✓					✓	~	~	
GETESS				✓					

Legend:

✓	supported
~	partially supported
	not supported

Figure 2.15
Evaluation of relevant tourism ontologies.

According to our evaluation, they show following shortcomings:

- Their vocabulary covers only a limited set of tourism concepts, due to a restricted application scope.
- They hardly integrate existing, domain-independent ontologies, that already define commonly used concepts such as time, currency or geo-location.
- They often contain a diversified mix of concepts, thus making ontology mediation and (automatic) exchange of data a difficult task.

As the tourism ontologies are partly based on upper-level ontologies, (Table 2.9) provides some details about them.

The QALL-ME Ontology.

The QALL-ME Ontology. QALL-ME [94] is an EU-funded project that aims at establishing a shared infrastructure for multilingual and multimodal question answering in the tourism domain. Thereby, it allows users to pose natural questions in different languages using a variety of input devices and returns a list of answers in the most appropriate modality. The QALL-ME ontology provides a conceptualized description of several aspects of the tourism domain, including *tourism destinations*, *tourism sites*, *tourism events* and *transportation.* It contains 122 classes and 107 properties that indicate the relationships among the classes. QALL-ME is encoded in the ontology language OWL-DL.

The CRUZAR Ontology.

The CRUZAR Ontology. CRUZAR [9] is based on the upper-ontology DOLCE in order to model *visitor's profiles*, *travel routes* and *attractions.* To describe attractions, it further reuses properties from the Dublin Core, FOAF and SKOS-Core.The *CRUZAR ontology* [9] contains the concept *tourism resource* in order to depict the places of interest which tourists tend to visit. In detail, 8 subclasses are distinguished that

Upper-level Ontologies	
DBpedia	DBpedia extracts structured data from Wikipedia.
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering. It aims at capturing the ontological categories underlying natural language and human commonsense.
FOAF	The Friend of a Friend (FOAF) project describes people, the links between them and the things they create and do in a machine-readable way.
GETTY	The Getty vocabularies contain terms, names, and other information about people, places, things, and concepts relating to art, architecture, and material culture. For example, the Art & Architecture Thesaurs (AAT) comprises terms related to fine art and architecture, while the Thesaurus of Geographic Names (TGN) contains around 1 million records about names, place types and coordinates.
OpenCyc	OpenCyc is the open source version of the Cyc technology, one of the largest general knowledge bases.
SKOS	The Simple Knowledge Organization System (SKOS) is a common data model for sharing and linking knowledge organization systems via the Semantic Web.
SUMO	The Suggested Upper Merged Ontology (SUMO) was created by merging publicly available ontological content into a single and comprehensive structure. As it is also aligned with WordNet, it is used in text search and linguistics and reasoning.
UMBEL	The Upper Mapping and Binding Exchange Layer(UMBEL) is a lightweight ontology for relating external ontologies and their classes to UMBEL subject concepts. It links to the most fundamental concepts from within OpenCyc.
W3C Geo	The W3C Geo is a basic RDF vocabulary that provides namespaces for representing geo locations using WGS84 as a reference datum.
YAGO	YAGO is part of the YAGO-NAGA project at the Max-Planck Institute for Informatics in Saarbrücken/Germany. It stores about more than 2 million entities (like persons, organizations, cities, etc.) and knows about 20 million facts about these entities.

Table 2.9 *Overview of upper-level ontologies and their purpose.*

are categorized in *monuments* and *green zones*. As depicted in (Figure 2.17) the subclasses of *monument* are rather mixed. One might argue whether *buildings* should be categorized under *monuments*.

The SPETA Ontology.

The SPETA Ontology. SPETA [52] exploits concepts from the *e-tourism ontology* [27] in order to describe tourist services. In addition, it links to concepts from DBpedia and YAGO to describe concepts such as *attractions* or *activities* and FOAF to describe *social links* of tourists. The *SPETA ontology* [52] models more detailed knowledge regarding different types of attractions. The main categories are depicted in (Figure 2.18). These categories contain more subcategories that are not shown in this figure due to space restrictions. The ontology allows to represent different characteristics of attractions. An attraction can be related to a *specific theme* (culture, leisure, modern art, etc.). The *context* of an attraction can be modeled, including *weather constraints* as well as *location* (indoor or outdoor) and *time* context (opening-hours). Rules can be defined to

Figure 2.16 QALL-ME ontology [94]. Fragment depicting different types of attractions.

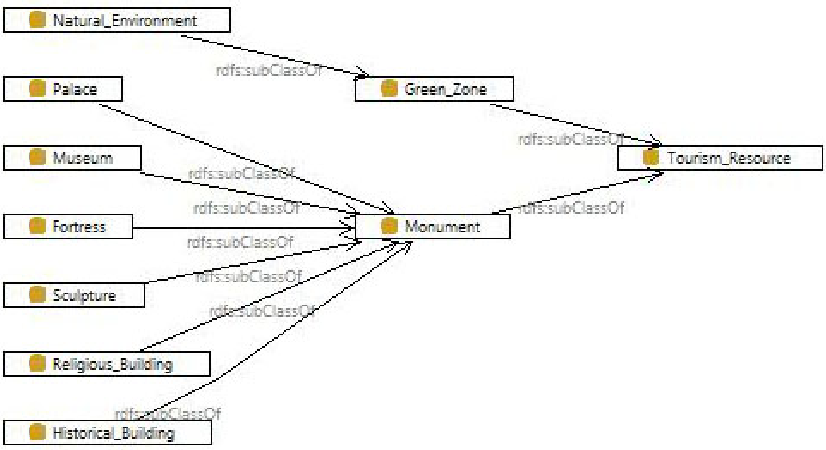


Figure 2.17 CRUZAR ontology [9]. Fragment depicting different types of attractions.

constrain the assumption that an attraction is always open. They allow to define that an attraction is closed for a specific time frame each day or is even closed for a whole day.

The INREDIS Ontology.

The INREDIS Ontology. INREDIS [26] classifies *points of interest* in 4 different types, comprising *events*, *attractions*, *restaurants* and *accommodation* (cf. Figure 2.19). Thereby, each point of interest can be further described in form of *feature types*. However, only feature types that describe *attractions* have been modeled in the ontology as only data about tourism attractions was available during the development of the ontology. Attractions are thus represented through following *feature types*. The *entrance type* describes whether the attraction provides free or discounted entrance. The *POI facility type* describes different services offered at the attraction's site (e.g., restaurants, shops) as well as whether the attraction is accessible for disabled visitors. The *educational subject* categorizes the attraction by the type of education offered to the tourist, including sports, music, archaeology, art, history, religion and science.

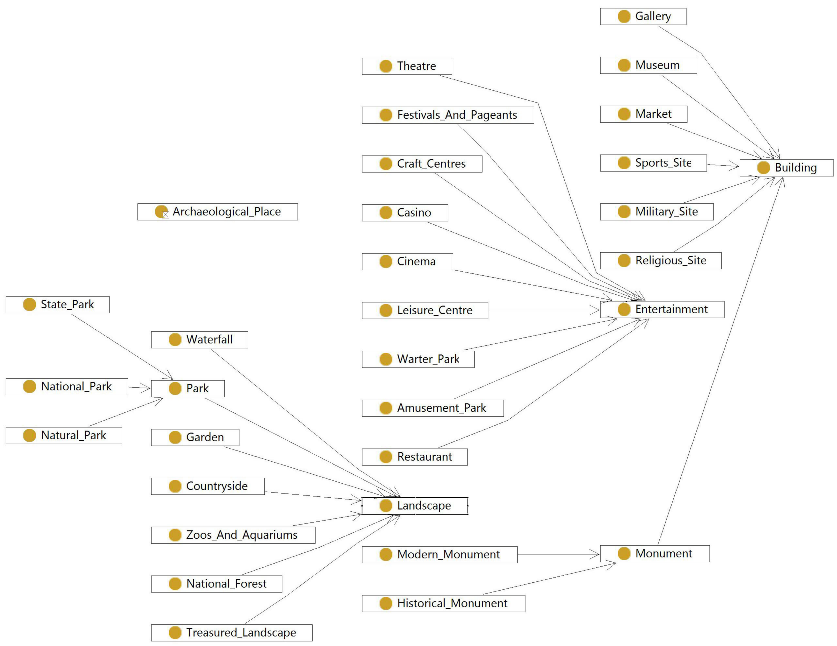


Figure 2.18 *SPETA ontology [52]. Fragment depicting different types of attractions.*

The *functional type* classifies attractions according to their function. It comprises the cultural, architectural, natural and recreational type. For example, the architectural type includes different kinds of buildings (maritime, civil, urban, religious buildings) as well as distinguishes among different architectural styles such as romanesque, ghotic or renaissance.

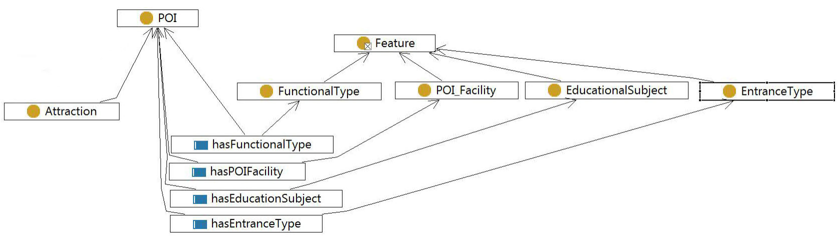


Figure 2.19 *INREDIS ontology [26]. Fragment depicting different types of points of interest and their features.*

The Harmonise Ontology.

The Harmonise Ontology. The Harmonise [49] ontology, first developed within the Harmonise project, is now the central element within the HarmoNET (Harmonisation Network for the Exchange of Travel and Tourism Information) that aims to create an international network for harmonization and data exchange in the tourism industry. The ontology focuses on two sub-domains of the tourism domain, namely *events* (e.g., conferences, sport) and *accommodation* (private rooms, hotels, guesthouses), modeled in the language RDFS. Members of this network can share data by mapping their specific data model to the Harmonise ontology, which acts as the central data model of the network. The mapping proceeds at the site of each individual member, since there is a proprietary mapping between the member's legacy system and the Harmo-TEN ontology.

The GETESS Ontology.

The GETESS Ontology. The *GETESS* (German Text Exploitation and Search System) [132] project focused on retrieving information

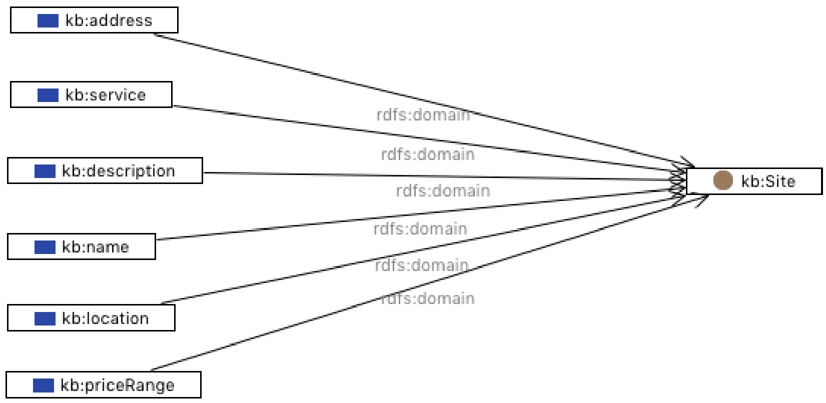


Figure 2.20 *Harmonise ontology [52]. Fragment depicting different attributes of a site.*

from touristic websites. This information is semantically interpreted and can be queried by the user through natural language processing techniques. It contains 1043 concepts and 201 relations. As depicted in (Figure 2.21), the *GETESS ontology* [132] represents the concept *sight*, but lists all different types of sights more or less as a flat list under this concept. For example, the concept *museum*, *church* and *fortress* appear at the same level of the hierarchy (all of them are direct subclasses of *sights*). No further classification is available to group them into different categories according to their individual characteristics.

Further ontologies.

Further ontologies. The *Hi-Touch ontology* [83] was developed during the IST/CRAFT European Program Hi-Touch, which aimed at establishing Semantic Web methodologies and tools for intra-European sustainable tourism. The goal was to formalize knowledge on travelers' expectations and to propose customized tourism products. The ontology was mainly developed by the company 'Mondeca' and is encoded in the ontology language OWL. The ontology classifies tourist objects, which are linked together in a network by semantic relationships. The semantic network is provided by a Topic Map. The top-level classes of the Hi-Touch ontology are *documents* (any kind of documentation about a tourism product), *objects* (the tourism objects themselves) and *publication* (a document created from the results of a query, e.g., a PDF document). The tourism objects can be further indexed by keywords using the thesaurus on tourism and leisure activities by the World Tourism Organisation [99]. This standard terminology ensures the consistency of the tourism resources categorization managed on distributed databases and enables semantic query functionalities. *DTG's ontology* is built leveraging some existing taxonomies from DAML and GETTY. The *DERI e-Tourism ontology* [39] was developed by STI Innsbruck. The ontology focuses on the description of accommodations and infrastructure and should facilitate a user who queries a tourism portal to find a package of relevant accommodations and infrastructure. The *Travel Agent Game in Agentcities (TAGA)* [137] is an agent framework for simulating the global travel market on the Web. The TAGA ontology defines travel concepts such as *itineraries*, *customers*, *travel services*, and *service reservations* as well as different types of *auctions*. The *EON Travelling Ontology* [43] was developed by the Institut National de l'Audiovisuel in France. It

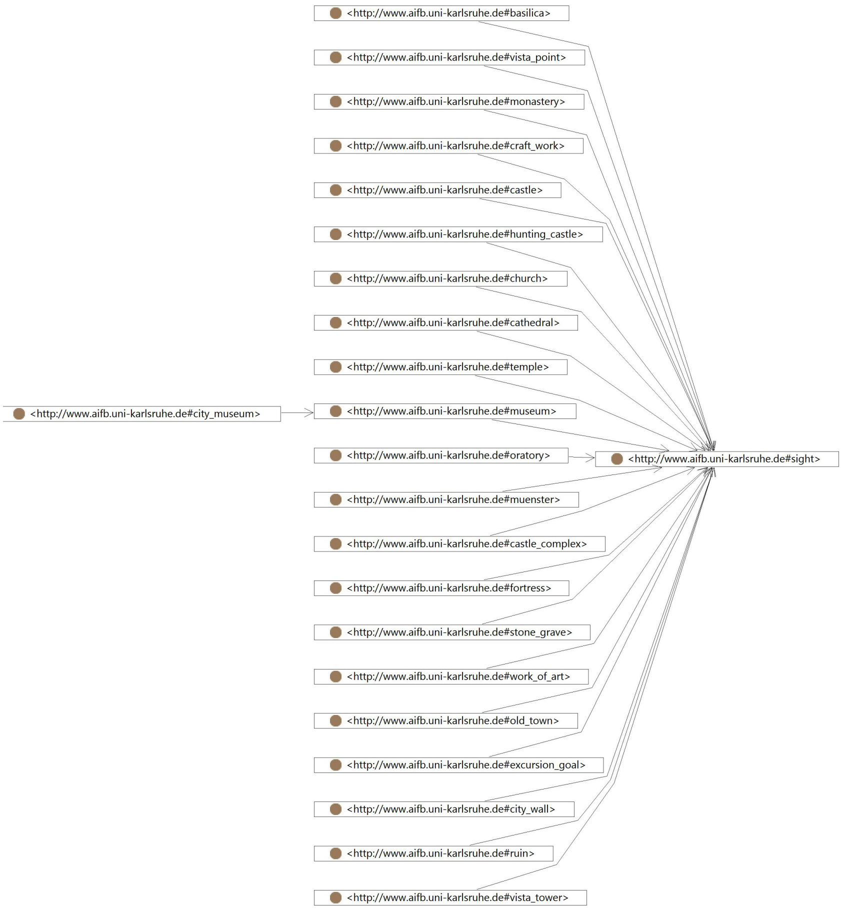


Figure 2.21 *GETESS ontology [132]. Fragment depicting different types of sights.*

describes tourism concepts that are divided into *temporal entities* (e.g., reservations) and *spatial entities*, which further comprise *dynamic artefacts* (e.g., means of transportation) as well as *static artefacts* which comprise *town sights* or *lodging facilities.*

Ontologies revisited

In the following, we briefly summarize the results of the evaluation of existing tourism ontologies [16].

Semantic matching between tourism objects and tourist types. Existing tourism ontologies neglect to model user preferences, which are fundamental to provide personalized information about tourism objects. In the tourism domain, these user preferences can be aggregated to a set of tourist types.

Temporal context of tourism objects. Information about tourism objects is of little value without a valid temporal description. Receiving up-to-date information about events, opening-hours of points of interest, or delays of means of transportation is essential for tourists in order to have a successful trip. Existing tourism ontologies rather neglect temporal information. Some of them create proprietary concepts to model time or date periods.

Location context of tourism objects. Information needs in the tourism domain are typically assigned to a geographical context. Current ontologies rather use a simple geospatial model.

Boolean values and the open world assumption (OWA). Organizations tend to store their data in relational databases, thereby following a closed world assumption. This means that if something cannot be found in these databases, it is assumed to be absent and therefore false. As the uplifting of existing resources to an ontological level is a very labor-intensive process, using Boolean values in ontologies is still a common way, although it is regarded as a bad style. Following example does not model the intended semantics:

```
CafeSacher live-music 'true'
```

Does this mean that there is live music? Every day or at certain recurrent days? How do we get information about the band? What, if the boolean value is missing at all? In this case, a database would infer that the `Café Sacher` does not offer live-music. However, the OWA assumes incomplete information by default, which means that a fact can be true even if it is not present in the knowledge base. Something is false only if it can be proved to contradict other information in the ontology.

2.4.4 Semantic similarity measures

In this section, we give an overview of relevant semantic similarity metrics based on a comprehensive literature review. In detail, we outline their basic principles and for each metric, we give an example to illustrate their approach and discuss advantages and disadvantages.

In literature, a number of models have been proposed to define similarity measures between objects of interests, focusing, for example, on similarity between strings [65], vectors [120] or trees [150]. Similarity measures are heavily utilized in different fields such as data-mining, information retrieval, matchmaking or recommendation systems [6].

In the following we restrict to the set of *semantic measures* that have become popular with the development of the Semantic Web and present a set of approaches that have been proposed in literature. Such measures are used in different areas. For example, they are used to enhance information search in unstructured documents. In this way, ontological knowledge (e.g., in form of a thesaurus such as WordNet) is exploited to improve the similarity measure between documents by evaluating the semantic relations between relevant terms.

If we, for instance, consider the terms `Austrian Capital` and `Vienna` and apply string-based similarity measures, the two terms would be rather dissimilar. If, however, we take into account the information from WordNet that defines the term `Austrian Capital` as synonym for the term `Vienna`, the result can be greatly improved. In this case, semantic-based measures clearly outperform purely syntactic methods, which have been previously used in information-extraction.

Semantic measures exploit ontological knowledge (i.e., descriptions of objects and their relationships, for a more formal definition cf. [59]) to find objects which are conceptually close but not identical.

Single ontology vs. cross ontology measures.

Basically, one has to distinguish between *single ontology* semantic measures that compute the similarity between nodes within the same ontology and *cross ontology* methods that compare nodes or set of nodes from *different ontologies* [147]. Methods that enable the comparison of nodes between different ontologies belong to the field of *ontology matching* (cf. [44] for an overview of ontology matching). In our case, we assume that all required domain knowledge is modeled within a single ontology.

Semantic relatedness is the broader term of semantic similarity.

In literature, different terms are used to describe the closeness between nodes in an ontology, comprising terms such as *semantic relatedness*, *semantic similarity* and *semantic distance* [19]. Semantic relatedness evaluates the resemblance between nodes based on the *strength* of their semantic link. The strength of the semantic link depends on their *relationship with other nodes*, the *number of shared properties*, and especially their *level* (general vs. specific) within the hierarchical component of an ontology [127].

The strength of a semantic link between nodes is composed of the type of relation, the shared properties and their depth in hierarchy.

In contrast, semantic similarity calculates the closeness between two nodes based on a subset of semantic links (e.g., considering only is-a relations). In this way, semantic similarity can be considered as a special case of semantic relatedness. For example, the terms `car` and `gasoline` seem to be more closely related than `car` and `bicycle`, but `car` and `bicycle` are more similar according to their taxonomic context [112], as both are some means of transportation. In the following, the term *semantic similarity* is used to describe the closeness between two nodes. If a measure is more complex and utilizes different forms of semantic links and thus rather belongs to the class of *semantic relatedness measurements*, it is explicitly stated.

The *semantic distance* is an inverse notion to the semantic relatedness and semantic similarity. If the similarity value between two given nodes scales between [0..1], the semantic distance is stated as 1 - similarity value and vice versa.

According to [161], ontology-based measures can be classified into three main categories:

Path-based measures.

- *Path-based measures:* These measures, also called *edge-based measures*, are based on the path length between nodes. They make use of the fact that nodes within an ontology are usually organized hierarchically in form of a tree. More specific nodes are located nearer to the leaves, while more general nodes are located nearer to the root of the tree. The most intuitive similarity measure of two nodes in an ontology is to calculate the shortest path between them (i.e., by counting the number of edges on the path connecting them). In this simple case, every edge represents the same semantic information and is weighted with value 1. The shorter the path between these nodes is, the more related they are.

  For example, `science museums` are closer related to `natural history museums` than to `palaces`. To model this statement in an ontology, `science museums` and `natural history museums` might be both subclasses of the node `museum`, which has, together with the node `palace`, the parent node `building`. In this case, only

the taxonomic is-a relations have been considered to calculate the semantic similarity. In contrast to a taxonomy, which is restricted to is-a relations, an ontology allows to use a diverse set of relation types. Hence, nodes can be linked in an arbitrary way (e.g., utilizing user-defined relations such as 'part-of' or 'related-to'). These various relation types obviously have different meanings, which should be reflected by assigning different weights to the various relation types, which are represented through edges in the graph.
In addition, some exploit the level of generality or specificity of nodes in an ontology (i.e., their taxonomic position) to compute weights on edges. Nodes that are close together and lower in the hierarchy (more specific ones) should have a higher similarity than nodes that are equally close but more nearer to the root (more general ones) [147].

- *Information-based measures:* Such measures are similar to the path-based measures, but assign weights to nodes rather than to edges. The weight represents the information content of the node, which is expressed as its probability of occurrence in a text corpus. The more specific a concept is, the higher is its information content.

Information-based measures are also called node-based measures.

- *Feature-based measures:* In contrast to the path-based measures, which exploit the taxonomic relationships between nodes, these measures look at the set of properties (i.e., the features) that are used to describe the characteristics of each node. The similarity betweeen two nodes depends on the number of common and distinct features. While common features increase the similarity, non common features rather decrease it.

Feature-based measures rely on the shared properties between nodes.

- *Hybrid measures:* Such measures combine the advantages of the different measurements [77]. Othman et al. propose a hybrid measure in [100] that combines path-based and information content measures, and considers node depth as well as node density within the taxonomy. Cross ontology methods make use of hybrid and feature-based methods as the structure of different ontologies are not directly comparable and therefore edge-based methods may not be adequate [147].

Hybrid measures combine edge- and node-based measures.

Following parameters have an impact on the quality of the similarity measures [119]:

- *Node density:* The density of an ontology (distribution of nodes within the graph) is in most cases not equally balanced. Some parts in an ontology are more dense than other parts. Consequently, an edge in a dense part represents a smaller distance than an edge in a more wider part. For example, the ontology might contain a high number of different `museum` categories, each category having again many subclasses. In contrast, the category `palace` might only have two subclasses. The distance between these two subclasses is obviously greater than the distance between two subclasses of the category `science museum`. Hence, if there are irregular densities of edges between the nodes, this might have a negative impact on the similarity metric.

Node density.

- *Depth in hierarchy:* The depth of the nodes within the hierarchical component of the ontology is an important means to indicate the

Depth in hierarchy.

similarity between nodes. The reason for this is that the distance shrinks as one goes down in the hierarchy. Adjacent nodes, located at the top, are more general and therefore have a wider distance than nodes located nearer to the leaves, which are more specific.

- *Strength of connotation between parent and child nodes:* The weights for a specific type of relation (e.g., is-a relation) between a parent and its child nodes might not be equal. In [119], an example is stated whereby the parent node `life form` is more strongly connected with its child nodes `animal`, `plant` and `person` than with other nodes such as `plankton`. A solution to this problem is to estimate the weights of the various is-a relations by taking into account the information content of the single nodes.

  Strength of connotation.

- *Type of relations:* Edge-based measures are typically based on is-a relations to calculate the similarity between two nodes. However, to improve the similarity metric, other types of relations (e.g., domain-specific ones) should be considered as well. This certainly leads to a more realistic estimation of the semantic strength between two nodes.

  Mixed relations.

- *Parent vs. sibling [13]:* The distance between siblings should be greater than the distance between parent and child nodes.

  Parent vs. sibling similarity.

In the following, we discuss 11 similarity measures (cf. Table 2.10), which are classified with respect to the categories listed above.
We describe their theoretic foundation and outline which parameters they use in order to calculate the semantic link strength between adjacent nodes, and finally the distance between arbitrary nodes. For that, we use the definitions depicted in (Table 2.11).

Table 2.10 *Summary of similarity measures discussed in this section.*

Measure	Basic principle
Path-based	
Rada et al. [109]	Shortest path length between nodes
Wu & Palmer [154]	Shortest path length to most specific common parent (ccp), scaled by depth of ccp
Zhong et al. [162]	Shortest path length to most specific common parent (ccp), scaled by weighted depth of nodes
Sussna [136]	Shortest path length between nodes, considering type & number of relations as well as depth within hierarchy
Information-based	
Resnik [112]	Information content of most informative subsumer (mis)
Seco et al. [126]	Calculation of an intrinsic information content for each node
Feature-based	
Tversky [142]	Common vs. distinctive properties of nodes
Knappe et al. [81]	Ratio shared nodes vs. 'own' nodes, scaled by generalization and specialization factor
Hybrid	
Jiang & Conrath [77]	Extension of Resnik, accounting for path between nodes and most informative subsumer (mis)
Lin [85]	Extension of Resnik
Mazuel et al. [90]	Extension of Jiang & Conrath and Knappe et al., shortest mixed-relation path between nodes, which is semantically correct

Path-based measures

Rada et al. [109]. This distance metric is one of the most simple approaches and is based on computing the shortest path on an is-a hierarchy between nodes in the Mesh ontology, which is distributed by the National Library of Medicine and used to describe articles from biomedical periodicals. Consequently, the distance is equal to the minimum number of edges that connect two particular nodes.
The formula and the result of calculating the distance between nodes c_1, c_2 are as follows:

$$\begin{aligned} dist_{Rada}(c_1, c_2) &= \min_{path \in \pi_{paths}(c_1,c_2)} len_e(path) \\ &= 2 \end{aligned} \tag{2.3}$$

In this approach, all edges represent uniform distances. The resulting distance also adheres to the similarity: the shorter the distance between two nodes, the more similar they are. However, this approach ignores the depth of the nodes within the hierarchy. In this way, if two specific nodes and two rather generic nodes have the same distance, they also share the same similarity. This contradicts the principle, that specific

c	a certain node within the ontology
ccp	closest common parent, i.e. the most specific common parent of two given nodes
$depth(c)$	the depth of node c within the taxonomy
$ic(c)$	the information content of node c
$len_e(path)$	the length of a *path* as the number of edges
$ls(c_1, c_2)$	the link strength of the edge between nodes c_1 and c_2
max_r	maximum weight assigned to relation r
$milestone(c)$	a weighting factor for the depth of node c within the hierarchy
min_r	minimum weight assigned to relation r
mis	the most informative subsumer (normally equals to the *cpp*
$n_r(c)$	number of relations of a given type leaving node c
p	parent node of node c
$P(c)$	set of features (i.e., properties) of node c
$path$	a certain path in the graph
$pr(c)$	probability that a node c occurs in a corpus
$prop$	a RDF property (similar to a relation r)
$propCard(c, prop)$	the cardinality of a specific property *prop* leaving node c
r	relation of a given type between two adjacent nodes
r'	inverse relation of r
$root$	root node of the hierarchical ontology fragment
TC_r	a static weight factor, which represents the strength (weight) of a particular relation type
w_r	weight of a directed relation between two adjacent nodes
λ	a weighting factor for generalization vs. specialization
π_{nodes}	set of nodes
π_{paths}	set of paths

Table 2.11
Definition of variables used by the listed similarity measures.

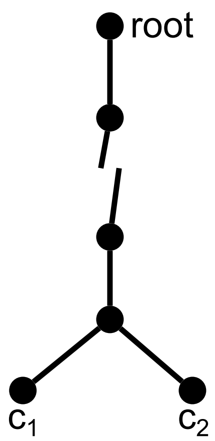


Figure 2.22
Distance measure by Rada et al. [109]. The distance is based on the shortest path between two given nodes by counting the number of is-a edges between them.

nodes are more similar than more generic ones. Simply summing up the number of edges between nodes is therefore insufficient.

Wu & Palmer [154] propose a measure to describe the similarity of verbs which is used to translate English to Chinese verbs. They define the similarity between the nodes c_1 and c_2 based on the closest common parent *ccp* (i.e., the most specific common node that subsumes both nodes). However, opposed to the approach of Rada et al., the relative position of the nodes within the hierarchical structure is regarded.

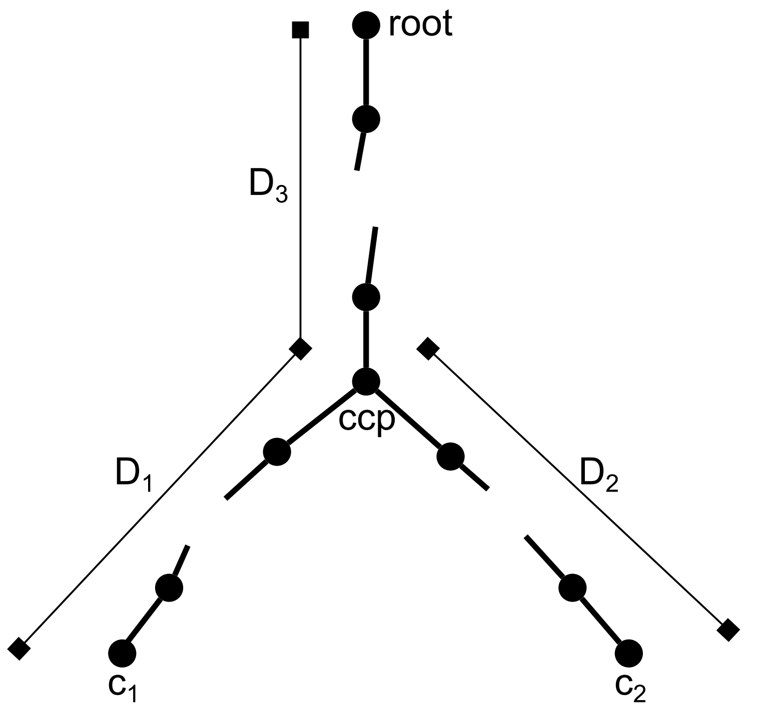


Figure 2.23 *Similarity measure by Wu & Palmer [154]. The similarity is calculated by the path length to the ccp, scaled by its depth within the hierarchy.*

$$sim_{Wu\&Palmer}(c_1, c_2) = \frac{2 \cdot D_3}{D_1 + D_2 + 2 \cdot D_3} \tag{2.4}$$

cpp is the closest common parent of c_1 and c_2. D_1 is the number of edges on the path from c_1 to *cpp*. On the other hand, D_2 is the number of edges on the path from c_2 to *cpp*. D_3 is the number of edges on the path from *cpp* to the root node of the hierarchy [154]. The values range between 1 (similar nodes) and 0.

Zhong et al. [162]. In this measure, the similarity between two nodes c_1 and c_2 is determined based on their distance, which is obtained by calculating their relative positions in the hierarchy. Thereby, the distance measurement reflects following two principles, i) the difference between upper level nodes is larger than that between lower level nodes, and ii) the distance between siblings is larger than that between child and parent nodes. Each node in the hierarchy has assigned a *milestone value*, which is calculated with the formula:

$$milestone(c) = \frac{1/2}{k^{depth(c)}} \tag{2.5}$$

k is a predefined value that expresses the rate at which the value decreases along the hierarchy. *depth(c)* indicates the depth of the node c

in the hierarchy. The root node of the hierarchy is assigned the depth 0.
The distance between two nodes can thus be calculated based on their

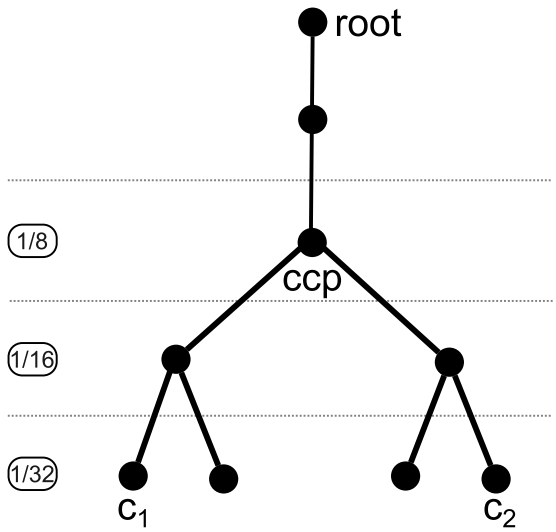


Figure 2.24 *Distance measure by Zhong et al. [162]. The distance is calculated based on the path length to the ccp. The depth of the nodes within the hierarchy is incorporated through assigning a milestone value for each hierarchical level.*

milestone values and the shortest path between them and their closest common parent node *ccp*. For obtaining the milestone values, k was set to value 2:

$$\begin{aligned} dist_{Zhong}(c_1, c_2) &= dist(c_1, ccp) + dist(c_2, ccp) \\ dist_{Zhong}(c, ccp) &= milestone(cpp) - milestone(c) \\ dist_{Zhong}(c_1, c_2) &= (\frac{1}{8} - \frac{1}{32}) + (\frac{1}{8} - \frac{1}{32}) = 0.18 \end{aligned} \tag{2.6}$$

The distance between two nodes has always a value between 0 (nodes are similar) and 1 (nodes are dissimilar). This is ensured by setting the numerator of (Equation 2.5) to $\frac{1}{2}$ so that the distance between the two deepest nodes with the root as their closest common parent will be 1. As the distance values are in the range of [0..1], *similarity* can be defined as

$$sim_{Zhong}(c_1, c_2) = 1 - dist(c_1, c_2) \tag{2.7}$$

Hence, the similarity between the two nodes can be stated as

$$sim_{Zhong}(c_1, c_2) = 1 - 0.1875 = 0.8125$$

This approach was applied by Bizer et al. [5] in the job recruitment domain. Thereby, they compared job descriptions and applicants' profiles based on their semantic similarity.

Sussna [136] proposes a more advanced measure to calculate the similarity between two nodes in the WordNet network. This measure cannot only calculate the distance of nodes based on the shortest path of is-a links between them, but allows to determine the distance based on any type of relation. Moreover, it considers the hierarchical position of the particular nodes within the taxonomy graph. In addition, it takes into account the number of relations of a certain type leaving each of the corresponding nodes. The number of relations reflects the principle of connotation. It is assumed that the more relations are connected to a certain node, the less is the importance of each of these relations. For example, the `St.Stephen's Cathedral` of the city of Vienna has different architectural styles as it contains elements from the ghotic, baroque and renaissance epoch. On the other hand, the `Schönbrunn Palace`, the former imperial summer residence in Vienna, is rather built in baroque style. Therefore, the palace is assumed to be more closely related to baroque than is the `St.Stephen's Cathedral`.

In addition, the different types of relations can have assigned a weight range to reflect their importance for the similarity of nodes. Thereby, for each relation of type r, a weight vs. a weight range $[min_r, max_r]$ is defined by stating minimal and maximal values (e.g., Sussna set the weights for the relation types hypernymy, hyponymy, holonymy and meronymy to [1,2]). The exact weight value depends on the number of relations of type r (n_r) leaving a given child-parent node pair and reflects the connotation between them. Other relation types, where the number of relations is not important, can be assigned a constant weight. For example, Sussna set the weight for the relation type antonymy to the value 2.5.

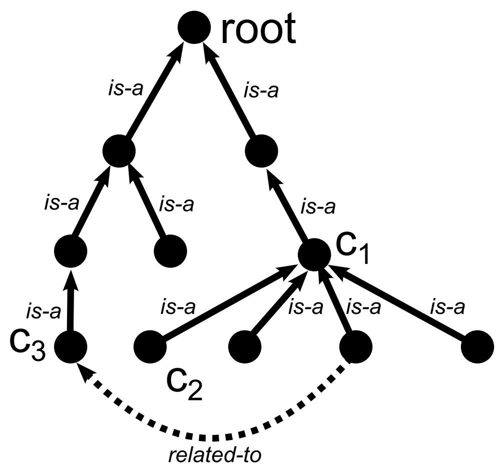


Figure 2.25
Distance measure by Sussna [136]. It is based on the shortest path between nodes. Its strength is the consideration of different types of relations. In addition, the number of relations of each node which belongs the the shortest path is taken into account. Finally, the distance is scaled by the depth of the more specific node.

The weight of a relation r connecting node c_1 with node c_2 is calculated by:

$$w_{Sussna}(c_1 \rightarrow_r c_2) = max_r - \frac{max_r - min_r}{n_r(c_1)} \tag{2.8}$$

For each directed relation of type r, a weight w is calculated by considering the number of relations of type r which leave a particular node. As pointed out, the exact weight is within the given interval $[min_r; max_r]$. If a node has many relations of type r, the value of weight w approaches max_r. In (Equation 2.9) an example of such a weight calculation is given. For this example, the weight range of [1,2] is set for the is-a relation. According to (Figure 2.25), node c_2 has only 1 is-a relation, which links to node c_1, $w(c_2 \rightarrow_r c_1)$ converges to the minimum value of the weight range. On the other hand, node c_1 has 4 incoming is-a relations, one of them being c_2, $w(c_1 \rightarrow_r c_2)$ gets a higher weight range.

$$\begin{aligned} w_{Sussna}(c_2 \rightarrow_r c_1) &= 2 - \frac{2-1}{1} \\ &= 1 \\ w_{Sussna}(c_1 \rightarrow_r c_2) &= 2 - \frac{2-1}{4} \\ &= 1.75 \end{aligned} \tag{2.9}$$

The edge between adjacent nodes c_1 and c_2 has distance or weight

$$w_{Sussna}(c_1, c_2) = \frac{w(c_1 \rightarrow_r c_2) + w(c_2 \rightarrow_{r'} c_1)}{2 \cdot max \left[\min\limits_{path \in \pi_{paths}(c_1, root)} len_e(path); \min\limits_{path \in \pi_{paths}(c_2, root)} len_e(path) \right]} \tag{2.10}$$

The (total) weight w of the edge between these nodes is calculated based on their relation r and its inverse r'. It is the average of the weights of these relations and is scaled by the depth of the edge within the overall semantic graph.

$len_e(path)$ is the length in number of edges of the path and $\pi(c_1, root)$ are all paths from node c_1 to the *root* node of the taxonomic hierarchy. From this set, the path with the smallest number of edges that connects one of the corresponding nodes with the root node is selected. If both nodes have many relations of type r, weight w gets larger. Hence, the numerator in (Equation 2.10) increases, which implies an increase of the total weight (distance) as well. (Figure 2.26) depicts the weights (distances) of all edges within the graph. The weight range of an is-a relation is set to [1,2]. The domain-specific relation *related-to* is set to the constant weight 1.5.

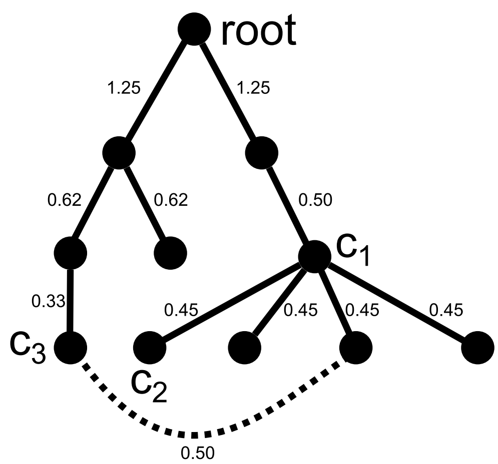


Figure 2.26 *Distance measure by Sussna [136]. The edge values represent the distance between the connected nodes. The lower the distance value, the higher is the similarity.*

Sussna has defined this measure for adjacent nodes. Consequently, to calculate the distance between nodes that are not directly connected, a path between these nodes has to be found. Commonly, the shortest path (i.e., the path where the sum of the distances is minimal) is used. Using the shortest path, the distance between node c_2 and node c_3 is obtained by

$$dist_{Sussna}(c_2, c_3) = 0.45 + 0.45 + 0.50 = 1.40$$

Information-based measures

Resnik [112]. A problem with simple edge-counting measures is that they assume that an edge between two nodes always has the same weight and thus represents the same distance. However, as Resnik points out in [112], parts of a taxonomy (e.g., biological categories) are much denser than others and therefore have varying edge distances.

To counteract this problem, Resnik proposes in [112] an approach to measure semantic similarity between two nodes based on the notion of *information content*, which expresses how specific and informative a node is. The information content ic of a node c is quantified as negative the log likelihood, $-log\,pr(c)$. $pr(c)$ is the probability of encountering an instance of the node c in a specific text corpus and is evaluated with a value in the interval between [0,1].

Resnik used the taxonomy of WordNet for his experiment, whereby the nouns represented the nodes. Node probabilities were computed as the relative frequency of occurrence within the Brown Corpus of American English, which is a collection of different text articles. A node (noun) with a high ic is very specific as the probability of its occurrence is rather low. $pr(c)$ increases monotonically as one moves up in the taxonomy. However, as the node's probability increases, its informativeness or information content decreases. Thus, more general concepts have a lower ic. The root of the taxonomy has the probability 1 and consequently, its ic is evaluated with 0 (as log(1) equals to 0). To put it in another way, if instances of a node appear frequently within a text, the informativeness of this node is rather low. In (Figure 2.27), the amount of informativeness of a node is visualized through the size of the grey circle surrounding the node. More specific nodes have a higher informativeness than more general nodes, which are located nearer to the top of the taxonomy.

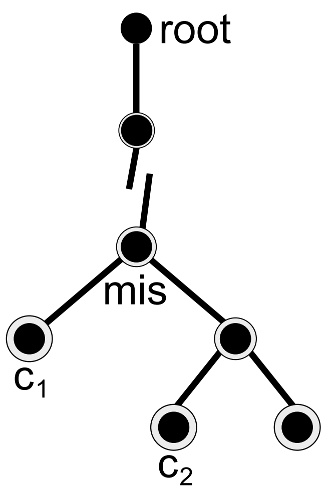


Figure 2.27 *Similarity measure by Resnik [112]. The size of the grey circles represent the information content of the nodes. The more specific the nodes, the higher is their information content.*

Resnik argues that the similarity between two nodes is dependent on the amount of information they have in common. The information shared by the nodes is indicated by the maximal information content of the (lowest) node in the taxonomic hierarchy that subsumes both nodes.

Hence, the similarity between c_1, c_2 is defined by:

$$sim_{Resnik}(c_1, c_2) = \max_{c \in \pi_{nodes}(c_1, c_2)} [\log^{-1} pr(c)] \tag{2.11}$$

$\pi_{nodes}(c_1, c_2)$ is the set of nodes that subsume both c_1 and c_2. The node that is selected based on (Equation 2.11) is termed the *most informative subsumer (mis)*. This is normally the *closest common parent (ccp)*, but other approaches exist as well. As nodes can have various *ccp* nodes, Couto et al. propose in [34] an approach to take the average *ic* of all *ccp* nodes. Resnik's approach does not include the information content of the nodes themselves, whose similarity should be computed, nor the path length to the *mis* node. In this case, all nodes that share the same *mis* are assigned the same similarity value. Hence, some extensions have been introduced that focus on these issues (cf. Jiang & Conrath [77]).

Seco et al. [126]. Measuring the information content of nodes based on their frequency within a text corpus is time consuming. Besides, the content of the text corpus has to correspond to the terms of the ontology. Seco et al. state that *ic* values obtained from very general text corpora maybe different from those derived from rather specific text corpora. To overcome these limitations, they suggest an approach to calculate the intrinsic *ic* of a certain node based on the taxonomic structure of the ontology and not from a text corpus. More specific, the measure to calculate the *ic* of a specific node is based on its number of hyponyms. Nodes, representing leaves in the taxonomy, are assigned an *ic* value of 1, as they do not have any hyponyms. The more hyponyms a node has, the higher is its probability of appearing and thus, its *ic* has a lower value.

Feature-based measures

In contrast to the path-based measures (similarity determined based on the path length) and the information-based measures (similarity determined based on the information content of a node), feature-based measures define the similarity between nodes by looking at their shared and distinctive features. The intuition is that common features tend to increase the perceived similarity of two corresponding nodes whereas distinctive features tend to decrease their similarity.

Tversky [142] proposes a method to measure the similarity of nodes based on the ratio of their common and distinctive features. The

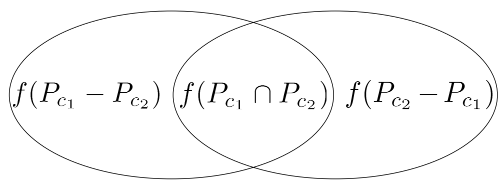


Figure 2.28 *Similarity measure by Tversky [142]. The more properties two nodes have in common, the more similar they are.*

ratio model is given by (Equation 2.12) and expresses similarity between

c_1 and c_2 through a value between 0 and 1:

$$sim_{Tversky}(c_1, c_2) = \frac{f(P_{c_1} \cap P_{c_2})}{f(P_{c_1} \cap P_{c_2}) + \alpha f(P_{c_1} - P_{c_2}) + \beta f(P_{c_2} - P_{c_1})}, \quad \alpha, \beta \geq 0 \tag{2.12}$$

P_{c_1} and P_{c_2} are sets of features (i.e., properties) that belong to the corresponding nodes c_1 and c_2. $f(P_{c_1} \cap P_{c_2})$ represents the set of features shared by both concepts, $f(P_{c_1} - P_{c_2})$ represents the features held by c_1 but not by c_2 and $f(P_{c_2} - P_{c_1})$ represents the features held by c_2 and not by c_1. f measures the prominence of the various features. Commonly, the cardinality of the features is taken. The constants α and β are used to specify the importance of each nodes. If $\alpha = \beta = \frac{1}{2}$, Tversky's measure can be simplified to $2f(P_{c_1} \cap P_{c_2})/(f(P_{c_1}) + f(P_{c_2}))$.

Knappe et al. [**81**] define a similarity measure, which is not only based on taxonomic relations but also considers other, property-based relations between nodes. They agree that approaches calculating the shortest path based on the number of is-a relations between nodes (e.g., Wu & Palmer [154] or Zhong et al. [162]) are straightforward and do not exhibit high computational costs. However, not only the shortest path between nodes may contribute to similarity as multiple paths between nodes may exist.

For example, the degree of similarity between node c_1 and node c_2 should be higher if they both have a property to node c_3 in addition. Hence, node c_1 and node c_2 are connected over two paths. This example shows that a measure that considers multiple paths between nodes, returns a better similarity value, but also raises issues of computational issues as calculating all possible paths between nodes is a complex task.

To solve the complexity problem, Knappe et al. exploit a subset of all possible paths for measuring similarity, which is based on the notion of *shared nodes*. Shared nodes are all nodes, that are (upwards) reachable from the corresponding nodes and reflect the set of nodes $\pi(c_1) \cap \pi(c_2)$, they have in common. In (Figure 2.29), an example ontology is given, in which all shared nodes between nodes c_1 and c_2 are visualized through non-colored circles.

Knappe et al. define the similarity between these nodes as:

$$\begin{aligned} sim_{Knappe}(c_1, c_2) = \lambda \quad &\cdot \quad \frac{|\pi_{nodes}(c_1) \cap \pi_{nodes}(c_2)|}{|\pi_{nodes}(c_1)|} + \\ (1 - \lambda) \quad &\cdot \quad \frac{|\pi_{nodes}(c_1) \cap \pi_{nodes}(c_2)|}{|\pi_{nodes}(c_2)|} \end{aligned} \tag{2.13}$$

λ is a weighting factor with range [0,1] that defines the relative importance of generalization vs. specialization. The similarity function outputs a value which scores between 1 (for similar nodes) and 0 (for dissimilar nodes). Setting λ to $\frac{4}{5}$, the similarity of nodes c_1 and c_2 according to the ontological fragment shown in (Figure 2.29) is

$$\begin{aligned} sim_{Knappe}(c_1, c_2) &= \frac{4}{5} \cdot \frac{4}{5} + (1 - \frac{4}{5}) \cdot \frac{4}{6} \\ &= 0.77 \end{aligned}$$

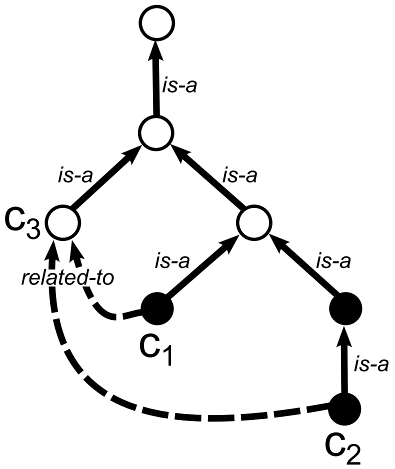


Figure 2.29
Similarity measure by Knappe et al. [81]. This measure is based on the set of shared nodes between two given nodes. This set is represented in this Figure through non-colored circles.

Hybrid measures

Jiang & Conrath [77] present a mixed approach that combines a path-based with an information-based method. They use the approach proposed by Resnik to measure the similarity between nodes based on the information content of their *mis*. However, this approach does not consider how distant the nodes are from their common *mis*. Therefore, Jiang & Conrath extended the approach of Resnik by exploiting not only the information content of the *mis*, but also take into account the distance between the nodes and the *mis*.

To measure the distance, weights are assigned to the edges along the shortest path to the *mis* of the two corresponding nodes. The weight of an edge is defined as the link strength ls, which is computed by taking the difference of the information content ic values between a child node c and its parent node p:

$$ls(c, p) = |ic(c) - ic(p)| \tag{2.14}$$

Next, the semantic distance between nodes c_1, c_2 can be computed by adding all the edge weights along the shortest path. It is therefore defined as:

$$\begin{aligned} dist_{Jiang\&Conrath}(c_1, c_2) &= \sum_{c \in \{path(c_1,c_2) - mis(c_1,c_2)\}} ls(c, p)) \\ &= ic(c_1) + ic(c_2) - 2 \cdot ic(ccp(c_1, c_2)) \end{aligned} \tag{2.15}$$

Lin [85] proposes a similar measure that takes into account the ic of the individual nodes and the ic of the *mis*. Instead of calculating the difference, he uses a ratio in the formula.

Mazuel et al. [90] propose a similarity measure, which can be seen as a generalization of the approach by Jiang & Conrath [77]. On the other hand, it is closely related to the approach presented by Knappe et al. [81]. It is related to the latter as it emphasizes that two nodes can be connected to each other over various paths following different kind of relations, such as is-a, part-of or domain-specific ones. However, they argue that not every path is semantically correct and thus valid. To identify semantically correct paths, they refer to rules defined by Hirst & St-Onge [69].

Hirst & St-Onge [69] distinguish three kinds of relations, namely upward relations [U] (corresponding to generalization), downward relations [D] (corresponding to specialization) and horizontal relations [H] (representing domain-specific relations and carrying a specific meaning). Based on these relations types, they present three rules to describe semantically correct paths, namely a) no other direction may precede an upward relation, b) at most one change of direction is allowed, and c) it is permitted to use a horizontal relation to make a transition from an upward to a downward direction. Finally, they outline 8 patterns to build correct paths, which correspond to these three rules. The patterns are U, U-D, U-H, U-H-D, D, D-H, H-D, H.

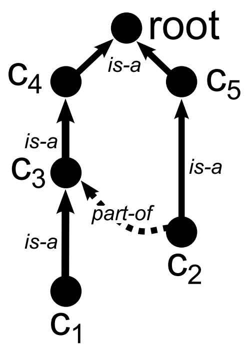


Figure 2.30 *Distance measure by Mazuel et al. [90]. It is based on the shortest, semantically correct path between two nodes. The path might consist of different types of relations.*

According to (Figure 2.30), there are two correct paths between nodes c_1 and c_2:

$$\begin{aligned} path_1(c_1, c_2) &= \{c_1, is{-}a, c_3, is{-}a, c_4, is{-}a, root, contains, c_5, contains, c_2\} \\ path_2(c_1, c_2) &= \{c_1, is{-}a, c_3, has{-}part, c_2\} \end{aligned}$$

The relation *contains* is the inverse of the relation $is{-}a$ and the relation $has{-}part$ the inverse of the relation $part{-}of$. Next, the weight (i.e., cost) of the semantically correct paths has to be calculated. For a path that only consists of is-a relations (e.g., $path_1$), the information-based measure proposed by Jiang & Conrath is used. If a path consists of mixed relations (e.g., $path_2(c_1, c_2)$ with $\{c_1, is{-}a, c_3, has{-}part, c_2\}$), it is splitted into single-relation sub-paths. Hence,

$$path_2(c_1, c_2) = path_{2_1}\{c_1, is{-}a, c_3\} \oplus path_{2_2}\{c_3, has{-}part, c_2\}$$

The weight of a mixed-relation path is calculated by summing up the weights of its single-relation paths. As the information-content based measure computes the weights regarding to the hierarchy structure of the ontology, it cannot be used to compute the weight of non-hierarchical relation types such as $part{-}of$ ones. Therefore, Mazuel et al. suggest a new approach to derive a weight for such kind of relations. The weight of a path $w_{path_r}(c_1, c_2)$ when r is a non-hierarchical relation, is defined as

$$w_{path_r}(c_1, c_2) = TC_r \cdot \frac{|path_r(c_1, c_2)|}{|path_r(c_1, c_2)| + 1} \tag{2.16}$$

Mazuel & Sabouret use a $\frac{n}{n+1}$ function to simulate the log function of the information-content based approach. (According to this approach, the most specific nodes in an ontology are considered high informative and have assigned a value near to 1. If going up in the taxonomy, the values of the parent nodes decrease with the log progression of their informativeness.) Based on the $\frac{n}{n+1}$ function, the weight of a path increases with the length of the path (the increase follows the log progression). It is bounded by the factor TC_r (i.e., the weight of the path $path_r$ can never score higher than TC_r).

TC_r is a static weight factor, which represents the strength (i.e., weight) of a given relation type. A low value of TC_r represents a high strength of this relation type, which means that it is very informativeness. This results in a low weight (i.e., cost) of $path_r$ (cf. Equation 2.16). The higher the value of TC_r is, the lower is the informativeness of this relation type and the higher is the weight of the path.

If we assume the information content ic of the various nodes depicted in (Figure 2.30) as $ic(c_1)$=1, $ic(c_2)$=1, $ic(c_3)$=0.68, $ic(root)$=0.08 and $TC_{part-of}$=0.4, we can calculate the weight w of the paths $path_1$ and $path_2$ as follows:

$$w_{path_1}(c_1, c_2) = 1 + 1 - 2 \cdot 0.08 = 1.84 \qquad \text{(cf. Equation 2.15)}$$
$$w_{path_{2_1}}(c_1, c_3) = 1 - 0.68 = 0.32 \qquad \text{(cf. Equation 2.14)}$$
$$w_{path_{2_2}}(c_3, c_2) = 0.4 \cdot \frac{1}{1+1} = 0.2 \qquad \text{(cf. Equation 2.16)}$$
$$w_{path_2}(c_1, c_2) = w_{path_{2_1}}(c_1, c_3) + w_{path_{2_2}}(c_3, c_2)$$
$$w_{path_2}(c_1, c_2) = 0.32 + 0.2 = 0.52$$

In the last step, the distance between nodes c_1 and c_2 is defined as follows:

$$dist_{Mazuel\&Sabouret} = \min_{path \in \pi_{paths}(c_1, c_2) | HSO(path) = true} W(path) \qquad (2.17)$$

$\pi_{paths}(c_1, c_2)$ is the set of acyclic paths $path$ between nodes c_1 and c_2. The function $HSO : \pi_{paths}(c_1, c_2) \longrightarrow \mathbb{B}$ is defined such that $HSO(path)$ is $true$ iff $path$ is a valid path according to the Hirst & St-Onge rules. As defined by (Equation 2.17), the distance between c_1 and c_2 equals the minimum path weight. According to our example, $path_2$ has the minimum weight value and thus, the distance $dist(c_1, c_2) = 0.52$.

Lessons learned

(Table 2.12) gives an overview of the advantages and disadvantages of the various approaches that have been presented in the last sections.

Path-based measures.

Path-based measures calculate the shortest path between nodes. In the simplest approach (cf. Rada [109]), the edges on the path represent uniform distances. More advanced versions scale the distance value by the depth of the closest common parent (cf. Wu & Palmer [154]) or by a factor which calculates the depth for each level within the ontological hierarchy (cf. Zhong et al. [162]). The approach presented by Sussna [136] is more complex as it is sensitive to following four parameters: a) not only

Table 2.12 *Pros and Cons of the presented similarity measures.*

Measure		
Rada et al. [109]	Complexity	semantic distance
	Pros	simplicity
	Cons	all edges represent uniform weights, requires a consistent taxonomy
Wu & Palmer [154]	Complexity	semantic similarity
	Pros	simplicity, depth of ccp
	Cons	taxonomic relations only, requires a consistent taxonomy
Zhong et al. [162]	Complexity	semantic similarity
	Pros	simplicity, weighting factor for depth of nodes
	Cons	taxonomic relations only, requires a consistent taxonomy
Sussna [136]	Complexity	semantic relatedness
	Pros	considers different types and number of relations as well as depth of nodes
	Cons	requires parameters to be settled
Resnik [112]	Complexity	semantic similarity
	Pros	probability of word occurence in text corpora to enhance similarity
	Cons	WordNet specific, only mis, computational effort to calculate ic from corpora
Seco et al. [126]	Complexity	semantic similarity
	Pros	ic based on hierarchical ontology
	Cons	requires consistent ontology, taxonomic relations only
Tversky [142]	Complexity	semantic similarity
	Pros	considers properties
	Cons	does not exploit taxonomic relations and information about the object a property is linked to
Knappe et al. [81]	Complexity	semantic similarity
	Pros	considers properties
	Cons	relation types cannot be weighted
Jiang & Conrath [77]	Complexity	semantic similarity
	Pros	exploits path between nodes and common mis
	Cons	taxonomic relations only, WordNet specific
Lin [85]	Complexity	semantic similarity
	Pros	exploits path between nodes and common mis
	Cons	taxonomic relations only, WordNet specific
Mazuel et al. [90]	Complexity	semantic relatedness
	Pros	considers mixed-relation paths
	Cons	similarity is calculated based on the shortest path only

taxonomic relations (is-a) are exploited to determine the shortest path but also domain-specific ones, b) these relation types can be weighted, c) the number of relations of the nodes on the shortest path are taken into account, and d) the depth of the nodes is considered as well. All these parameters account for the strength of a link between adjacent nodes in an ontology and thus, have a significant effect on the length of the shortest path between particular nodes within the ontology. As the approach by Sussna utilizes different parameters to define the strength of a link, it can be classified as being a *semantic relatedness measure*.

Information-based measures.

As Resnik points out in [112], path-based measures require a consistent and well-organized ontology, as the semantic strength of a link is just computed by exploiting ontological information. He argues, that approaches that exploit information content from other sources (e.g., text corpora) are less sensitive to issues related to variable semantic distances or different node densities within the ontology [8]. Resnik only considers the information content of the most informative subsumer to calculate the similarity. Hence, all nodes that share the same most informative subsumer therefore have identical similarity values, although there might be some fine-grained distinctions between them. Therefore, Jiang & Conrath [77] and Lin [85] also consider how distant the individual nodes are from their most informative subsumer by incorporating the information content of the individual nodes, too. However, the information content value heavily depends on the type of text corpora used as well as its size. The Brown corpus, which is usually used together with the taxonomy of WordNet, is a general knowledge base and ideal to calculate similarities between words. In this way, its applicability for domain ontologies, which utilize a specific vocabulary, is restricted as WordNet does not cover very specific definitions. Moreover, information-based measures require to perform a rather time-consuming analysis of the text corpora to calculate the information content values. Of course, once the computation is finished, it does not need to be recomputed unless a new node (word) is added to the ontology. In addition, they just consider taxonomic relations to compute the similarity values. If domain-specific relations should be considered as well, a hybrid approach has to be chosen.

Property-based measures.

Property-based measures focus on the set of shared properties between nodes. In this way, the commonalities of two nodes can be estimated by the extent to which they share properties. Such a measure is proposed by Tversky [142]. However, one cannot assign specific weights to different domain-specific relations. In this way, all properties have the same weight. In addition, it does not account for properties a node may inherit from its parent node. A further, major disadvantage is that only the properties leaving a source node are taken into account to compute the similarity, but not the target nodes to which the properties are connected. To give an example, for calculating the similarity, it makes no difference whether the property `hasArchitecturalStyle` links to `BaroqueArchitecture` or `GhoticArchitecture`, nor the property `hasTopic` links to `ImperialVienna` vs. `GreenVienna`. Knappe et al. [81] suggests a similar approach, which is based on the notion of *shared nodes*. Its crucial drawback is that all relations, i.e., is-a relations and domain-specific ones, represent the same weights.

Hybrid approaches.

Hybrid approaches incorporate the advantages of different approaches, such as path-based and information-based measures. The approach of Mazuel et al. is worth-mentioning as it considers mixed-relation paths, i.e., paths that might include is-a and domain-specific relations. Thereby, different weights can be assigned to the varying types of relations. Basically, this measure calculates the whole set of semantically correct paths between two given nodes. In the next step, the shortest path, which is semantically correct is selected to define the similarity between two nodes. As the weight of taxonomic relations is computed based on the information-based approach by Jiang & Conrath [77], it exhibits the same disadvantages.

3 First matchmaking process

As outlined in (Section 1.5) we propose an *iterative matchmaking process* that consists of two sub processes. To provide personalized recommendations for tourists, it is crucial to model both the generic preferences of tourists and the characteristics of the related tourism objects: on the one hand, the tourist has to be characterized sufficiently in his/her preferences, and on the other hand, the supplier side, all tourism objects have to be quantitatively characterized, reflecting similarities and differences between them.

(Figure 3.2) details the *first matchmaking process*. We leverage the concept of *tourist types* (e.g., sightseer, adventurer) based on existing typologies in the scientific tourism literature (see Section 2.3) and use them in form of a *stereotypical approach* to generate a high-level user profile.

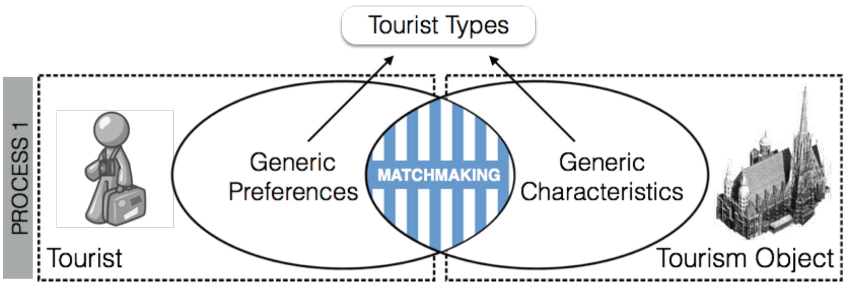


Figure 3.1
The first matchmaking process based on tourist types.

Such profiles can be represented in a *vector-space model*. Given that the tourism objects are represented in this vector space as well, *vector-based matchmaking* can be applied to relate user profile vectors against vectors of tourism objects.

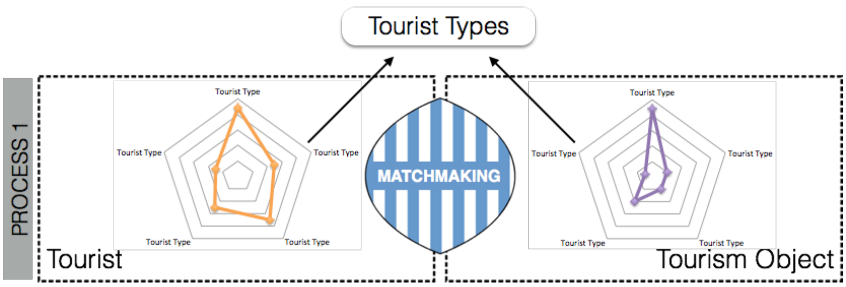


Figure 3.2
Representing generic profiles in a vector-space model.

As depicted in (Figure 3.2), a particular tourist and the exemplary tourism object 'St. Stephen Cathedral' are both represented through vectors that describe their generic preferences vs. characteristics. A vector based matchmaking method can then be applied to propose a list of top-N recommendations.

To achieve this goal, following research issues have to be tackled:

- a selection of an appropriate tourist typology, which is used as basis for the stereotypical user modeling approach
- a representation of tourist profiles in a vector-space model,
- a method to construct such tourist profiles in a convenient way for tourists as it is difficult to specify such vectors directly,
- a representation of tourism objects in this vector space, and
- a vector-based similarity metric to match tourist profiles against the profiles of tourism objects

Each of these research issues will be discussed in the following subsections.

3.1 Selection of an appropriate tourist typology

We use the approach by Werthner et al. [96] as basis to derive our set of tourist types and to generate a high-level tourist profile that predicts tourist preferences according to the set of activities offered at a specific destination. This work addresses the question whether a lower number of factors is sufficient to compose a travel profile. Thereby, it uses the 15 tourist roles proposed by Yiannakis & Gibson [54], which people are likely to be engaged in during vacation. In addition, personality traits in the form of the "Big Five", i.e. *extraversion*, *agreeableness*, *conscientiousness*, *neuroticism*, and *openness to experience* are taken into account to relate the personality types to travel behavioral patterns. Personality characteristics stay rather stable over time and thus constitute the *long-term* user profile. Based on a factor analysis, the tourist roles and the personality traits could be reduced to a set of 7 factors, comprising *sun loving and connected*, *educational*, *independent*, *culture loving*, *open minded and sportive*, *risk seeking* as well as *nature and silence loving*.

In our case, we refer to the set of the 7 factors as follows: *Sun Worshipper*, *Educational Buff*, *Sight Seeker*, *Cultural Visitor*, *Avid Athlete*, *Action Seeker*, *Nature Lover* (see Figure 3.4).

Along with this, we would like to emphasize the work by Moscardo et al. [93] that people evaluate destinations during their decision-making process with respect to the set of activities they can perform there. In this way, activities are very important for tourists to satisfy their needs during destination stay. Gretzel confirmed in her study [57] that tourist types (e.g., sight seeker) are a valid means to predict the set of activities in which tourists like to engage during their vacation.

The linkage between a particular tourist and the tourist types he/she engages in is not fixed, which means that in each decision-making vs. recommendation process he/she can refine the linkage and choose those tourist types that best reflect his/her travel needs and prevalent travel preferences.

3.2 Representation of tourist type profiles

Vectors are suited to model tourist types whereby each dimension corresponds to a certain tourist type while the value indicates how much a particular tourist identifies him- or herself with the corresponding type. To put it in other words, the value quantifies the degree of match between this category and the tourist.

Typically, however, individual tourists cannot be characterized by only one of these archetypes but have *mixed profiles* as they show attributes of several types, although to varying degrees. Thus, tourist types model the tourists' generic interests in an abstract form. (Figure 3.3)

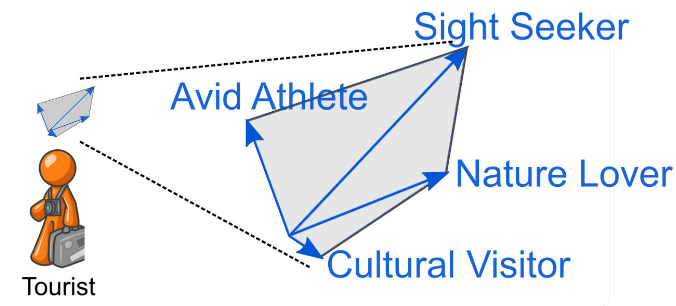


Figure 3.3 *Modeling the tourist profile through a vector.*

depicts an exemplary tourist, who likes to enact in the role of an sight seeker, followed by nature lover and avid athlete, and rather dislikes cultural activities.

3.3 Construction of an initial tourist profile

Putting personal data into the system is a cumbersome task for tourists and it is thus desirable to minimize user input for creating a user profile. As it is difficult to specify a tourist profile directly, a *stereotype approach* addresses this issue by using the *tourist factors* classification in order to facilitate the process of tourist profile creation. Gretzel et al. [57] confirmed that tourist types can be used as a shortcut in order to classify tourists into groups and use the predetermined preferences that characterize the individual tourist types to select appropriate tourism objects. Furthermore, they found out that tourists were able to select best-fitting tourist types and could identify those that did not match their preferences.

In this way, travel types can be used as a substitute for lengthy user questionnaires. One way to use them within a stereotype approach is to let users choose among a set of pre-defined travel types which are presented to them through a Web interface (cf. Figure 3.4). Users can use the rating bars to quantify how much they identify themselves with the several tourist factors. A mapping has to be defined between the ratings (number of filled in heart-symbols) and the numerical values of their vector representation. (Figure 3.5) shows such a mapping whereby the numerical values range between 0 and 100.

Another method is to derive the tourist profile from tourism-related photographs, by letting the tourists select from a couple of photos that reflect their tourism habits, and then infer their according tourist factors. This way, it is possible to make the traditional process of profile

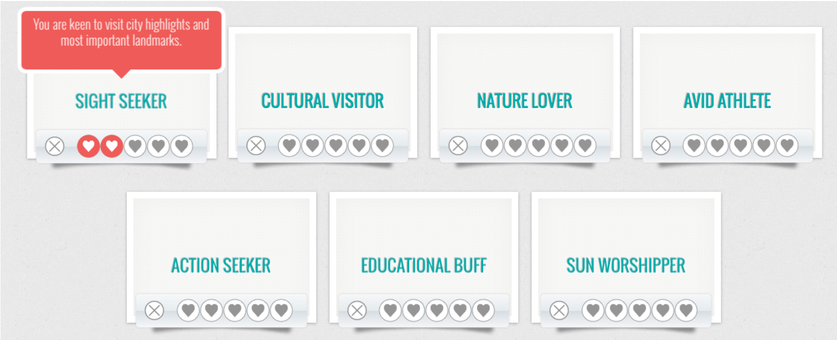


Figure 3.4
A Web interface allowing users to choose among tourist factors to initialize their profile.

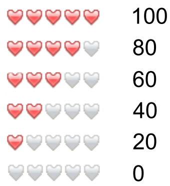


Figure 3.5
Mapping between rating symbols and numerical values.

generation more fun [95]. This approach has been implemented in a Web application called "PixMeAway" by the company PIXTRI OG. (Figure 3.6) depicts the picture selection process.

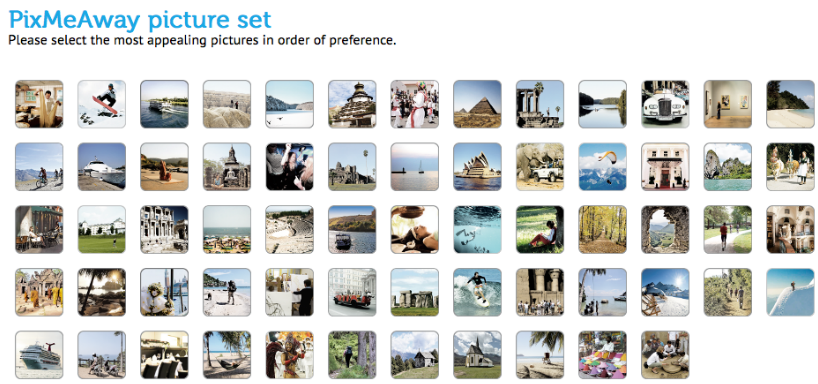


Figure 3.6
Online picture selection (www.pixmeaway.com).

3.4 Representation of tourism objects' profiles

The goal of the matchmaking process is finding, for a given tourist profile, those sights that are most attractive to a particular tourist. Therefore, it is crucial to model how the different *tourist types* can be linked to appropriate *tourism objects*. As destinations usually offer a large amount of tourism objects to their visitors, we propose a *semi-automatic, two-step process* to link the given tourist types to appropriate tourism objects.

First step: domain experts define scores for tourism objects manually.

In a first step, domain experts mark manually for each of the *prototypical tourist factors* (e.g., Action Seeker or Cultural Visitor) a small sample of *typical tourism objects* that are closely related to these types. The degree of relationship can be specified with different weightings. That is done individually for each of the tourist types. As depicted in

(Figure 3.7), the city destination *Vienna* offers a set of tourism objects that fit the interests of the *Action Seeker*, including *Bungee Jumping* at the Danube Tower, visiting the *Third Man Tour* and riding the *Roller Coaster* at the Prater. Another set of tourism objects such as the *Spanish Riding School* or the *Albertina Museum* are related to culture and thus fit the *Cultural Visitor* type.

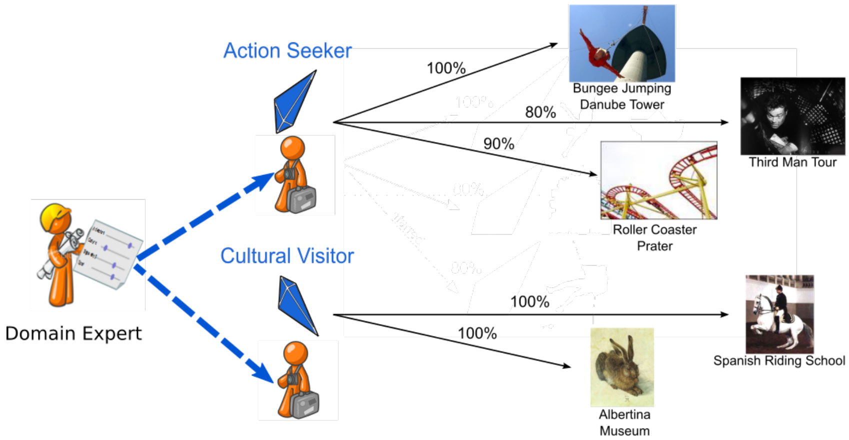


Figure 3.7 *Linking prototypical tourist factors to typical tourism objects of the destination Vienna.*

After this step, certain tourism objects (e.g., the Spanish Riding School) are linked to their most corresponding tourist factor (e.g., the Cultural Visitor). The weight of the linkage is specified through a numerical score. The higher the score, the closer is the linkage between the tourism object and the respective prototypical tourist type. In our case, the score attached to a tourism object scales between the values 100 (maximum value) and 0 (minimum value). The scores defined for each tourism object can be visualized in a matrix. (Table 3.1) shows the scores of the `Spanish Riding School`. In this case, there is only 1 entry inserted by the domain expert. Moreover, some tourism objects

Table 3.1 *A matrix depicting the weighting scores for the Spanish Riding School that have been inserted by the domain expert.*

Tourist Type	**Spanish Riding School**
Sight Seeker	?
Cultural Visitor	100
Nature Lover	?
Avid Athlete	?
Action Seeker	?
Educational Buff	?
Sun Worshipper	?

(e.g., the Clock Museum) might not have even been linked to any of the tourist types by the domain expert (cf. Table 3.2). As these examples show, some weighting scores are missing. For example, we do not know whether the `Spanish Riding School` is also relevant for other tourist types such as the *explorer* tourist.

Hence, we need a procedure that automatically inserts all missing scores of the tourism objects. Another requirement is that this should

Tourist Type	Clock Museum
Sight Seeker	?
Cultural Visitor	?
Nature Lover	?
Avid Athlete	?
Action Seeker	?
Educational Buff	?
Sun Worshipper	?

Table 3.2 *A matrix depicting the weighting scores for the Clock Museum. As the domain expert has not inserted any values, it is yet unknown which tourist types might be interested in this attraction.*

be done in an automatic way in order to disburden the domain expert.

A promising solution is based on the idea to consider the similarity between tourism objects. If two objects are similar, we can also argue that they are of same relevance for a specific tourist type. Once we have computed the degree of similarity, we can propagate the scores which have been determined by the domain expert to all related tourism objects. For example, if we know that the `Albertina Museum` is highly relevant for cultural tourists and given the RDF statements `Albertina_-Museum rdf:type Museum` and `Museum_Modern_Art rdf:type Museum` it can be predicted that the `Museum of Modern Art` might also be relevant for cultural tourists as it is rather similar to the former one. One reason that we can claim that both museums are similar is the fact that both belong to the same concept (`Museum`). Of course, the *amount* of score that is propagated to 'similar' objects should reflect the *degree* of similarity. This means that with decreasing similarity value, the score propagated should decrease as well.

Second step: a similarity metric is used to predict the missing values in an automatic way.

Thus, in a second step, we need to come up with a similarity metric for tourism objects. *Semantic similarity measures* quantify similarities based on the proximity between the nodes in an *ontology*. If the tourism objects are semantically annotated according to a tourism ontology and organized in form of a semantic network, we can exploit this kind of ontological knowledge. (Figure 3.8) illustrates this approach and shows how ontological concepts are used to describe the tourism objects. In

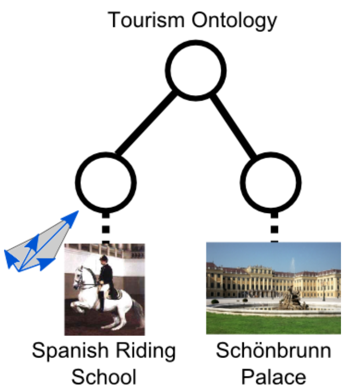
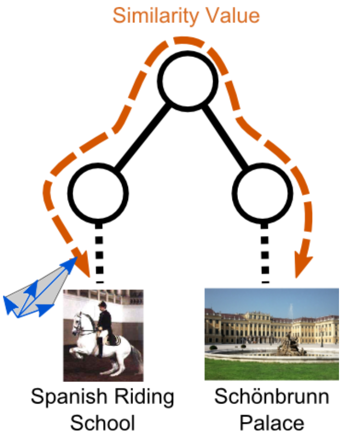
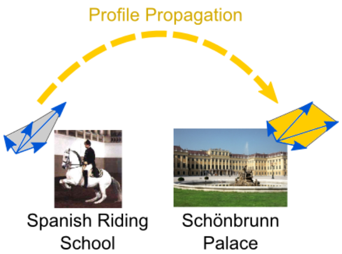


Figure 3.8 *Tourism objects can be described semantically based on a tourism ontology. The circles represent concepts within the ontology.*



(a) Defining a similarity measure between tourism objects.



(b) Profile progation based on the semantic similarity between tourism objects.

Figure 3.9 *Leveraging semantic similarity between tourism objects to propagate weighting scores.*

this Figure, the `Spanish Riding School` and the `Schönbrunn Palace` are shown, which are annotated through ontological concepts and act as elements in the semantic network. The 'leave' symbol next to the *Spanish Riding School* denotes that this instance has already been linked by the domain expert to some tourist types.

Once we have *annotated all tourism objects with semantic concepts* and *defined a semantic similarity metric* (cf. Figure 3.9a), we can use it to propagate the weightings (cf. Figure 3.9b) throughout the semantic network of tourism objects. After this step, for every tourism object a score is obtained that expresses the correlation with each of the given tourist type in a quantitative way.

In the following, we first present our tourism ontology that we have used to annotate the tourism objects with ontological concepts. We then show which semantic similarity measure (for an overview of different measures see Section 2.4.4) fulfills our requirements and thus is appropriate to be used as basis to calculate the similarity between tourism objects. Finally, we show how we adapt this measure to fit our framework and we visualize the process that is used to automatically fill the missing scores of the tourism objects. The tourism ontology is also used in the second matchmaking process (see Chapter 4) to generate specific profiles as overlays of the ontology.

3.4.1 Developing an e-tourism ontology

Due to the heterogeneity of the tourism sector, the process of developing and maintaining a *single tourism ontology* that covers the whole tourism market, including geographical-, temporal-, and user-related information would be very tedious [28] and would require an agreement of the shared vocabulary between the different tourism organizations. Hence, in order to cover the semantic space of the tourism domain and to facilitate interoperability between the different tourism services, a bundle of ontologies may be required. However, these ontologies are not disconnected, but should be integrated around a *core domain ontology*, as proposed by the methodology in Stuckenschmidt et al. [135]. In detail, the core ontology should contain the common vocabulary of the tourism sector and can be *extended by other ontologies* in a modular way.

Ontologies can be described according to their level of genericity (cf. Figure 3.10), using the classification model developed by Guarino [63].

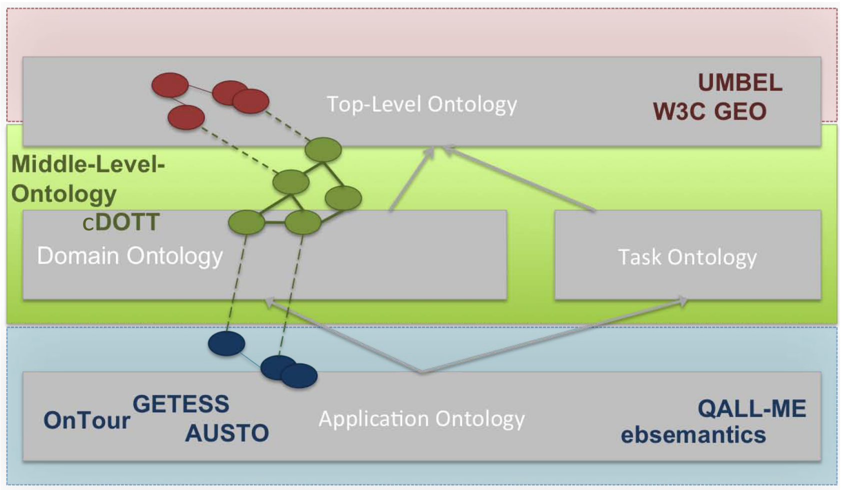


Figure 3.10 *Classification of ontologies.*

Top-Level (upper-level) ontologies describe very general concepts like space, time, matter or object. Ontologies on this level are domain- or problem independent. Several top-level ontologies are available, such as UMBEL, SUMO, OpenCyc, W3C Geo or W3C Time. A *domain ontology* describes the vocabulary related to a generic domain. Thereby, the concepts introduced in the top-level ontologies are further specialized. *Task ontologies* describe a generic task or activity, such as the process of booking a package tour, including flight and rental car. *Application ontologies* are a specialization of both, domain and task ontologies. The concepts of an application ontology correspond to roles played by domain objects while performing a certain activity. An example would be an ontology to assist tourists in planning a complete travel solution.

We have developed a core domain ontology called cDOTT (core Domain Ontology of Travel & Tourism) [16]. cDOTT was developed based on the methods proposed by Uschold & King [55] and Noy & McGuinness [98]. As depicted in (Figure 3.10), cDOTT is a domain ontology that contains tourism concepts on a higher-abstract level which can be further specialized by application ontologies. The main concepts to be defined are *tourism objects* (i.e., attractor, food&service, accommodation, transportation and infrastructure), *tourism events* (e.g., music festival) as well as *tourism destinations* (e.g., national park or lake region)[110]. In addition, it is related to concepts of upper-level ontologies, thus achieving the linkage of lower-level and upper-level ontologies. cDOTT is aligned to concepts of the upper-level ontology UMBEL, a lightweight, subject concept reference structure for the Web. In addition, concepts of other upper-level ontologies are integrated, including the W3C Geo & Time-, the FOAF -, a currency- and a country ontology. cDOTT is based on the published Harmonise ontology [49] as well as on the EON Traveling ontology [43], while including relevant terms from additional ontologies. The ontologies were investigated with respect to their level of maturity, because each of them has been developed under a slightly different purpose. According to this fact, the ontologies show only a small overlap of concepts and one has to cope with different naming and distinctive granularity of concepts. In order to counteract these obstacles, cDOTT is defined on a more abstract level, so that existing, more fine-grained tourism ontologies such as the QALL-ME [94] or the DERI OnTour ontology [39] can be integrated.

The integration of other ontologies enables concept expansion and reasoning. The benefit of this methodology is the possibility to leverage existing concepts and enrich tourism objects with semantic annotations using controlled vocabularies, thereby assigning new semantic roles to tourism objects.

Towards the integration of concepts related to tourism objects within cDOTT

As our main goal is to match tourist profiles with tourism objects it is necessary to add specific concepts that are necessary in order to semantically describe tourism attractions. This is done by extending cDOTT through a more fine-granular vocabulary that focuses on the domain of tourism attractions.

Tourism attractions are a very specific field within the tourism industry. A vast amount of attractions exists worldwide that attracts millions of tourists each year. The side-effect of the variety and quantity of attractions is nearly the same amount of different classification schemes that are used to categorize and describe these attractions. This not only confuses tourists searching for attractions, but also has a negative effect on destination management organizations which are forced to use different categorizations. Different classification schemes are used not only on a country level, but also on a regional level and it also happens that local tourist boards belonging to a certain tourism region use different categories to describe their attractions. Hence, it is not suprising that there is no standard yet on European level to classify tourism attractions. Due to the diversity and complexity of attractions, tourism research has come up with a large volume of definitions of tourism attractions.

According to Walsh-Heron & Stevens [78], a visitor attraction is a *feature in an area that is a place, venue or focus of activities*, which, among other things,

- sets out to attract visitors
- provides a fun and pleasurable experience
- provides an appropriate level of facilities and services to meet and cater to the demands, needs and interests of its visitors.

Based on this definition, tourists visit such attractions where they expect to derive high pleasure. These either can be either natural attractions, such as the 'Grand Canyon in the USA', buildings created by humans such as the 'Colosseum in Rome', venues such as the 'Football World Cup' or locations that offer specific kinds of activities such as sports-related activities (e.g., rafting) or cultural-related activities (e.g., watching an opera). Besides, attractions might not realize their potential themselves, but require infrastructural and other kinds of services to satisfy their visitors.

Modeling tourism objects in cDOTT

This section details the vocabulary that we used to describe tourism objects and explains its integration in the overall cDOTT ontology. As the cDOTT ontology already contains the concept *attractor* for describing tourism attractions, this concept served as starting point to derive a more comprehensive set of concepts which is needed to describe different types of attractions. To define this set of concepts, a top-down and bottom-up approach has been followed. The task of the top-down approach is to build a conceptual layer of attraction types based on existing taxonomies and classification schemes. Within the bottom-up approach, descriptions of tourism attractions, taken from Websites of tourism organizations and travel guide books, have been studied to identify missing features.

In order to refine the *attractor* class and form a hierarchical structure of attraction types, we exploited the taxonomy developed within the PICTURE project [143]. In (Figure 3.11), the sub-concepts of the class *attractor* are depicted. On a coarse-grained level, *attractors* can be classified in *objects* (such as monuments and buildings), *places* (such as parks, gardens or urban areas) and events (such as concerts or festivals). The concept *city highlight* is used in order to mark certain attractors explicitly as 'worth seeing' and to bring them into prominence. Each of

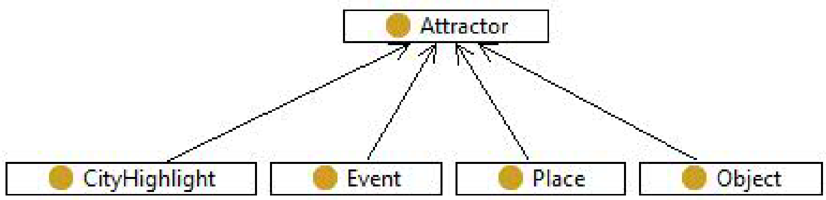


Figure 3.11 *Classification of attractors within cDOTT.*

these concepts is further splitted into several sub-concepts. The resulting hierarchical structure reaches a depth up to six levels and contains about 120 classes.

Based on information from existing tourism Websites and travel guide books, we found out that quite often, tourism attractions are grouped according to destination-specific *themes*. For example, the Viennese tourism board bundles tourism attractions in certain topic related to *the Imperial Vienna, Green Vienna, Vienna Design* or *Art Nouveau*. In order to relate attractors with destination-specific topics, we included the class *topic* into the cDOTT ontology as well. (Figure 3.12) depicts the semantic description of the tourism attraction `Schönbrunn Palace` located in Vienna. It is classified as *palace* which belongs to the class *historical architecture*. Furthermore, it belongs to the theme *imperial vienna* as it is closely related to the House of Habsburg. Moreover, it is built by the architects `Nicolo Pacassi` and `Johann Bernhard Fischer von Erlach`. The FOAF ontology was imported in order to represent

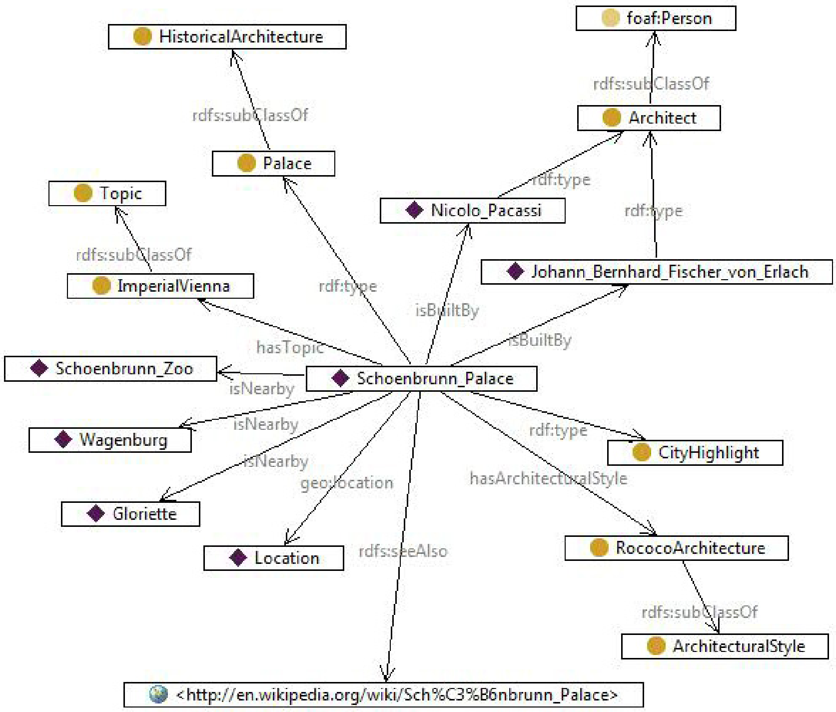


Figure 3.12 *Fragment depicting the semantic description of the Viennese attraction Schönbrunn Palace in cDOTT.*

the concept *architect* as a *foaf:Person*. Furthermore, as the `Schönbrunn Palace` is one of the most favorite Viennese attractions, it is annotated as a *city highlight*. Besides, it is built in *rococo style*. It is closely located to the attractions `Schönbrunn Zoo`, `Wagenburg` and `Gloriette`. Additional information about this attraction is given by the link to its Wikipedia Website.

3.4.2 Adaptation of the semantic similarity measure for the tourism domain

In (Section 2.4.4) we presented a list of semantic similarity measures and discussed their advantages and shortcomings. In order to select the similarity measure which is best suited for our problem domain, we need to identify a set of requirements that should be accomplished by the measure. First of all, as we have modeled all related domain knowledge within a single tourism ontology, we can restrict to the set of *single ontology* measures that compute the similarity between objects within the same ontology.

Single ontology similarity measure.

A further critical requirement is that the measure should not only exploit taxonomic relations to predict a similarity value between objects, but incorporates domain-specific ones as well. A reason for this is the fact that destinations may classify attractions not only based on their functional type (e.g., museum or palace) but also assign them to specific topics with respect to their distinctive features (e.g., the tourism office of Vienna classifies attractions that are related to the imperialism to the topic 'Imperial Vienna'). To reflect such facts in the knowledge base, domain-specific relations have to be used. Relations such as `hasArchitecturalStyle` or `isBuiltBy` are further examples of domain-

Consideration of different types of relations.

specific relations, which might influence the similarity values. These values are thus dependent on the relationships within the semantic network. Hence, the similarity measure needs to support several kinds of relations.

Consideration of the depth within the hierarchical ontology.

Next, the similarity measure should consider the depth of the nodes within the hierarchical ontology fragment. Hence, the similarity between general nodes (e.g., `Event-Object-Place`) should be lower than between rather specific nodes at the bottom of the hierarchy (e.g., `ArtGallery-CulturalHistoryMuseum-TechnologyMuseum`). Another requirement, which should be fulfilled, is that the distance between siblings should be greater than the distance between parent and child nodes. In other words, the `Natural-History Museum` and the `Science Museum`, which are both subclasses of the class `Museum`, should be less similar than the `Natural-History Museum` compared to the class `Museum`.

Varying distances between sibling and child-parent nodes.

All these requirements influence the semantic strength of a link between two nodes within the ontology, and thus, the resulting similarity value. To sum up, the weight of a link should be affected by the following:

- the kind of relations (hierarchical and non-hierarchical)
- the depth within the hierarchy
- the sibling-parent similarity

As we do not exploit knowledge from a text corpora, we cannot rely on information-based measures. In addition, they are rather used in connection with upper-level ontologies and not with domain-specific ones. Our research was subsequently considerably influenced by that of Sussna [136] as it incorporates all needed requirements. As it provides the possibility to specify weights for different kinds of relations (comprising non-hierarchical ones) it allows for the comparison of domain concepts based on their properties as well. Hence, we abstain from combining it with a feature-based measure.

Billig et al. adapted in [18] the measure by Sussna for the RDF environment. The weight between two adjacent nodes c_1 and c_2 is defined by:

$$w_{Billig}(c_1, prop, c_2) = \frac{w(c_1, prop) + w(c_2, prop)}{2 * max(len_e(c_1), len_e(c_2))} \tag{3.1}$$

$$w_{Billig}(c_1, prop) = max_{prop} - \frac{max_{prop} - min_{prop}}{propCard(c_1, prop)} \tag{3.2}$$

$$len_e(c) = length(MinPathWeight(c, root, f, rdfs{:}subClassOf)) \tag{3.3}$$

$(c_1, prop, c_2)$ describes an RDF triple, whereby the property $prop$ links node c_1 with node c_2. Billig et al. [18] define the property weight range of `rdf:label` as [0,0] and the ranges of `rdfs:subClassOf` and `rdf:type` as [1,2]. The property weight range of domain-specific relations can be either set to a similar range or to a constant value that is determined based on the results of experiments. $propCard$ is the number of properties of a given type leaving the corresponding node c. $len_e(c)$ depicts the minimum path weight from node c to the root node of the ontology based on the number of is-a relations. As the path weight f is equal to the path length, the minimum path weight is calculated by summing up the number of is-a edges on the path from node c to the root node.

3.4.3 Process to calculate similarity between tourism objects and quantify their attractiveness

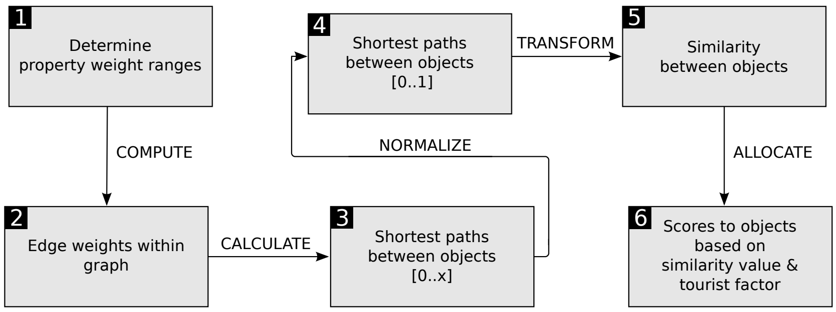


Figure 3.13
Process steps in order to calculate the similarity values between tourism objects and derive a score for each object that depicts its attractiveness for a specific tourist type.

The adaptation of Sussna's approach to our tourism domain and the propagation of scores to the tourism objects incorporates following process steps (cf. Figure 3.13):

1. *Determination of the property weight ranges:* According to [136], we set the weight range of `rdf:label` as [0,0], the range of `rdf:type` and `rdfs:subClassOf` as [1,2] and the weight range of `hasTopic` to the constant weight 1.5. These weights can be set in an external RDF file, which can be read in programmatically. (Figure 3.14) depicts a fragment of our tourism ontology and shows the usage of these relations.

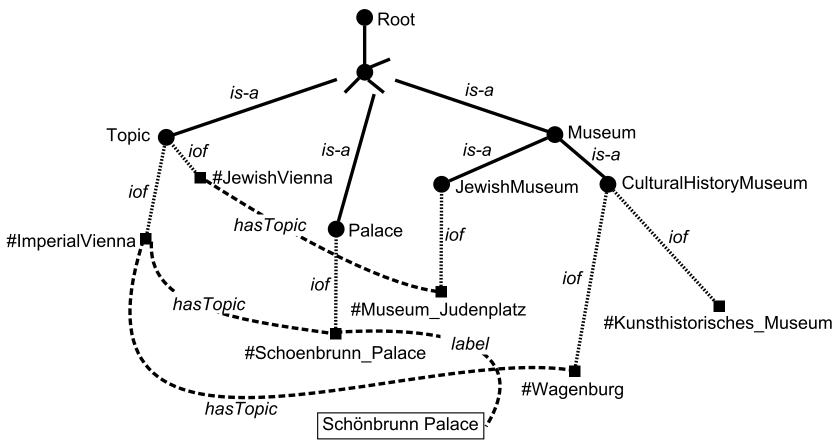


Figure 3.14
Ontological fragment of our tourism ontology, depicting different kinds of relations: rdf:type equals iof (instance of) and rdfs:subClassOf equals is-a.

2. *Computation of the edge weights within the ontological graph:* In this step, the weight definitions are used to define the weights of selected triples (subject, predicate, object) within the ontological graph. Hereby, all triples will be selected, whose property belongs to the set of properties with defined weights. The weights are calculated based on the approach by Sussna [136] & Billig [18]. According to (Equation 3.1), the maximum of the shortest paths between both nodes and the root node is calculated by counting the number of is-a edges to the root node.

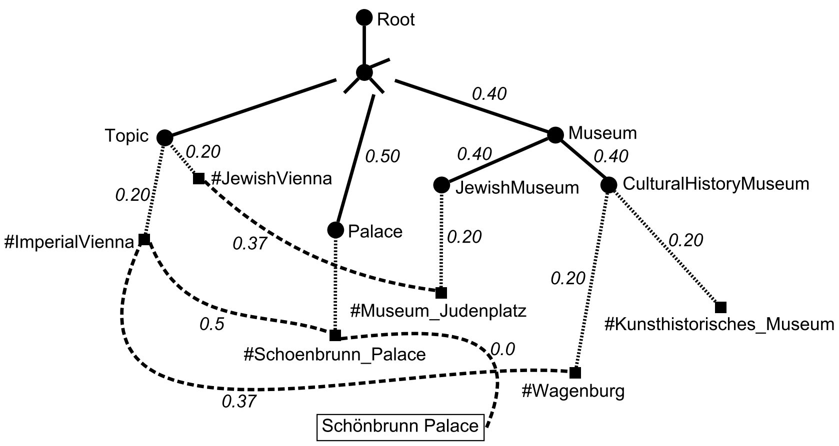


Figure 3.15 *Ontological fragment of our tourism ontology, depicting the relations with attached weights, which are calculated based on the approach by Sussna [136].*

3. *Calculation of the shortest paths between the tourism objects:* Using the edge weights from step 2, the shortest path lengths between all tourism objects can be calculated. To find the shortest path between a pair of objects, the algorithm of Dijkstra is applied in this step.
4. *Normalization of shortest path lengths:* As the length of a path between two given tourism objects reflects their distance (the longer the path, the larger is their distance), we need to apply normalization in order to compare the different path lengths. The distance between two objects should scale to a value between [0..1]. The value 0 means that the objects are identical, whereas value 1 denotes that they are dissimilar. Normalization is applied by a) identifying the maximum length of all paths between the objects, and b) calculate the distance values based on the ratio of the individual path lengths and the maximum path length.
5. *Transformation of the distance values into similarity values:* The normalization is the prerequisite in order to define the similarity values between tourism objects. As semantic similarity is the inverse notion of the semantic distance, its value is derived by $1 - \textit{distance value}$.
6. *Allocation of scores to the tourism objects:* In this step, a numerical score is allocated to all objects with respect to the predicted similarity values. Those objects that have been marked by the domain expert get that score that has been assigned to them for a certain tourist factor by the domain expert. The scores of all other objects are derived based on their similarity values with the marked tourism objects. For example, if the score of object A is manually set to 100 by the domain expert for the tourist factor *Cultural Visitor*, the score of object B for this tourist factor will be 60, if the similarity between A and B is predicted to be 0.6. In case that the domain expert marked several objects, a particular object would get assigned different score values. To prevent this happening, an object will always receive the maximum possible score. In addition, a boost factor can be set to increase the score of tourism objects,

which are classified as `city highlights`. The reason for this is the fact that if the domain expert marks only one object, then objects that are semantically apart from this specific object would get a low score even if they are important city attractions. The purpose of the boost factor is to increase their score as they might be attractive for tourists anyway.

Following the 6-steps procedure, we can fill the missing values of the linkage between the tourism objects and the tourist factors by propagating the scores given by the domain experts based on the similarity between the tourism objects. In this way, each tourism object is linked to each of the tourist factors whereby the strength of the linkage is expressed through a numerical value. Likewise the tourist profile, the tourism object profiles can be represented by a vector as well.

3.5 A vector-based distance metric

As soon as a particular tourist profile and the profiles of the given tourism objects are available and modeled in the same way, a function has to be defined that compares these profiles. If a tourist profile matches the characteristics of an object, this object should be recommended to the respective tourist. Therefore, a matchmaking measure is needed that examines whether they share similar features.

To estimate the similarity degree between a particular tourist profile and a set of tourism objects, a *vector-based matchmaking approach* is applied whereby a given profile and each tourism object constitute vectors and are compared in a vector space. The dimensions of the vector space correspond to the predefined set of tourist factors such that each distinct tourist factor represents one dimension in that space. (Figure 3.16) depicts an exemplatory representation of two tourism object vectors (TO vectors) and a tourist profile vector in a 3-dimensional vector space that is defined by the dimensions *Cultural Visitor*, *Nature Lover* and *Sight Seeker*.

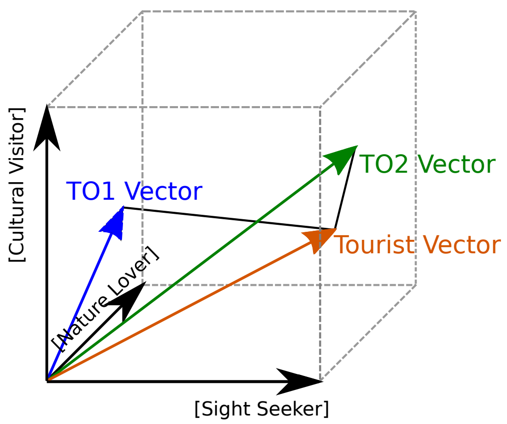


Figure 3.16 *A representation of vectors in the 3-d space.*

The *Euclidean metric* is frequently used in the information extraction domain. We use Euclidean distance as we are rather interested in the actual difference between a particular tourist profile vector (T) and the individual tourism object vectors (O). The Euclidean distance takes the magnitude of difference between the vectors in account. The Euclidean metric is defined as follows:

$$Distance_{Euc}(\vec{T},\vec{O}) = \sqrt{(t_1 - o_1)^2 + (t_2 - o_2)^2 + ... + (t_n - o_n)^2} \quad (3.4)$$

This measure calculates the distance between two given vectors. The resulting value of this metric is a distance value. The smaller the distance the more similar the vectors are. In (Figure 3.16) the distance between the tourist vector and the tourism object vector *TO2* is smaller than the distance between the tourist vector and the tourism object vector *TO1*. Hence, *TO2* is more similar to the tourist vector and should be recommended to the tourist in the first place.

In general, the distances between the tourist vector and each tourism object vector can be calculated and then sorted from the least to the greatest distance. The top-N candidates of this list can then be recommended to the tourist as shown in (Figure 3.17).

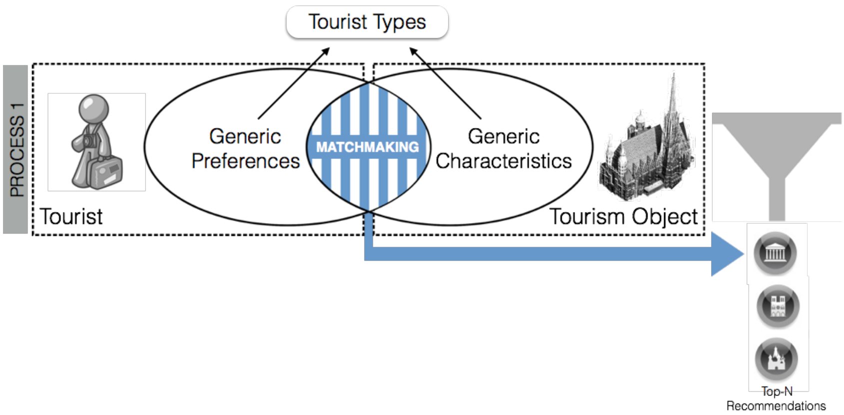


Figure 3.17 *A list of top-N recommendations based on the first matchmaking process.*

3.6 Summary

In this chapter we demonstrate the first matchmaking process. We show that tourist factors can be used to generate a high-level user profile and represent it in a vector-space model whereby the factors form the dimensions and the values depict how much the tourist identifies him- or herself with the individual factors. We propose a semi-automatic way to represent the generic characteristics of tourism objects in this vector-space as well. Thereby, domain experts mark for each of the tourist factors a small set of tourism objects that are related to these types. In a second step, the ratings of these experts are propagated to other tourism objects that are similar to the ones rated by the domain experts based on a semantic similarity measure. A vector-based similarity measure is then used to calculate the similarity between the tourist vector and the vectors of the tourism objects in order to compute their degree of matching and propose a first list of recommendations.

4 Second matchmaking process

In this chapter, the *second matchmaking process* is outlined. For an overview of both matchmaking process see (Section 1.5). The focus of this matchmaking process is to *refine* the generic tourist profile and to enrich the generic preferences of a tourist through more specific interests (e.g., a tourist may be a *Sight Seeker* but may express a dislike of churches). This is achieved by exploiting *tourist feedback* on the proposed top-N list of objects and by using this information to derive a more fine-granular profile that is capable to model statements such as a 'dislike of churches'. In contrast to the *tourist factors* utilized in the first recom-

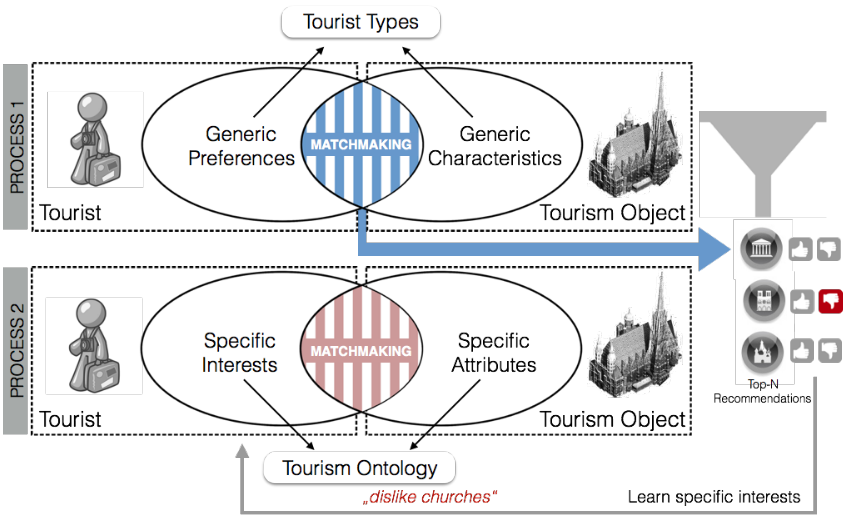


Figure 4.1
The second matchmaking process based on specific interests.

mendation process, an *ontology-based approach* is favored to model the specific profiles of the tourist as well as the tourism objects (see Figure 4.1). This approach is based on the tourism ontology developed in (Section 3.4.1). The advantage of an ontology model is that it allows a *fine-granular tracking* of user interests. For example, an interest expressed in a particular object (e.g., the `Imperial Furniture Collection` located in Vienna) may increase the *level of interest* in objects related to this category (i.e., `cultural history museum`), to its parent category `museum` or to a related topic such as the `House of Habsburg`. This dynasty is one of the most important royal houses of Europe. The museum `Imperial Furniture Collection` gives insight into how the imperial family used to live. In this way, different kinds of relations within the domain model can be exploited to predict the interests of yet unexplored concepts (i.e., concepts that are not yet rated by the user) [23]. Hence, the ontological model facilitates the propagation of user interests between parent and child concepts or any other related concept. Without having such a semantic network, predicting the interest scores of other concepts would not be possible. For instance, if the user expresses interest in an object connected with the concept `art gallery`, the interest level of this concept would be increased but not the level of the concept `museum` as it is unknown how these concepts are related.

Such user interests are typically represented as an *overlay* of the conceptual domain model, whereby each concept is associated with a (numerical) interest value that quantifies the level of interest in this concept. In this way, the ontological user profile can be represented as a high-dimensional vector whereby the dimensions represent the concepts and the values represent the interest value (see Figure 4.2).

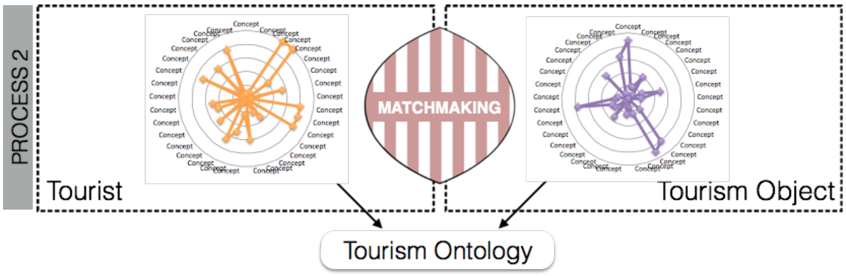


Figure 4.2
The specific profiles can be represented as high-dimensional vectors.

Not only the user interests, but also the characteristics of the tourism objects can be represented as overlays of the domain model. Hence, the tourism objects can be modeled in the same vector space. The ontological concepts represent the dimensions while the values indicate how much the tourism objects correspond to the different concepts. In order to examine whether the tourism objects match a particular user profile, their high-dimensional vectors can be related against the high-dimensional vector of the tourist. To achieve this goal, following issues have to be tackled:

- a study to explore construction techniques of overlay models, modeling the specific interests of tourists,
- a representation of the specific user and tourism object profiles in form of overlay models
- a measure to match the high-dimensional tourist profiles against the high-dimensional profiles of the tourism objects

4.1 An overview of existing overlay construction techniques

A number of techniques can be used to build user profiles which are represented as overlay models of ontologies. Basically, they all require a way to determine the concepts that are interesting for a particular user, e.g., based on his/her feedback. The degree of interest is typically expressed through a quantitative weight value attached to each concept.

One construction technique to build such user profiles is based on the *tree coloring* method [138]. Tree coloring is a method that colors or tags nodes in a tree with information. This method is applied in the system

Persona, which is a personalized Web search tool that enables users to retrieve Web documents that are tailored to their individual preferences. The tree used in this system is the Web taxonomy of the Open Directory Project (ODP). Thereby, the nodes of the tree represent the semantic contexts of the pages. Each time, Persona delivers a page to a user, they are asked to rate this page in a positive or negative way. This feedback is then used to update their profiles, which represent the collection of all concepts of the ODP Web taxonomy. As each page has already been annotated with concepts of the taxonomy, the corresponding nodes of that page are 'colored' by the number of times they have been visited and the number of times the user has provided positive or negative feedback. (Figure 4.3) depicts a fragment of the taxonomy used by Persona.

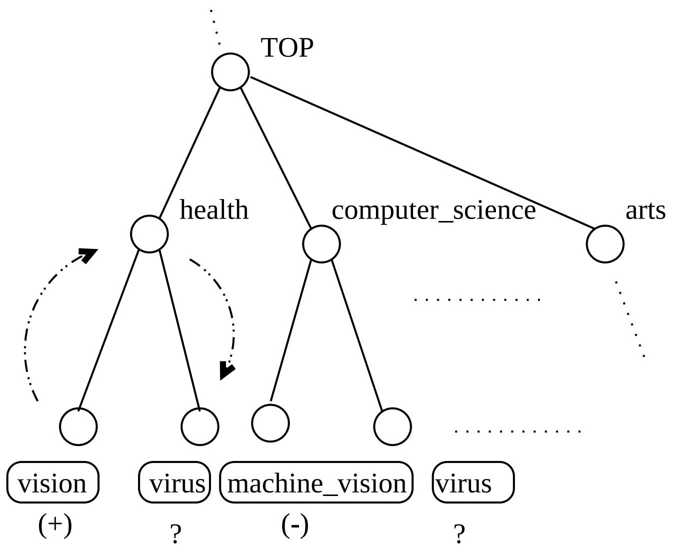


Figure 4.3
Tree coloring method. A taxonomical fragment depicting colored nodes [138].

The nodes `vision` and `machine_vision` have already been colored based on user feedback (i.e., the user preferred information concerning 'vision' related to 'health' and expressed disinterest in the concept 'machine_vision'). In case that the user now queries for pages about 'virus', the system does not have any information regarding the user preferences for this word. In such a situation, the system exploits the taxonomical structure to find nodes that are semantically similar and that are already colored by user information. As `vision` has already been colored with positive user interest and as `vision` and `virus` share the parent node `health`, the system might propose pages that are related to 'virus' in the context of 'health'. In this way, Persona exploits ontological information to infer information about yet unknown user preferences.

Another similar method is *domain inferencing* [48], which utilizes the taxonomical or ontological structure of the domain model to propagate interest probability values to related concepts. Kobsa et al. use this method in a personalized tourist guide to populate user models more quickly. They distinguish between two techniques, namely *sidewards* and *upwards propagation*. *Sidewards propagation* assumes that if the user is interested in a minimum percentage of direct sub-concepts of a given concept, then he/she is also interested in the remaining sub-concepts. Thereby, the probability value of these remaining sub-concepts is calculated as the means of the probabilities of the current sub-concepts.

Upwards propagation works similar, i.e., if the user is interested in a minimum percentage of direct sub-concepts of a given concept, then he/she is assumed to be also interested in this concept with a particular interest percentage. Kobsa et al. used a percentage threshold value of 75% for sidewards inferences and a threshold value of 60% for upwards inferences.

Table 4.1
An example depicting sidewards propagation, adapted from [48].

Museums			
Modern Art	Applied Art	Contemporary Art	Fine Art
p=0.3	p=0.5	p=0.4	

According to Table 4.1, sidewards propagation can be applied as the percentage of marked concepts equals the threshold value of 75%. The user's presumable interest in `Fine Art` is calculated as follows:

$$p = \frac{0.3 + 0.5 + 0.4}{3} = 0.4$$

The *spreading activation algorithm* proposed by Sieg et al. [128] is a generalization of the previous stated techniques in that the relations between the concepts within the ontological user profile have pre-computed weights, which influence the amount of activation (interest value) that is propagated to the neighbour concepts. As depicted in (Figure 4.4), the

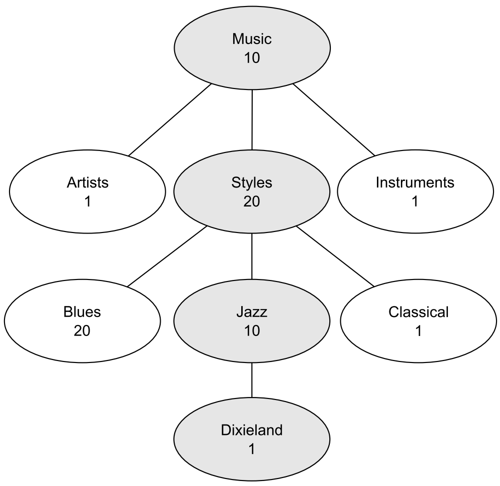


Figure 4.4
Fragment of an ontological user profile where interest scores of concepts in color grey are updated based on spreading activation. Adapted from [128].

concepts contained in the user profile are assigned an initial activation (interest) value. Let's assume that the user is interested in the concept `Dixieland`. Hence, the interest score of this concept will be increased. Based on a spreading activation algorithm, the interest scores of all related concepts (that are super concepts of the current one) will increase as well, comprising the concepts `Jazz`, `Styles` and `Music`. The amount of increase in score depends on the weight of the concept relationships. The weights scale between the values 0 and 1 whereby a value of 0 indicates that the respective concepts are rather dissimilar whereas a value of 1 indicates complete overlap of the concepts.

After increasing the interest scores of the corresponding concepts, the interest scores for all concepts in the user profile are normalized in order to prevent them from continuously escalating throughout the semantic network. The normalization works under the assumption that if the user expresses interest in a specific set of concepts, the score for other concepts decreases. For that, the scores of all concepts are treated as a vector, whose length is normalized to a pre-defined constant value. The concepts of the user profile are then updated with the normalized scores.

4.2 Creating overlay models for tourist and tourism object profiles

In order to model the specific interests of tourists and the specific characteristics of tourism objects we follow the main ideas of the spreading activation algorithm proposed by Sieg et al. [128] and the taxonomy-driven profile generation by Ziegler et al. [163]. In our case, a profile is represented as an overlay of the domain model describing the tourism attractions. Such a profile can be modeled as a vector, whose number of dimensions correspond to the *number of concepts* of the domain model and whose scores depict the *degree of match* with the various concepts.

In the following, it is explained how the profiles of the user and tourism objects are generated and how the underlying domain model is exploited to enable *score inferences* between concepts in a profile vector.

Let us assume that the user is interested in the `Imperial Furniture Collection` of Vienna. The `Imperial Furniture Collection` is semantically annotated with the concepts *Cultural History Museum* and *Imperial Vienna*, which are leaf nodes of the ontological hierarchy (cf. Figure 4.5). If the user explicitly states interest in the `Imperial Furniture Collection`, he/she might also prefer objects that are classified under the same categories. Hence, we can assign a certain amount of interest score to the leaf concepts with which the Imperial Furniture Collection is annotated, namely *Cultural History Museum* and *Imperial Vienna*. We can further exploit the ontological hierarchy to infer an interest score for those concepts that are super-concepts of the concepts that have been implicitly rated by the user. The user profile is thus formed by the numerical scores assigned to the concepts of the ontology. Such an ontological approach successfully tackles the cold-start problem of recommender systems as it is hard to generate recommendations when hardly any user information is available. Utilizing the semantic relations within the ontological graph allows to infer interest scores for concepts that have not been explicitly rated by the user yet. If a certain amount of interest score is assigned to the concept *cultural history museum*, this score can also be propagated to its parent concept *museum*. However, not the full amount of score that is assigned to the leaf concept should also be assigned to the concepts that are located on upper layers. First, if the parent concept gets assigned the same amount of score as its child concept, this would contradict the *is-a relationship* between these classes. Such a case would require a *same-as relationship*. A child class is more specific than its parent class and exhibits certain features that its parent class is missing. To correspond to this fact, a *decay factor* is required to decrease the amount of score as one goes up in the hierarchy. In addition, the amount of score that is propagated to the parent concept also

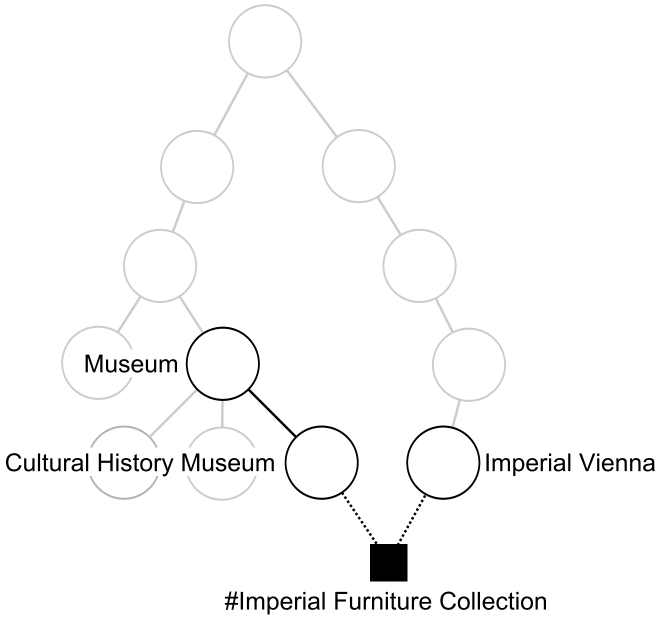


Figure 4.5 *An ontological fragment depicting the semantic annotation of the tourism object 'Imperial Furniture Collection'.*

depends on the number of siblings of the child concept. The smaller the number of siblings the more amount of score can be propagated to the parent concept [163]. For example, if the concept *cathedral* is the only sub-concept of *church*, than the full amount of score (with regard of the decay factor) can be assigned to the concept *church* as this concept is not influenced by any other sub-concept. The concept *museum*, however, has many sub-concepts (e.g., art gallery, natural history museum, technology museum, military museum). If the user has rated the concept *cultural history museum*, only a small amount of score can be propagated to the parent concept *museum* as the information is yet missing whether he/she is also interested in the other museum types. Following equation expresses the score propagation from a given child concept c to its parent concept p [163]:

$$score(p) \quad = \quad df \cdot \frac{score(c)}{nr(c)}$$

df expresses the decay factor and nr is a function that depicts the number of child concepts of a given parent concept p. (Figure 4.6) depicts the assignment of interest scores to concepts of the user profile after conducting score propagation. Let us assume that the the score value of the leaf concepts has been set to the value 15 and the decay factor to the value 0.8. The real number, however, depends on how often the user has already rated this concept and can either be positive (the user has shown interest) or negative (the user has shown disinterest). See (Section 5.2) for more details. Based on the values given above the value of 4 (0.8 * 15 / 3) is assigned to the concept *museum* and the value of 12 (0.8 * 15 / 1) to the concept *imperial vienna*. These scores are further propagated upwards in the ontology.

The overlay models of the individual tourism objects are created in a similar way. First, the set of leaf nodes that directly describe a tourism object is identified. In a next step, a certain amount of score is assigned

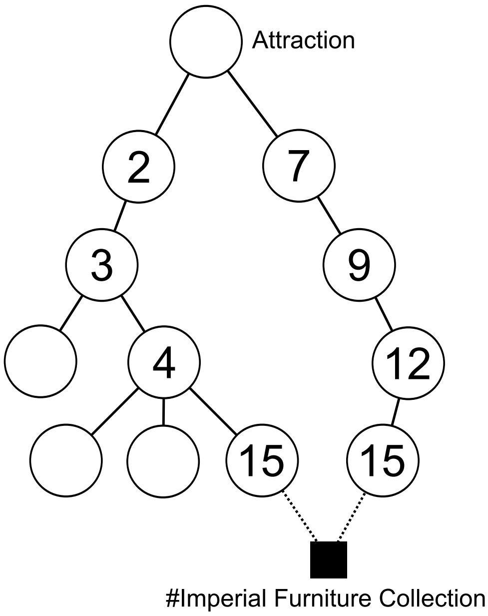


Figure 4.6
An ontological fragment depicting the score propagation after the user has expressed interest in the tourism object 'Imperial Furniture Collection'.

to them, which is then propagated to related concepts in the ontological hierarchy. However, in contrast to the score allocation process in case of a user profile, the score that is assigned to the leaf nodes is set to a maximum score max_s. max_s is the score that can be achieved at most in the user overlay model if the user rates this concept a vast number of times. (Figure 4.7) shows the score propagation to generate the overlay model of the tourism object `Imperial Furniture Collection`. max_s is set in this example to the value 50. According to (Figure 4.6), user interest in the concept *cultural history museum* scores 15. If the user rates this concept repeatedly, the score increases and eventually reaches also the value 50.

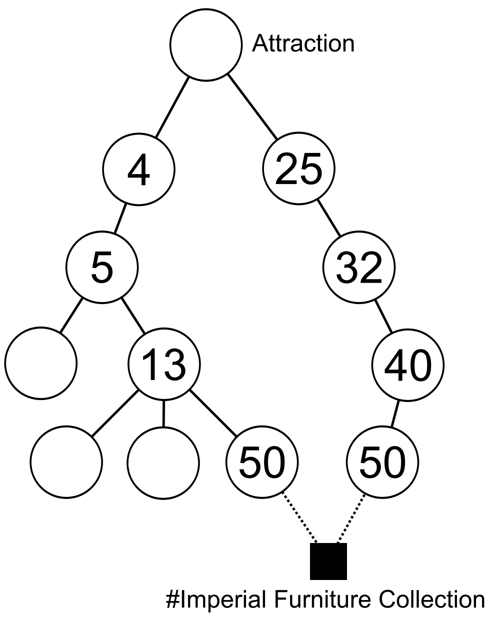


Figure 4.7
An ontological fragment depicting the score propagation to generate the profile of the tourism object 'Imperial Furniture Collection'.

4.3 Pearson similarity measure

In order to measure the similarity between the user profile and the profiles of the various tourism object profiles we use Pearson correlation. Pearson correlation allows to discover positive and negative correlation, thus detecting similar pattern of variation in the vector profiles. It implicitly normalizes the values of the vectors to their arithmetic mean. For a given user (u) and a particular tourism object (to) Pearson correlation is defined as follows:

$$sim(u,to) = \frac{\sum\limits_{i \subset C} (v_{u_i} - \bar{v_u}) \cdot (v_{to_i} - \bar{v_{to}})}{\sqrt{\sum\limits_{i \subset C} (v_{u_i} - \bar{v_u})^2} \sqrt{\sum\limits_{i \subset C} (v_{to_i} - \bar{v_{to}})^2}} \tag{4.1}$$

C depicts the number of concepts in the domain model. Hence, the vectors v_u and v_{to} have C dimensions. $\bar{v_u}$ and $\bar{v_{to}}$ depict the mean values of both vectors. The similarity value sim(u,to) ranges between -1 and 1, depending on whether the vectors correlate in a positive or negative way. If there is a large overlap in common concepts between the two given vectors, the match between the profiles is high, whereas in case of no overlap or contradicting scores the similarity between the profiles is low. Ontological inference of scores facilitates the increase in overlap of concepts. For example, if the user shows interest in the concept *cathedral*, a specific amount of score is allocated to concepts related to religious architecture as well and therefore, the user profile gets more similar to the profiles of other religious attractions such as churches or synagogues as the profiles have a large overlap in concepts related to religion.

4.4 Summary

This chapter outlines the second matchmaking process. The focus of the second matchmaking process is to refine the generic tourist profile and to enrich the generic preferences of a tourist through more specific interests. This is achieved by exploiting both positive and negative tourist feedback on the proposed top-N list of objects and by using this information to derive a more specific profile. In contrast to the tourist factors utilized in the first recommendation process, an ontology-based approach is used to model the specific profiles of a tourist as well as the tourism objects. In order to generate these specific profiles, the main ideas of spreading activation are applied. The specific profiles of both the tourist and tourism objects are represented in a vector-space, whereby the ontological concepts form the dimensions of this vector space. The values depict the interest of the tourist and respectively, the relatedness between the tourism objects and these concepts. In order to measure the similarity between the user profile and the profiles of the various tourism objects Pearson correlation is used.

5 Combined matchmaking

Let us have another look at the matchmaking framework depicted in (Figure 5.1). In the previous chapters we have explained the two matchmaking processes in order to model both the generic and specific preferences of tourists and how to match them against the profiles of tourism objects. In the next sections, we explain how the two processes interact with each other to generate a top-N list. A greedy approach is introduced so that the tourism objects of the generated top-N list fit the choice of tourist types the user has made at the beginning of the recommendation process. Moreover, it is shown how tourist feedback is exploited to refine the specific interest profile of a user and deliver a more accurate list of tourism objects. Finally, we show how diversification of the proposed recommendations may lead to better user satisfaction. To sum up, following issues are discussed in the following:

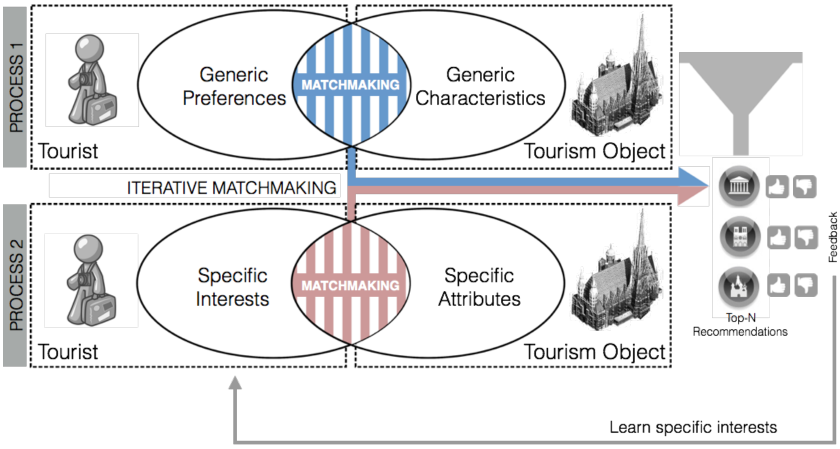


Figure 5.1 *Overview of the matchmaking framework.*

- ❏ a combination of the two matchmaking processes to generate a top-N list,
- ❏ an introduction of a greedy approach to guarantee that the objects within the top-N list correspond to the user's choice of tourist types,
- ❏ an exploitation of user feedback, and
- ❏ a recommendation diversification

5.1 Combination of the matchmaking processes

In order to calculate the similarity between a user and a particular tourism object, a weighting function is used that controls the influence of the two matchmaking processes on the resulting similarity value between a user (u) and a certain tourism object (to). The weighting factor α controls the weight of the respective similarity measure. It scales between the values 0 and max_{weight}, which is set to a value ≤ 1. In our case, max_{weight} is set to the value 0.8.

$$sim(u, to) = (1 - \alpha) \cdot sim_{generic\ preferences} + \alpha \cdot sim_{specific\ interests}$$

At the beginning of each recommendation session, the *tourist types* are used as a stereotypical approach to initialize the generic preference profile of a user. At this moment, information about more specific interests is not available as no feedback has been given by the user. Hence, the specific interest profile contains no essential information and is not helpful in the recommendation process. The weighting factor α is set to the initial value 0 as no user ratings are available. Hence, the first recommendation of top-N objects is solely generated by using the first matchmaking process based on the tourist factors. As soon as feedback from the user on the proposed objects is obtained, the respective factor α increases and the second matchmaking process based on interest propagation can be incorporated in the overall recommendation process. A new list of top-N objects can be recommended based on both processes. For example, let's assume a user, who is both a *Sight Seeker* and a *Cultural Visitor* and who has already rated a certain amount of concepts related to *sightseeing*. If the similarity between the user and a tourism object needs to be calculated, both matchmaking processes can be applied as some specific interests of the user are known (and thus, the weighting factor α with respect to the sightseeing type is set to a value higher than 0).

The factor α increases following a linear function, the more tourism objects are rated (see Figure 5.2). We assume that as soon as the user has rated 5 tourism objects, α is set to the value of max_{weight}. In this case, enough user interests are elicited to rely more on the second matchmaking process to determine the top-N tourism objects.

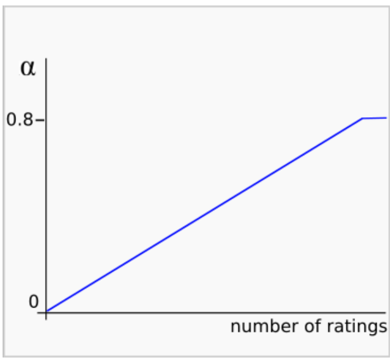


Figure 5.2
Using a linear function to set the value of the weighting factor α.

5.2 Exploiting tourist feedback to learn specific interests

The user has the possibility to rate the proposed tourism objects by giving either positive or negative feedback. This kind of explicit feedback is used to refine the user's specific interests and deliver a new set of objects that better suit the user's interests. To achieve this goal, an approach is needed that

- relates the *explicit* ratings of tourism objects to *implicit* ratings of concepts that are used to annotate these objects, and
- transforms the number of positive/negative ratings into a numerical *interest score*

To tackle the first aspect, as soon as an object receives a user rating, the set of leaf concepts is identified that semantically describe the corresponding object. In a second step, a sigmoid function (cf. Figure 5.3) is defined that calculates a numerical interest score for these concepts based on the number of ratings that have already been given to these objects by the user. Following consideration constitutes the choice for a sigmoid function to model both positive and negative ratings. As soon as a particular concept obtains (implicit) ratings, a rather high amount of interest score should be assigned to this concept so that the follow-up recommendations immediately reflect the user's current interests. However, the more often this concept is implicitly rated, the less additional score should be assigned to this concept as the recommendations are already aligned towards this interest concept. (Figure 5.3) depicts the sigmoid

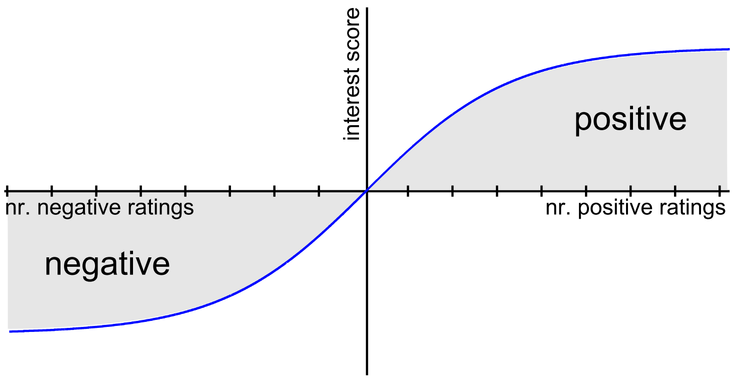


Figure 5.3
A sigmoid function is used to transform user ratings into interest scores of ontological concepts.

function. As shown, the interest score adds up with respect to the number of ratings, but stabilizes after some time according to a maximum value. The sigmoid function should be defined in such a way that the maximum value corresponds to the maximum score max_s that is assigned to the leaf nodes during the generation of an overlay model of a tourism object (cf. Chapter 4).

5.3 Preserve the right selection of tourism objects

At the beginning of the recommendation process, the list of tourism objects should reflect the choice of tourist factors the user has selected beforehand. Let us assume that a particular user has the profile that is shown in (Figure 5.4). He/she strongly identifies with the factor *Sight Seeker*, but also has certain interests in sports and recreational activities. Obviously, the set of recommended tourism objects should match the selection made by the user. However, simply taking the top-N elements

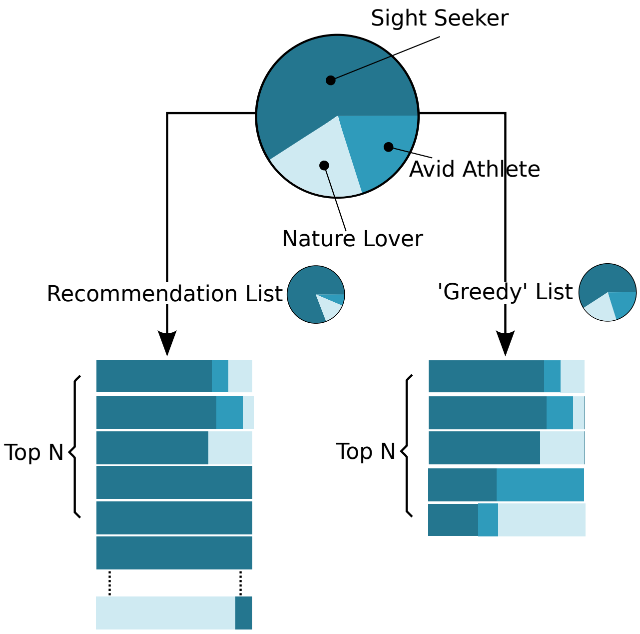


Figure 5.4
Preserving the right proportion of tourism objects with respect to the user's choice of tourist factors.

of the ranked list of tourism objects that result from the matchmaking processes may not fully match the user's choice of tourist factors. This case is depicted on the left side of (Figure 5.4) and labelled as *recommendation list*. As one can see, the top-N elements are much more related to sightseeing (cf. small circle on the left side) than declared by the user (big circle on the top) and thus contain too few sports and recreation features. This can happen when a destination offers a huge set of objects that are highly related to the *Sight Seeker*. As the user has expressed a high correlation with the factor *Sight Seeker*, the profiles of these objects are very similar to the user profile and therefore receive a high rank in the list. To counteract this situation, a greedy-based approach is introduced, which is depicted on the right side of (Figure 5.4). It calculates a new set of top-N tourism objects by a) considering their position in the original recommendation list and b) seeking to select the right mix of tourism objects that reflects the user's choice of tourist factors.

Algorithm 1: A greedy-based approach to preserve the tourist factor ratio of the top-N objects in R.

L = list of tourism objects (to) sorted according to similarity with user u;
$R = \{\}$ result set R, which is empty at the beginning;
F = ranked list of tourist factors according to tourist factor ratio in R;
k = number of top-N objects that R should contain in the end;
Insert first to of L in R;
Delete first to in L;
foreach $i = 2$ *to* k **do**
    F = Calculate TouristFactorRatio(R);
    to = Identify TourismObject(L,F);
    Insert (to) in R;
    Delete (to) in L;
end
return R;

At the beginning, the algorithm inserts the first tourism object to of the recommendation list L into the result set R. The ratio of the tourist factors related to this tourism object is calculated. A ranked list of tourist factors F is generated, whereby the first tourist factor in this list depicts the factor that is most under-represented within the result set R, whereas the last one is the type that is most related. For example, the first tourism object may be very much related to sightseeing. However, the factor *Avid Athlete* may be under-represented with respect to the user's choice depicted by the big circle on top of (Figure 5.4). Hence, the next object that should be inserted in the list should be related to sports so that the user's choice in tourist factors is reflected as accurately as possible (therefore, the first object in the F list is the factor *Avid Athlete*). For that all objects of the recommendation list are examined and the next object which is related to the factor *Avid Athlete* is inserted in the list. (If no further object related to this factor can be identified, the next tourist factor of the F list is selected and the whole process is repeated). Following that, the tourist factor ratio F of the objects already contained in the result list is once again computed and used as basis to select the next object. This process is repeated until the result set contains the required number of top-N elements. As depicted on the right side of (Figure 5.4), the greedy-based approach ensures that the top-N tourism objects roughly correspond the user's choice of tourist factors.

This greedy-based approach is crucial to obtain the right selection of tourism objects with respect to the user's choice in tourist factors at the beginning of a recommendation session. However, as soon as the user provides more feedback on the proposed tourism objects, preserving the right proportion of tourism objects with respect to the user's choice of tourist factors is less important. For example, if the user has positively rated a number of churches and thus seems to be very interested in buildings related to the concept `church` the user certainly wants to receive recommendations that rather comprise buildings of this type. Hence, the top-N elements of the recommendation list should be shown in favor of selecting a mix of tourism objects that reflects the user's choice of tourist factors.

We use a weighting factor α to control how much the proposed set of tourism objects should reflect the tourist factor ratio. At the beginning, α is set to the value 1 so tourism objects are proposed based on the user's choice of tourist factors. As soon as specific concepts (e.g., concept `church`) obtain (implicitly) positive ratings (see Section 5.2), the value of α decreases. We set up a linear function that assigns a value of 0 to α as soon as a concept has received at least three positive ratings. Hence, the tourist factor ratio does not play any role to construct the top-N elements. Instead, the elements of the original recommendation list are taken and thus, the user receives recommendations that are highly related to the concepts that have been positively rated before.

5.4 Improvement of recommendations through diversification

Recommendation systems have the goal of filtering from a huge set of possible objects a set of top-N objects that perfectly match the user profile. Obviously, these objects are good candidates to satisfy the user's needs as their profile is very similar to the one of the user. However, as a side effect, they are also very similar to each other. Hence, they provide poor coverage of the space of relevant recommendations [20]. For example, when the user expresses interest in culture, he/she may not be satisfied when the set of proposed objects comprise only museums. To give another example, when the user rates churches, he/she might not only want to receive recommendations about churches, but a mixture of tourism objects he/she has stated interest in and which are partly related to religious architecture such as cathedrals, synagogues or cemeteries. Hence, recommendation systems should seek to deliver a diverse set of recommendations that are nevertheless related to the user profile. A diverse and relevant set of recommendations certainly enhances user satisfaction as the user gets a more valuable insight into the space of objects that might be of interest for him/her.

Different strategies have been proposed that consider both similarity and diversity during the recommendation process. As pointed out in [20], similarity and diversity are orthogonal measures. *Similarity* between a tourism object profile and a user profile is defined through a function, which is independent from the similarity of another object. In contrast, *diversity* refers to the similarity/dissimilarity of the objects that are part of the result set. In this way, if a new object should be inserted in the result set, the diversity depends on the preceding similarity measurements. In the following, the *greedy selection algorithm* is explained as we use this strategy to achieve recommendation diversification for the top-N proposed tourism objects.

Algorithm 2: The greedy selection algorithm [20].

L = list of tourism objects (to) sorted according similarity to user u;
L' = subset of L, containing top b tourism objects;
$R = \{\}$; result set R, which is empty at the beginning;
k = number of tourism objects that R should finally contain;
foreach i = *1 to* k **do**
    Sort L' by quality(u,to,R) for each to in L';
    Insert first(L') in R;
    Delete first(L');
end
return R;

The greedy selection algorithm first selects the *top b* tourism objects that are most similar to a given user profile. A quality function is then used in order to guide the construction of the result set R based on the *top b* tourism objects in an incremental way [20]:

$$quality(u, to, R) = (1 - \alpha) \cdot similarity(u, to) + \alpha \cdot diversity(to, R) \quad (5.1)$$

The diversification factor α defines the importance of the respective tourism object to's dissimilarity with respect to the objects that are already inserted in R. A low diversification factor produces a result set that is close to the original ranked list of objects. A large factor favors diversification to similarity and outputs a rather different result set. The first object that is inserted in the result set R is always the one with the highest similarity. In the subsequent iterations, the remaining objects of L' are ordered according to their *quality value*. The object with the highest quality is then inserted into R. The process is repeated until the result set contains enough objects. The diversification factor α depends on the number of positive rates given by the user to certain tourism objects and thus, implicitly to certain concepts in the ontology. The more positive rates a certain concept has obtained, the lower the diversification factor should be so that the result set rather consists of objects that are related to this concept. As discussed in (Section 7.2.4), the factor α is set at the beginning to the value 0.6. It decreases based on a linear function with respect to the number of positive rates a certain concept has received. If a concept is rated 3 times or more, α is set to 0. Hence, diversification is not used anymore when constructing the result set R.

5.5 Summary

This chapter presents the combination of the first and second matchmaking process. At the beginning, tourists state their generic preferences and obtain a first top-N list of recommendations. As long as they are not satisfied with the recommendations, they can criticize the proposed tourism objects by stating positive or negative feedback, which will be used to refine their profile and to deliver a new set of top-N objects. The combination of the two matchmaking processes is done with the help of a weighting factor that controls the influence of the two processes on the resulting similarity value between a tourist and a certain tourism object.

6 Implementation

The matchmaking approach has been implemented in form of a Java Web-based prototype. This prototype recommends tourists, who are interested in a certain destination, tourism objects that are tailored to their personal needs and travel behavior. The city destination Vienna has been chosen for the reference implementation. However, as most data is loaded from external data sources, the prototype can be adjusted to serve other destinations as well.

In the following section, the Web-based user interface and the provided functions are described. (Section 6.2) outlines relevant data sets which are integrated from different data sources. (Section 6.3) illustrates a basic route recommendation service which proposes the tourist a route for visiting his/her favorite tourism objects. (Section 6.4) describes the tool used for the construction of the cDOTT ontology. (Section 6.5) describes necessary steps which have to be followed when the prototype should be adapted for another tourism destination. Finally, (Section 6.6) gives an overview over the system architecture and describes the technology.

6.1 User interface & functionality

The application provides a landing page (see Figure 6.1) that facilitates the users to describe their predispositions/generic preferences through seven tourist factors (i.e. *Sight Seeker*, *Cultural Visitor*, *Nature Lover*, *Avid Athlete*, *Action Seeker*, *Educational Buff* and *Sun Worshipper*). The users are not limited to a single factor but can choose a mixture of different factors. They can use the heart icons below each factor to indicate how much they identify with the corresponding factor. They can select from 0 to 5 hearts. The former means that the user is not interested at all in this factor, whereas the latter indicates a high conformance with this factor. By default, no hearts are selected at all. When the user hovers with the mouse over a specific factor, an overlay is shown with a description of that factor (see Table 6.1).

When the user is ready to proceed, he/she has to click on the button “Start Your Tour“. This triggers the calculation of his/her generic profile based on the selection of the seven factors. This profile is then matched with the generic profiles of the tourism objects in order to produce a ranked list of tourism objects. Another page (see Figure 6.3) is shown which depicts on the left side the ten top ranked tourism objects which fit best to his/her profile. In addition, these tourism objects are depicted on a map at the right side.

Figure 6.1

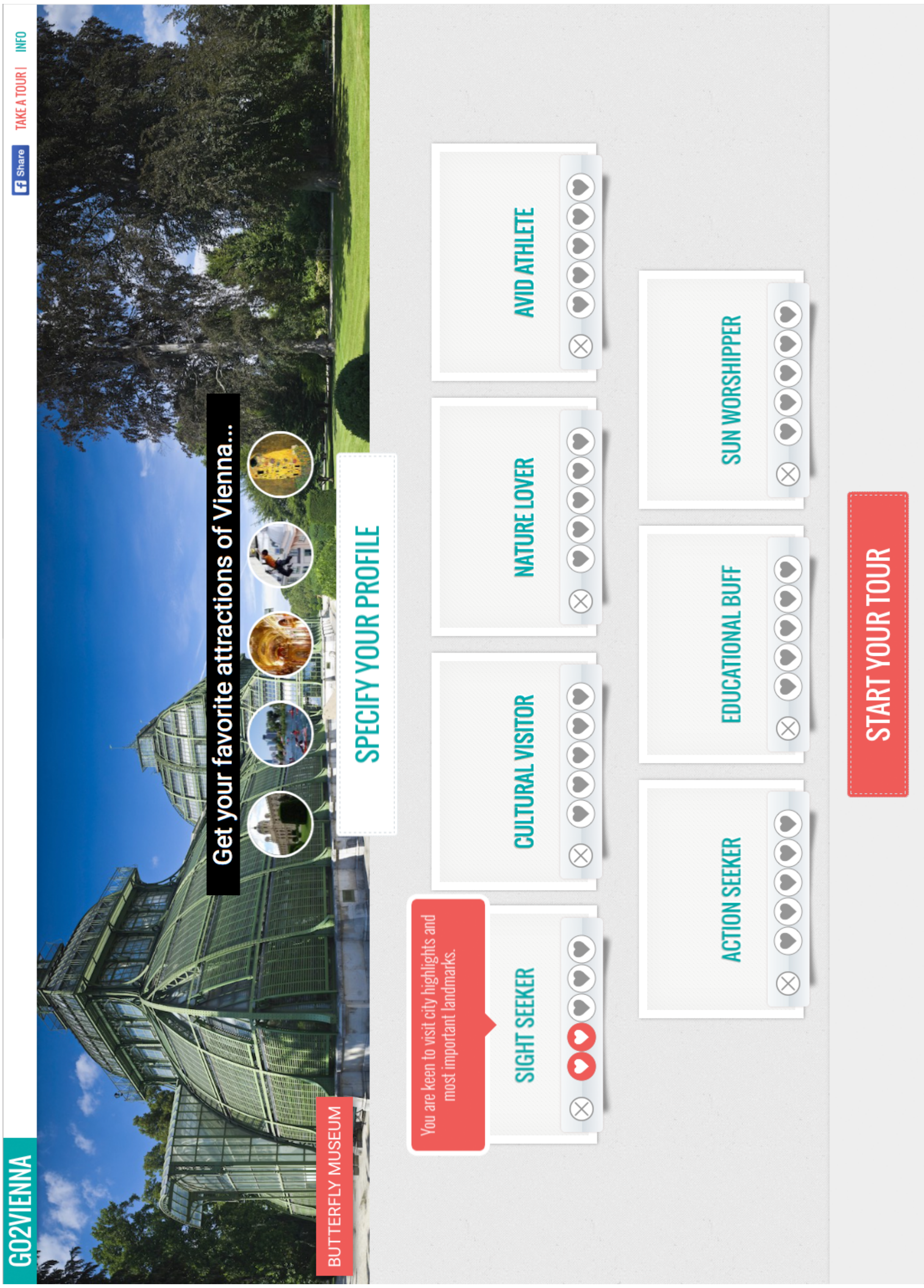

GO2VIENNA
Share
TAKE A TOUR |
INFO
Get your favorite attractions of Vienna...
BUTTERFLY MUSEUM
SPECIFY YOUR PROFILE
AVID ATHLETE
NATURE LOVER
CULTURAL VISITOR
SIGHT SEEKER
You are keen to visit city highlights and most important landmarks.
SUN WORSHIPPER
EDUCATIONAL BUFF
ACTION SEEKER
START YOUR TOUR

Figure 6.2

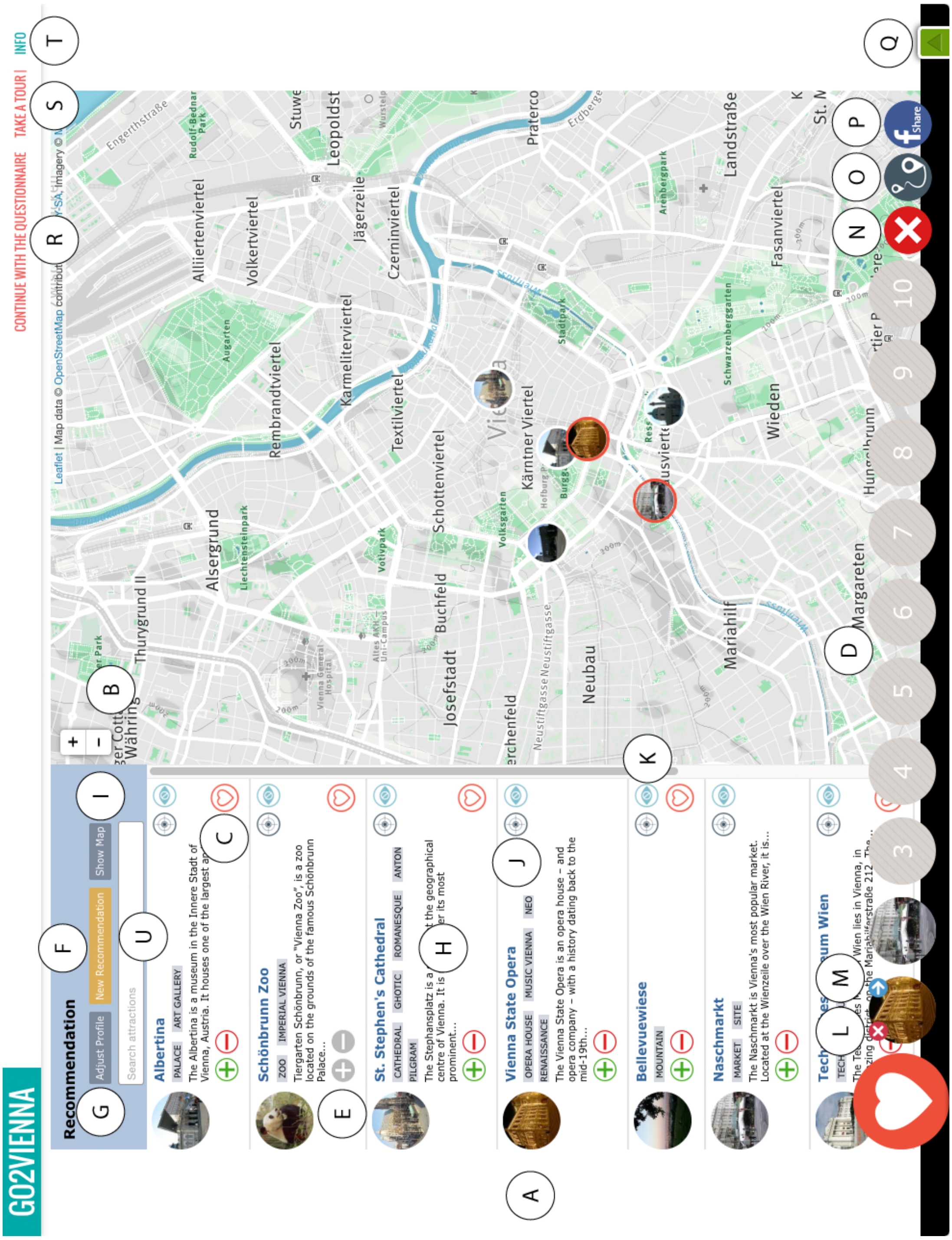

GO2VIENNA
CONTINUE WITH THE QUESTIONNAIRE
TAKE A TOUR!
INFO
Recommendation
Adjust Profile
New Recommendation
Show Map
Search attractions
Albertina
PALACE
ART GALLERY
The Albertina is a museum in the Innere Stadt of Vienna, Austria. It houses one of the largest and
Schönbrunn Zoo
ZOO
IMPERIAL VIENNA
Tiergarten Schönbrunn, or "Vienna Zoo", is a zoo located on the grounds of the famous Schönbrunn Palace...
St. Stephen's Cathedral
CATHEDRAL
GHOTIC
ROMANESQUE
ANTON
PILGRAM
Vienna State Opera
OPERA HOUSE
MUSIC VIENNA
NEO
RENAISSANCE
The Vienna State Opera is an opera house – and opera company – with a history dating back to the mid-19th...
Bellevuewiese
MOUNTAIN
Naschmarkt
MARKET
SITE
The Naschmarkt is Vienna's most popular market. Located at the Wienzeile over the Wien River, it is...
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
3
4
5
6
7
8
9
10
Share
Leaflet | Map data © OpenStreetMap contributors

Figure 6.3 *Detailed description of a tourism object.*

Tourist factor	Description
Sight Seeker	You are keen to visit city highlights and most important landmarks.
Cultural Visitor	You love everything cultural - theatres, museums as well as archaeological sites.
Nature Lover	You love to explore peaceful places & to immerse in the natural environment.
Avid Athlete	Whether spectator or participant, your ideal trip involves anything sports-related.
Action Seeker	Stopped looking for action? You are interested in risky, exhilarating activities & enjoying the nightlife.
Educational Buff	Acquiring new skills & knowledge are a crucial part of your trip.
Sun Worshipper	Relaxing & sunbathing in parks or recreational areas are important for you.

Table 6.1
Description of tourist factors on the landing page.

This page provides several functionalities which are described in the following. The letters refer to different elements on the page shown in (Figure 6.3).

A) The left side of the page depicts the top 10 objects that best fit the profile of the user. Each tourism object is described with a picture, a title, some keywords and a short description. The keywords represent the leaf concepts in the tourism ontology, with which the tourism object is annotated.
B) The right side of the page shows by default a map that depicts the top 10 recommended tourism objects. In addition, the tourism objects which are already selected by the user are shown as well and highlighted through a red circle. When the user hovers a circle the title of the tourism object is displayed.
C) The small heart icon provides the user the possibility to select the corresponding object as one of his/her favorites. By clicking it the respective tourism object is added to the favorites list (indicated through its picture) at the bottom of the page and inserted at the next free position. Coincidentally, the small heart icon is removed at the corresponding item in the list on the left side as it is already selected. If the user has selected 10 tourism objects, clicking on the heart icons triggers no action as the list with the favorites is full. He/she first has to delete one of the favorites before being able to mark another tourism object as favorite.
D) At the bottom all tourism objects marked as favorites are shown. In addition, he/she has the possibility to order them according to his/her preference from left to right (via drag & drop). The first tourism object is the most favorite one.

E) The plus/minus icons (more like this/less like this) allow the user to rate the corresponding tourism object. Clicking on the plus icon indicates that the user would like to see more of this kind of tourism object or any other objects that are semantically closely related or share certain attributes (e.g. the topic “Music Vienna“). If the user rates this item, its rating icons are disabled and the user cannot rate this object a further time in the current recommendation cycle. He/she first has to click on the button “New Recommendation“ in order to rate this object again (under the precondition that it is displayed another time). When clicking on the minus icon the user indicates that he/she does not like this specific object and has a weak interest in seeing other objects of this type that are semantically closely related. Objects that are directly negatively rated by the user will not be shown again in following recommendation cycles. The ratings are used to fine-tune the specific profile of the user.

F) The button “New Recommendation“ triggers a new recommendation cycle. This means that a new ranked list of tourism objects is generated by matching the (updated) user profile with the profiles of all tourism objects. The top 10 tourism objects are shown to the user. He/she can then proceed to mark objects as favorites or rate them in a positive or negative way.

G) The button “Adjust Profile“ takes the user back to the landing page where he/she can adjust the weighting of the tourist factors. For example, he/she can decrease the weighting of the factor *Sight Seeker* from 2 to 1 heart or set the weighting of the factor *Avid Athlete* from 0 to 3 hearts. By clicking the button “Start Your Tour“, his/her profile is updated and a new matchmaking is triggered which results in a new ranked list of tourism objects.

H) When the user clicks on the title or description of a tourism object, relevant details of this object are shown on the right side of the page instead of the map (see Figure 6.3). This includes information such as phone number, email, website, ticket price, transportation or opening-hours. In addition, a description and a Lonely Planet review is shown, followed by a selection of Flickr pictures related to this object.

I) By clicking the button “Show Map“, the map is displayed on the right side of the page. This functionality allows the user to activate the map view again in case he/she had a look at the detail description of one of the tourism objects.

J) This icon centers the map on the geographical position of this object.

K) By clicking this icon the user informs the system not to display this object in the following recommendation cycles again as he/she might have already seen this attraction and is not interested in this specific attraction anymore. Clicking this icon has no effect on his profile, this is in contrast to the plus/minus rating icons which trigger an update of the user profile.

L) By clicking this icon, the corresponding object is removed from the list of favorites.

M) By clicking this icon, the detail description of the corresponding object is shown on the right side of the page.

N) This icon triggers the removing of all selected favorites, resulting in an empty list of favorites.

O) This icon generates the best route between the selected tourism objects, taking into account opening hours and shortest transportation connection between the tourism objects.
P) This functionality allows the user to share his/her favorites with his/her friends on Facebook (see Appendix).
Q) Clicking this icon opens a pop-up that displays the most important guidelines on how to use this page (see Figure 6.4).
R) Clicking this link starts the online survey (see Appendix).
S) Clicking this link shows a web tour, i.e. it highlights the different icons/functions on this page with a short notice.
T) This page links to the newsletter that shows some guidelines as well (see Appendix).
U) This input fields provides a search feature with auto-complete functionality. The user can search for tourism objects by entering the title of a tourism object or part of the title. For example, inserting the string `Church` lists all objects that have the upper/lower case string `Church` in their title, e.g., *Church of St. Rupert*, *Augustinian Church* or *St. Mary on the Strand Church*.

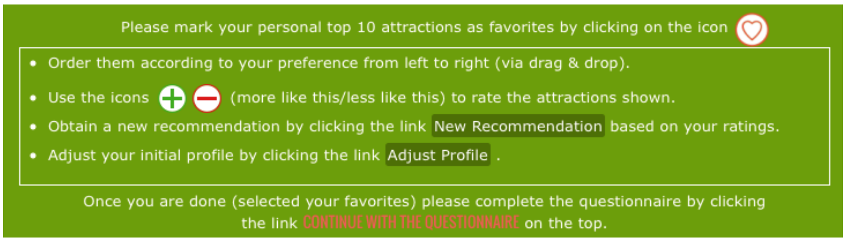


Figure 6.4
Usage guidelines.

6.2 Data sets

A total of 138 tourism objects of Vienna has been inserted in the tourism database and indexed under semantic concepts of the tourism ontology. Relevant information concerning these objects is exploited and integrated from several external data sources:

- *L*onely Planet: Following information is used for each of the tourism objects: phone number, website, ticket price, transportation, opening-hours and a review. This information is loaded over an HTTP/XML API and stored in the tourism object database. *www.lonelyplanet.com*
- *W*iener Linien: Wiener Linien is the transportation authority of the city of Vienna. It provides an HTTP/XML API to query the shortest transfer connection between two locations in Vienna and surroundings, either via public transit or walking or a combination thereof. This API is used to query the transfer connections between all combinations of all tourism objects. As this is a time-intensive task, it is carried out as a batch job, which can be repeated when new tourism objects are added into the database. *www.wienerlinien.at*
- *F*reebase: Freebase is an open, Creative Commons licensed graph database with more than 23 million entities which can be queried *www.freebase.com*

via different APIs. A major data source for Freebase is Wikipedia. Freebase is used to query the descriptions of the tourism objects from Wikipedia over a Java library.

- *F*lickr: Flickr is used to load additional pictures which are related to the tourism objects. The search is done based on tags and geolocation. When no images can be retrieved at the direct geographic location, the search radius is successively enlarged until a sufficient set of images can be found.

www.flickr.com

- *R*oute360: Route360 is a JavaScript API provided as service by the company Motion Intelligence GmbH to depict detailed routing information from source to target as polygons on a map. In Austria, the routing information is provided either as shortest walking or driving-by-car distance, whereas in Germany it already integrates public transportation information to show public transfer connections in certain cities.

https://www.route360.net

6.3 Route recommendation

In general, a route recommendation service's objective is to calculate a route between certain tourism objects that maximizes the satisfaction of a tourist, by taking into account the geographic location of the tourism objects, their opening-hours, the available time of the tourist and the travel time between the different locations. Such problems relate to the field of team orienteering problems with time windows that are explored within the domain of Operation Research (OR), which encompasses a wide range of problem-solving techniques in decision-making.

Traveling salesman problem

A well known example is the traveling salesman problem (TSP), where a traveling salesman must visit every city in his/her territory exactly once and then return home covering the shortest distance [92]. A 10-city TSP has already about 181.000 possible solutions. The TSP is an NP-hard problem, meaning that it is highly unlikely to solve it to optimality within polynomial time. Several heuristics have been developed to tackle such computational hard optimization problems such as greedy or genetic algorithms, ant colony optimization or local search [92].

Local search

Instead of searching the entire space of possible solutions (e.g., 181.000 solutions for the 10-city TSP), the *local search* algorithm picks a solution from the search space and uses an evaluation function that tells how good this solution is. The goal is to find a solution that maximizes the value of this function. Therefore, the local search algorithm applies a transformation to the current solution to generate a new solution and evaluate it again. If the new solution is better than the current solution, the current solution is discarded and replaced by the new solution. In this way, the algorithm gradually tries to improve the solution.

Iterated local search

The route recommendation service used in this prototype is based on a specific *iterated local search* algorithm which has been developed by Vansteenwegen et al. [146] to tackle the team orienteering problem with time windows in the tourism domain. Local search algorithms can stuck in a local maximum which means that the solution cannot be improved further based on the transformation function, although this solution might not be the best solution from a global perspective considering the whole solution space. Therefore, an iterated local search algorithm applies in addition a *perturbation step* to escape from the local optimum and then executes local search again based on the modified solution.

Perturbation to escape from local maximum

In the approach described by Vansteenwegen et al., each tourism object is assigned a quantitative score that expresses how good it fits the preferences of a certain tourist. In this way, the list of tourism objects that are included in the route is not predetermined but dependent on the algorithm, i.e. its evaluation function. For example, if 100 tourism objects are available, any of these objects might be included in the trip suggestion.

Our approach

In our approach, however, all tourism objects marked as favorites should be included in the route suggestion. Therefore, our evaluation function does not consider any preference score of the tourism objects. As starting and end point of the route serves the St. Stephen's Cathedral in Vienna. The begin time is predefined with 09:00 AM, the end time with 10:00 PM.

Insertion step

The algorithm used combines an *insertion step* and a *perturbation step* to escape from local optimum. The *insertion step* tries to add one favorite tourism object after the other into the route. An evaluation function is used to define the best insertion place of the respective favorite tourism object, based on the transfer time, waiting time, visiting time and opening-hours of the object. For example, if the route consists of the objects Start-A-B-End and object C should be inserted, it is examined, whether it should be inserted between the objects Start-A, A-B, or B-End. The option that requires the least amount of time (i.e., transfer time, waiting time, duration of visit) to add this visit into the route is selected. Thereby, the insertion must not make the visit of any succeeding objects infeasible. For example, if object B closes at 5 PM and its visit is scheduled for 3 PM and takes 1 hour, its visit can be delayed by maximum 1 hour but not more.

Via this *insertion step*, one after another of the favorite tourism objects is added to the route. If they cannot be all visited within a one-day route, an additional route for the next day is automatically generated and proposed to the user. If some objects cannot be inserted in any of these two route recommendations due to limited time, the user gets a corresponding notification.

Perturbation step

After all favorite tourism objects are inserted into one or two routes, a *perturbation step* is triggered in order to escape local optimum. Thereby, one or more tourism objects will be removed in each route and the insertion step is applied again with the removed tourism objects in order to find a better solution than the current one, i.e. which maximizes the number of included tourism objects while minimizing the route duration times. These iterations continue until no better solution could be found within a certain number of iterations or a specific time limit to calculate the routes is reached. The number and position of visits to be removed is dependent on the number of iterations already executed.

An example of the route recommendation service is depicted in Figure 6.5. In addition to depicting the route on the map, the user has the possibility to download a description of the route as PDF file. However, this service was excluded from the user online-survey as the main focus of the evaluation is the matchmaking approach between tourist and tourism object profiles.

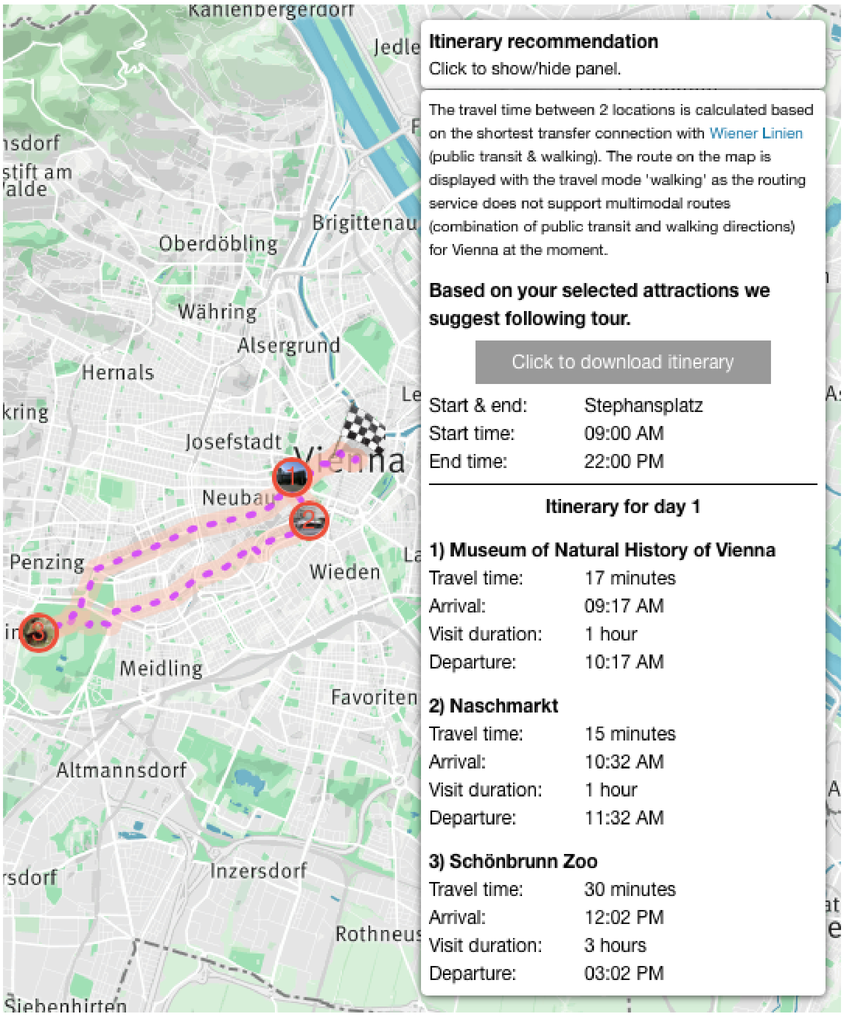


Figure 6.5
Routing service.

6.4 cDOTT ontology

The cDOTT ontology (see Section 3.4.1) has been created using the TopBraid Composer which is a commercial ontology editor of the company Topquadrant. The tourism objects have been inserted as instances in the ontology and annotated with the corresponding semantic concepts. Figure 6.6 shows as example how the tourism object *Imperial Treasury* is modeled as ontological instance within TopBraid Composer. Each instance is identified through an URI. It has a label (title) and belongs to certain semantic classes (e.g., `CityHighlight` or `CulturalHistoryMuseum`). It can be linked to further semantic concepts via specific relationships (e.g., the *Imperial Treasury* belongs to the topic `ImperialVienna`). In addition, the geographic location is stated in WGS84 format. In order to fetch further information from external data sources such as Freebase or Lonely Planet, the identifiers used by these data sources are listed as well.

www.topquadrant.com

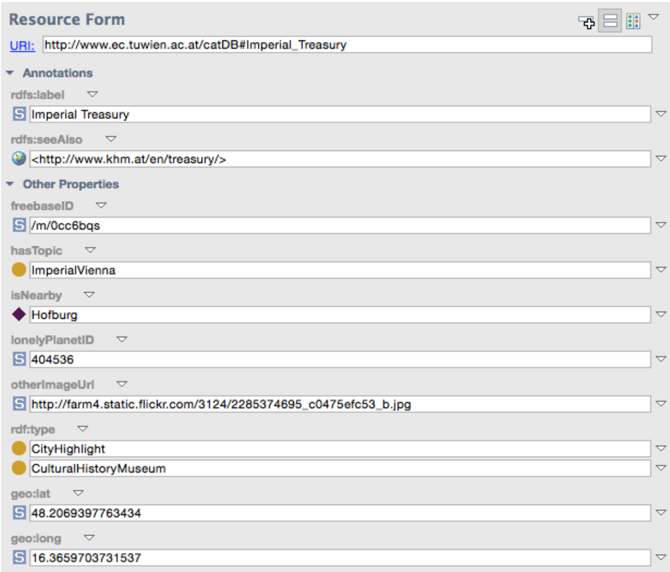


Figure 6.6 *Example of configuring an instance within TopBraid Composer.*

6.5 Customization

The prototypical application can be customized to recommend objects of another tourism destination by following these steps:

- A list of qualified tourism objects with title and geographic location has to be prepared and stored as instances in the Jena TDB store. In addition, each tourism object has to be semantically annotated with tourism concepts of the ontology.
- If the destination provides specific tourism services such as skiing or hiking, the ontology has to be extended to provide concepts that can serve as representatives for this kind of services.
- The tourist factors can be adapted by making changes in a specific database table and in a Java module that manages the factors.
- For each tourism object, its identifier in the Freebase and in the Lonely Planet data stores has to be noted in order to fetch the description from Wikipedia and information from Lonely Planet, including opening-hours, prices, transportation and reviews.
- If a route calculation service should be provided, a supplier of a service to obtain shortest transfer times between certain geographic locations in this destination has to be selected. Its service has to be integrated into the business layer in order to calculate the shortest transfer times between the different tourism objects.
- A configuration file is available that allows to adjust different metrics without having to change the source code directly, comprising the number of top-N recommendations shown, the score propagation factor or the weighting factor of the matchmaking function.

6.6 System architecture

The prototype has been developed as three-tier Java Web application, containing a front-end, business and data layer. For the front-end, Apache Wicket is used which is a component-based Web framework to ease the creation of Web sites. The business layer contains all the logic needed to fetch data from the database, load data from external data sources and provide services to the front-end to facilitate the interaction with the user. In detail, it comprises different modules for profile generation, matchmaking, interest propagation of positive and negative ratings to semantically related concepts within the ontology and route calculation. Hibernate is used as an Object Relational-Mapping (ORM) framework to handle the mapping between the object model and the relational database. Overall, the Spring framework is used to define the object dependencies and allow the injection of dependencies in Java objects. At the database layer, a PostgreSQL relational database is used to store the profiles and relevant attributes of the tourism objects. The ontology with the tourism-related concepts and the tourism objects as instances is stored in a Jena TDB triple store.

wicket.apache.org

hibernate.org

projects.spring.io/spring-framework

www.postgresql.org

jena.apache.org

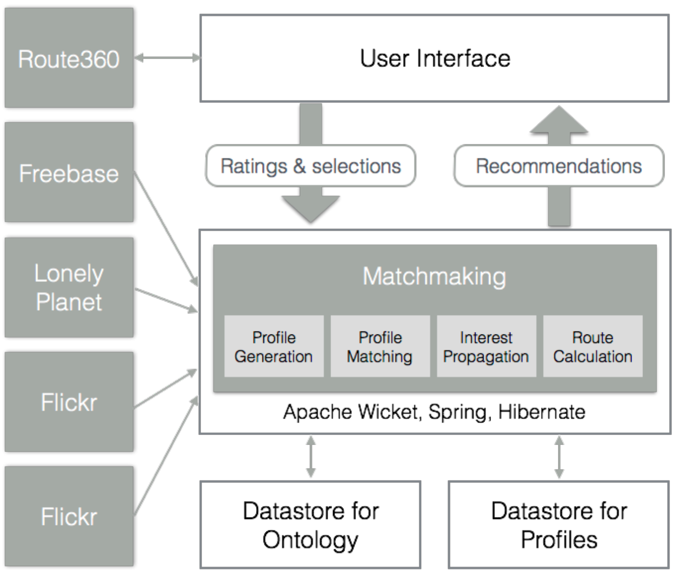


Figure 6.7
System architecture.

6.7 Summary

This chapter describes the prototypical implementation of the matchmaking process. The prototype consists of a Web application that recommends tourists, who are interested to visit Vienna, tourism objects that fit their interests. For that, 138 tourism objects are stored in a database and semantically annotated with concepts of the tourism ontology. Information regarding these objects are exploited from external data sources such as Lonely Planet, Wiener Linien, Freebase or Flickr.

7 Evaluation

This chapter presents the evaluation of the proposed recommendation algorithm, which aims to investigate the feasibility of the approach. Thus, the focus is on the user experience including the interactions of the users with the system and their overall satisfaction with the provided recommendations. We published the prototype in the Web and conducted a user study by asking users to interact with the system and fill in a questionnaire afterwards (see Appendix). In the following section the setup of the experiment and data used for the evaluation is described. In (Section 7.2) the results of the evaluation is presented, including a discussion regarding the relevance of the recommendations, the impact of using a semantic overlay model to facilitate score propagation based on user ratings and the influence of diversification. (Section 7.3) summarizes the responses of the participants to the questionnaire.

7.1 Experimental setup & dataset

The evaluation is based on a dataset extracted from weblog information of the recommendation system which had been available on the Web over a period of 3 months from June to September 2015. The target group consisted of tourists who had already visited Vienna or who plan to visit Vienna in near future or persons who know Vienna very well because they work or live in this city. In order to find a suitable range of subjects for the evaluation, postings had been placed on the homepage of the institute as well as in various Facebook groups (e.g., Foreigners in Vienna). In addition, a newsletter had been sent to members of the International Federation for IT and Travel & Tourism (IFITT) group as well as to students of the TU Wien PhD school. Furthermore, colleagues had been asked to distribute an invitation for participation via their personal social networks.

In total, 232 distinct user sessions were identified. In 137 sessions users interacted with the system by adding at least 1 recommended tourism object to their list of favorite items. In 70 sessions, users started to fill out the questionnaire. Data from 16 sessions had to be removed as the users had not completed the questionnaire or spent too few time on the questionnaire (only some seconds) which in our opinion is not sufficient to provide deliberate and well-considered answers. The questionnaire was fully completed by 54 users. These 54 users were used as the final dataset, which comprised 26 female and 28 male users. (Figure 7.1) shows the age distribution of these participants. Most users fall in the group between 25-34 years, no subject aged 50 years and over participated in the survey.

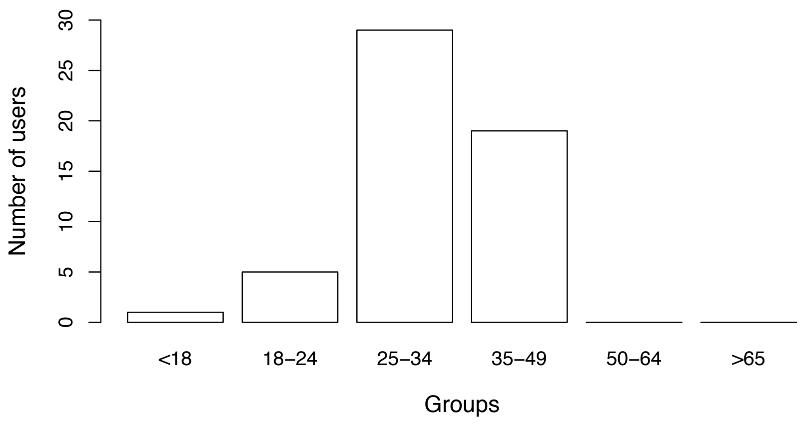


Figure 7.1
Age distribution of users within final dataset (n=54).

(Figure 7.2) shows the home countries of the users (n=53). One user refused to state his/her home country. Most of the users come from Austria (n=19), followed by Germany (n=5) and Croatia (n=3).

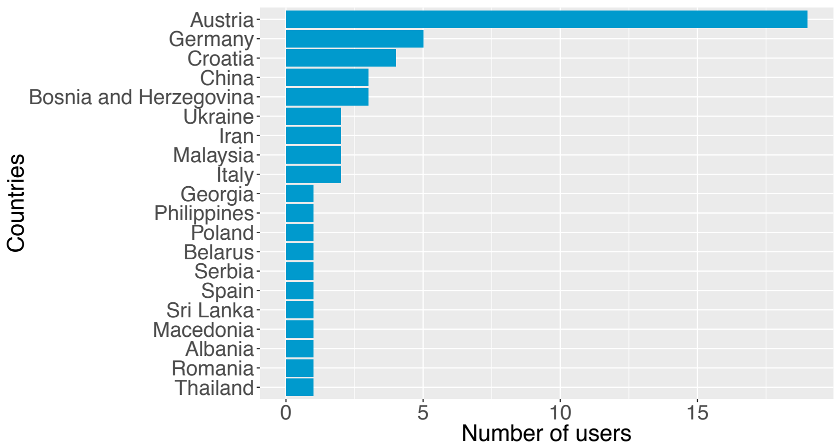


Figure 7.2
Home countries of the users (n=53).

7.2 Experimental results

First, we analyzed the selection patterns of the seven tourist factors in order to explore whether the users typically choose only a subset of the factors or rather identify themselves with a mixture of factors.

(Figure 7.3) shows how many factors were selected by the users. 32 users selected a mixture of all seven factors and about 75% of the users (i.e. 41 users) chose at least five factors. This indicates that users tend to select more than one tourist factor if they have the choice. On average, the users selected 5.7 factors to describe their predisposition and generic preferences.

(Figure 7.4) shows the distribution of the tourist factors on average. It seems that the users mostly identified themselves with the factors *Sight Seeker*, *Nature Lover* and *Cultural Visitor* whereas the factors *Avid Athlete*, *Action Seeker* and *Sun Worshipper* were less frequently selected. A reason might be the fact that Vienna is a city destination and therefore not so much associated with activities related to sports, action or sun bathing.

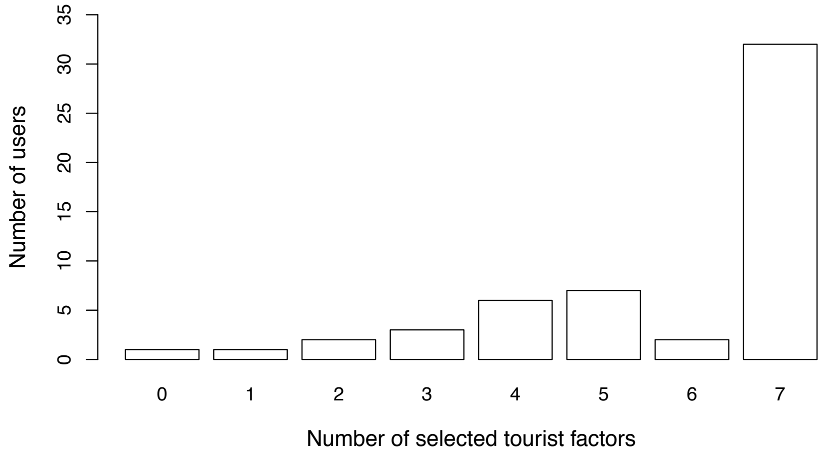


Figure 7.3
Number of tourist factors selected by the users (n=54).

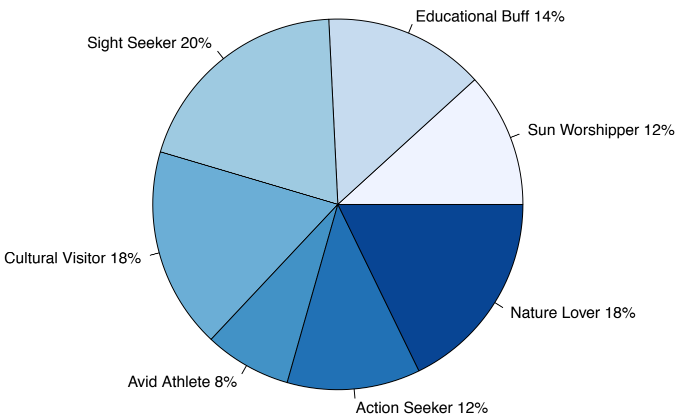


Figure 7.4
Distribution of the tourist factors on average.

We then looked at the number of tourism objects that had been marked as favorites by the users. At best, they should had selected 10 tourism objects during the recommendation process before proceeding to the questionnaire (see Appendix). As depicted in (Figure 7.5) a minimum of 2 tourism objects had been selected by all users, but about 25% had not marked 10 tourism objects as favorites.

In the next step, the number of recommendation cycles that had been produced by the users was investigated, i.e. how often each user requested to get a new set of recommended items from the system based on his/her feedback. (Figure 7.6) depicts the distribution of the recommendation cycles. About 70% of the users explored the recommendations proposed by the system within 1 to 6 recommendation cycles. The arithmetic mean is 5.5, the median is 3.

(Figure 7.7) shows at which recommendation cycles tourism objects were added or removed from the list of favorites. In total, 519 objects had been added to the list of favorites by the 54 users, i.e. on average 9.6 objects per user. About 75% of the objects had been added in the first 3 recommendation cycles. In contrast, only 37 objects had been removed again from the list of favorites by the users.

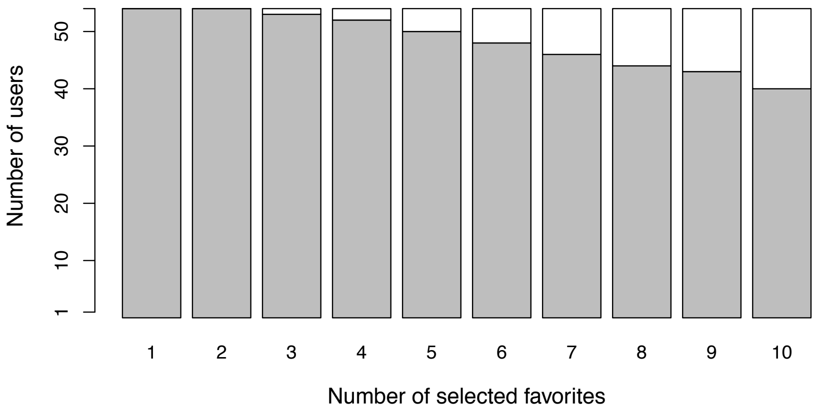


Figure 7.5
Number of selected favorites by the users (n=54).

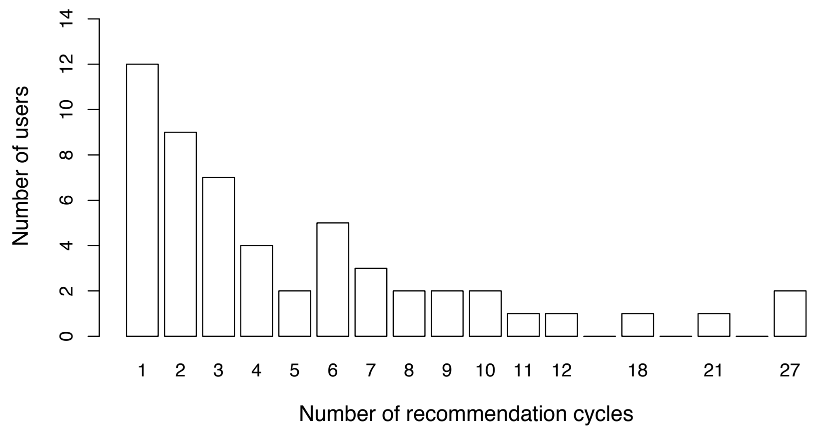


Figure 7.6
Number of recommendation cycles executed by the users (n=54).

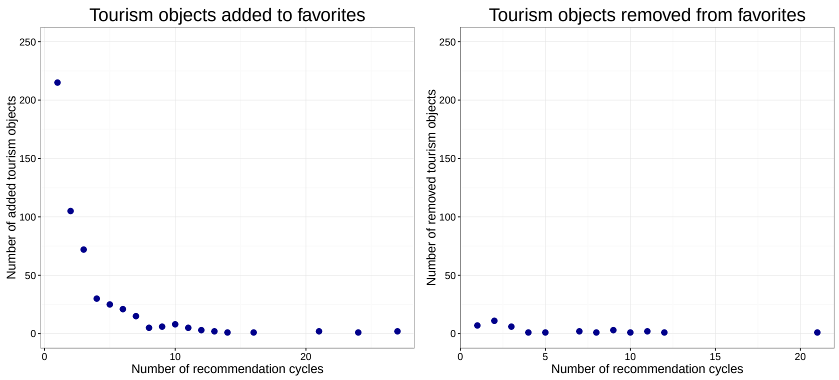


Figure 7.7
Distribution of added and removed tourism objects.

(Figure 7.8) shows the distribution of the positive and negative ratings over the recommendation cycles. On average, 8.48 ratings had been given by each user, whereby about 40% of the ratings were stated in the first recommendation cycle. In total, the 54 users produced 286 positive ratings (i.e. about 5.3 ratings per user on average) and 172 negative ratings (i.e. about 3.2 ratings per user on average).

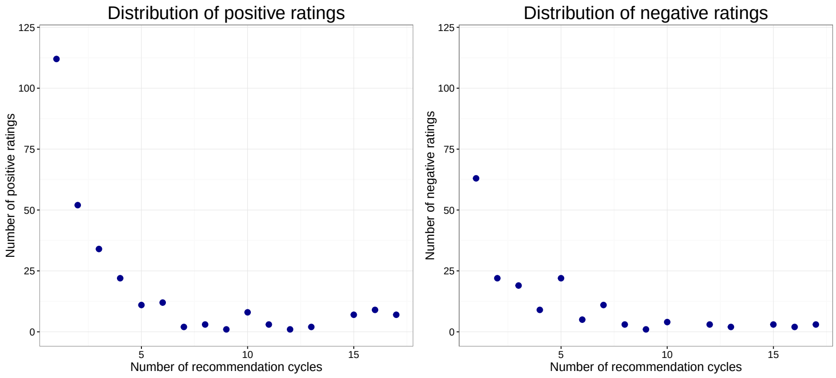


Figure 7.8 *Distribution of positive and negative ratings.*

(Table 7.1) shows the top 10 selected tourism objects based on the evaluation compared to the top 10 tourism objects based on the official visiting numbers. The *Prater*, *St. Stephen's Cathedral*, *Schönbrunn Zoo* and *Albertina* belong to the most liked tourism objects in our dataset as it is the case in the official ranked list. The *Danube Tower* ranks on place 11th in the official list. According to our evaluation, the *Danube Island*, *Volksgarten*, *Türkenschanzpark* and *Vienna State Opera* are also favorite objects. However, for these objects no official visiting numbers are stated.

Table 7.1 *Top 10 selected tourism objects.*

Top 10 selected tourism objects based on our evaluation	**Top 10 visited tourism objects in Vienna**[1]
Prater	St. Stephen's Cathedral
St. Stephen's Cathedral	Schönbrunn Palace
Danube Island	Schönbrunn Zoo
Volksgarten	Belvedere
Danube Tower	Cultural History Museum
Vienna State Opera	Imperial Palace
Schönbrunn Zoo	Museum of Natural History
Türkenschanzpark	Albertina
Albertina	Prater (Big Wheel)
Technical Museum Vienna	Aqua Terra Zoo

[1] http://www.wienkultur.info/wien_besucherzahlen.html

7.2.1 Linking the tourism objects to appropriate tourist factors

In (Section 3.4) we have proposed a method to semi-automatically link tourism objects to the seven tourist factors. First, domain experts mark for each of the tourist factors a small sample of typical tourism objects that are closely related to these factors whereby the weight of the linkage is specified through a numerical score. Second, a similarity metric is applied in order to propagate the scores given by the domain experts to the remaining tourism objects by exploiting the semantic relations between the objects.

We wanted to assess whether the linkage of the tourism objects to the tourist factors derived in this semi-automatic way is confirmed by the participants. In the questionnaire (see Appendix) the users were thus given up to five tourism objects and asked to quantify how much each of these objects matches the seven factors based on their personal opinion. The objects presented to the users were randomly chosen from the list of favorite objects they had selected before during the recommendation process. In total, 79 objects had been evaluated within 50 user sessions. On average, 4.7 objects had been evaluated during one session.

(Table 7.2) lists the objects that were evaluated by more than 5 users.

Table 7.2
Tourism objects which received scores with respect to the tourist factors from more than 5 users.

Tourism object	**No. of user ratings**
Prater	20
Danube Island	14
Volksgarten	12
St. Stephen's Cathedral	11
Danube Tower	9
Schönbrunn Zoo	9
Technical Museum Vienna	7
Türkenschanzpark	7
Vienna State Opera	6

(Figure 7.9) exemplary depicts six tourism objects and outlines the average ratings from the different users with respect to the seven factors. Based on the users' personal opinion, the *St. Stephen's Cathedral* and the *State Opera* are highly relevant for the factors *Cultural Visitor*, *Sight Seeker* and *Educational Buff*, which highly correlates to the values we have defined for these objects in the system: a value of 0.96 for the *St. Stephen's Cathedral* and a value of 0.85 for the *State Opera*. The correlation values between the users' evaluation and our quantifications for the *Türkenschanzpark* and *Volksgarten* are 0.81 and 0.56 respectively. We have defined that these objects are relevant for the factors *Nature Lover* and *Sun Worshipper*, which conforms to the opinions of the users. However, according to the users these objects are quite relevant for the other tourist factors as well. In our case, we have set those values at a lower level. The same is true for the objects *Prater* and *Danube Island*, where the users stated that they fit for nearly every tourist factor.

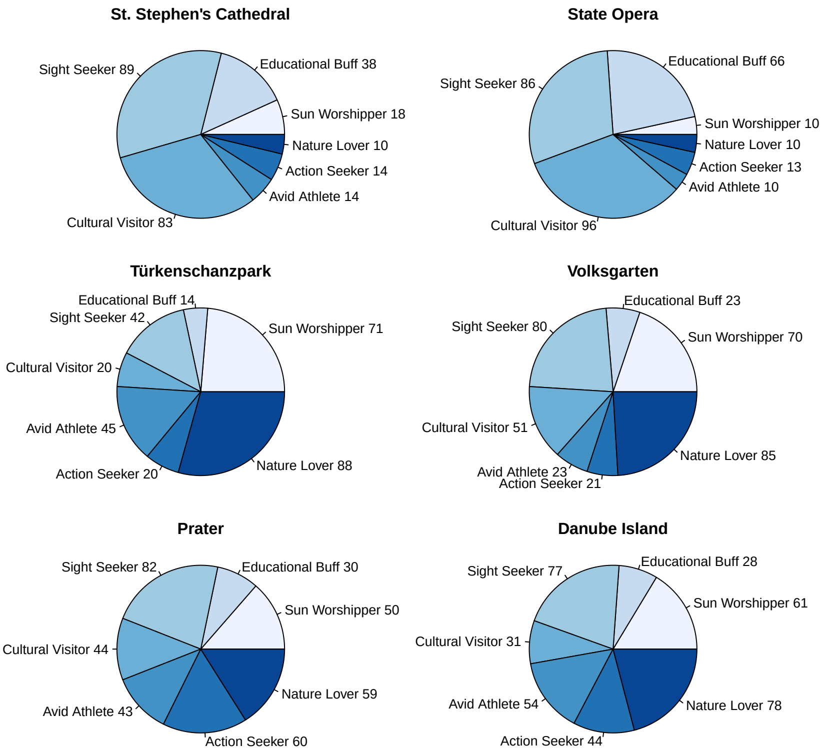


Figure 7.9
Users' average quantification values of the selected tourism objects w.r.t. the seven factors [0=no match between tourism object and given tourist factor; 100=complete match between object and factor].

It seems that certain objects can be clearly assigned to a few tourist factors while others might be relevant for a larger number of tourist factors. This might be especially the case for outdoor attractions that offer a great variety of different activities to tourists. In order to get a better understanding of the relevance of tourism objects for defined tourist factors one would need to conduct a separate study with tourism experts and a larger number of tourists to explore the attributes of tourism objects as well as the activities they offer and how they affect the mapping to specific tourist factors. This could be done on a generic level but has to be adapted to the destinations that comprise these tourism objects.

7.2.2 Relevance of the recommendations

Precision and *recall* are well known measures to evaluate the performance of recommendation systems [67]. They are utilized to assess how relevant a set of ranked recommendations is for a given user. *Precision* is defined as the ratio of relevant items selected by the user to all items selected and therefore represents the probability that a selected item is relevant. *Recall*, on the other hand, is defined as the ratio of relevant items selected to the total number of relevant items that are available. Hence, it represents the probability that a relevant item will be selected [67].

Precision and recall are complementary measures and therefore reported in pairs. For example, it would easily be possible to achieve a 100% recall by returning all items from the dataset to the user. But in this case precision would be obviously very low. Precision and recall are typically calculated by taking a set of user ratings which are divided into a training and test set. The training set is used to train the recommendation system whereas the test set is used to predict the relevance of the top-N items from this set to the given user. Precision is then calculated as the proportion of items out of the retrieved set that are relevant to the user, and recall is the percentage of known relevant items from the test set that are included in the retrieved set.

Many recommendation algorithms use existing datasets of user ratings (e.g., the MovieLense dataset) in order to evaluate their performance. In our case we did not have an existing set of user ratings that we could use as the ground truth to evaluate the performance of the algorithm. Establishing such a set would mean to present all available tourism objects listed in our database to a large group of users and let them determine which objects are relevant for them. Another approach would be to do this on an ad-hoc basis by letting the participants select their favorite tourism objects based on the recommendations given by the system in a first step and afterwards let them view all available objects from the database to identify those objects that are relevant but have not been shown to them in the top-N list. However, this approach would be very cumbersome for the participants as it takes much time to browse through the whole set of tourism objects to identify relevant but missed objects. Therefore, we decided to ease the burden of the users by limiting the number of objects to 10 tourism objects that were randomly chosen from the whole dataset and had not been recommended before. We presented this set to the users during the *evaluation* and let them decide if one or more items were relevant for them (see questionnaire in the Appendix). In this case, they had to add them to their list of favorites. (When they had already selected the maximum number of favorites, they first had to remove one or more tourism objects from their favorites that they did not like that much.) We assume that the fewer randomly shown objects were added to the favorites list, the more satisfied the users were with the initial recommendations given by the system and selected favorites.

http://grouplens.org/datasets/movielens/

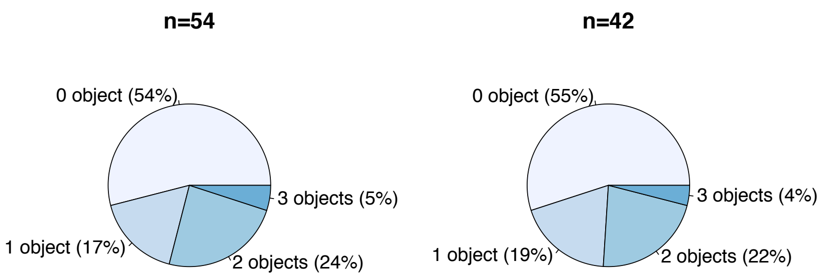


Figure 7.10 *Feedback from the users to the randomly shown tourism objects.*

The left part of (Figure 7.10) presents the feedback of the 54 users to the randomly shown tourism objects. About 54% of the users did not add any of the randomly shown tourism objects to their favorites, which could be an indication that they were quite satisfied with the recommendations and selected favorites. 17% added 1 object, 24% added 2 objects and 5% added 3 objects. No one added more than 3 objects. However, this distribution contains 12 users who requested just a single recommendation round so that the system could not use any feedback to fine-tune their preferences and deliver more relevant objects. Truncating this set of users results in the pie chart depicted on the right part of (Figure 7.10). It summarizes the feedback of users who requested more than one recommendation round. About 45% users added one or more of the randomly shown objects to their favorites. Why had these objects not been recommended to the users beforehand? The analyzation of the web log data yields some explanations:

- User did not select specific tourist factor. In some cases, the randomly shown object added by the user is associated with a tourist factor that was not selected by the user so this object did not match his/her profile. For example, the object *Pedal Power* is associated with the factor *Avid Athlete*. This object was added by a user who had not selected this factor beforehand.
- Objects of same category already recommended. In some cases, objects of the same category as the randomly shown object were already recommended to the user. For example, certain kind of museums, e.g. the *Liechtenstein Museum*, *Cultural History Museum*, *Leopold Museum* or *Albertina* were already shown to a certain user. One of the randomly shown objects was the museum *Secession* which is semantically closely related to the other museums. This object was added in addition by the user.
- Objects belonging to other categories were positively rated. In some cases, objects belonging to categories different to that of the randomly shown objects were positively rated. The ratings affected the user profile by aligning it more with the profiles of the positively rated objects. This resulted in a larger distance between the user profile and the profile of the randomly shown object and hence, this object was not included in the top-N list. The question why the user nevertheless selected the randomly presented object cannot be clearly answered without qualitative user feedback. An explanation would be that the user might have preferences towards certain objects but is also open to other objects as well. A separate study would be needed to explore this behavior in more detail.

7.2.3 The impact of score propagation using a semantic overlay model

In (Chapter 4) we have outlined the advantage of using a semantic overlay model to facilitate the propagation of user interests between child and parent concepts based on user ratings. The semantic relations within the ontology can be exploited to predict the interests of concepts that have not been yet rated by the given user.

In the following, we would like to demonstrate through a scenario how the propagation of user interests affects the position of relevant objects in the ranked list of recommendations. For this scenario, the second matchmaking process described in (Chapter 4) serves as basis as this matchmaking is used to match the high-dimensional vector of the user profile with the high-dimensional vectors of the tourism objects.

(Figure 7.11) and (Figure 7.12) depict the position of individual tourism objects (visualized in form of blue dots) within the ranked list after the user has rated the *St. Stephen's Cathedral* positively.

In (Figure 7.11), however, the propagation of user interests within the semantic model is *turned off*, i.e. only the concepts with which the *St. Stephen's Cathedral* is directly annotated, receive a certain interest score but not their parent concepts. The *St. Stephen's Cathedral*

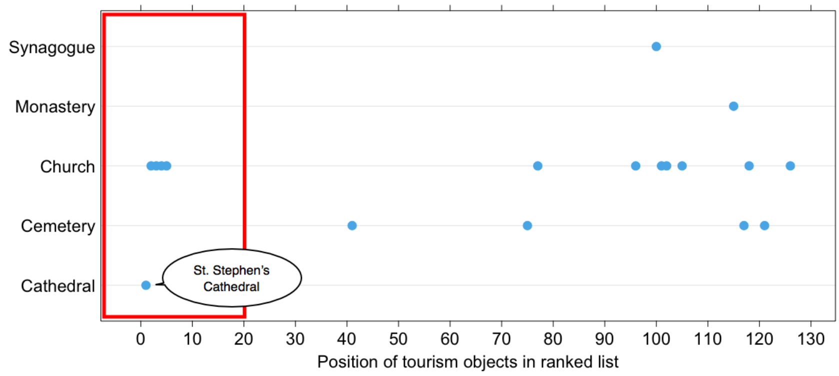


Figure 7.11 *Position of tourism objects related to 'religious architecture' in ranked list without score propagation.*

is annotated with the concepts `cathedral`, and `romanesque` and `gothic architecture`. Thus, these concepts receive a certain interest score based on the positive rating given by the user. As the *St. Stephen's Cathedral* is rated positively it perfectly matches the user's interests and is placed on the first position. On the next positions are four churches as they belong to the *gothique* or *romanesque architectural style* as well. However, other objects which belong to the category *religious architecture* such as monasteries, cemeteries or further churches are not on the top-20 positions of the ranked list. These objects might be of relevance to the user as well but would not be included in the top-N list.

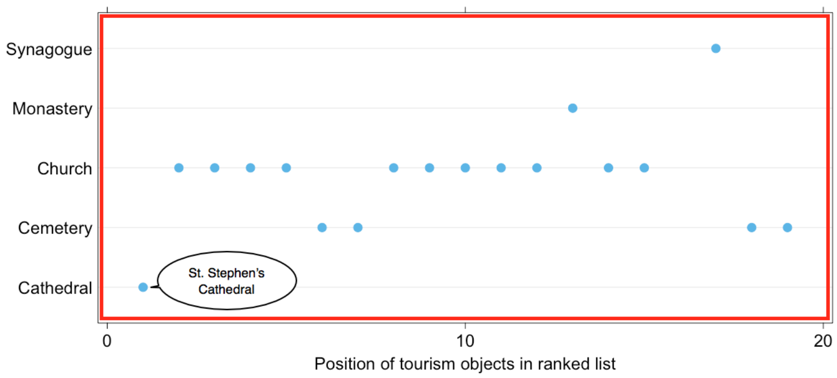


Figure 7.12 *Position of tourism objects related to 'religious architecture' in ranked list with score propagation.*

(Figure 7.12) depicts the ranking of the objects related to *religious architecture* when the propagation of user interests within the semantic model is *turned on*. The *St. Stephen's Cathedral* is again placed on the first position. But in contrast to (Figure 7.11), all the other objects of the category *religious architecture* are now directly placed on subsequent positions within the list. The reason for this is that the user interests scores are now propagated from the leaf concepts to their parents and further up in the ontological hierarchy. The user profile gets more similar to the profiles of other religious tourism objects as the profiles have a larger overlap in concepts related to religion and therefore, these objects are ranked on top-20 positions of the list. Thus, the recommendations are better aligned to the category *religious architecture*.

This scenario shows that score propagation within the ontology is highly useful to predict the interests of concepts that have not been yet rated by the given user and to propose similar objects in the next step.

7.2.4 The influence of diversification

In this section we discuss how the diversification method defined in (Section 5.4) influences the diversity of the recommendations provided by the system. The initial list of recommendations computed by the system includes tourism objects which are ranked according to their adequacy to the user, which is based on the similarity between the user profile and the individual profiles of the tourism objects. When the diversification method is applied, the final list of the objects returned to the user will not contain the *first* top-N elements of the list, but rather a subset of all elements of the list. Hence, the variety of the elements is increased but at the same time their accuracy with respect to the user profile decreases. This way, relevance is traded for variety, which nevertheless has a positive effect on the user satisfaction, assuming that tourists are interested to see various quite different tourism objects at the beginning of the recommendation process.

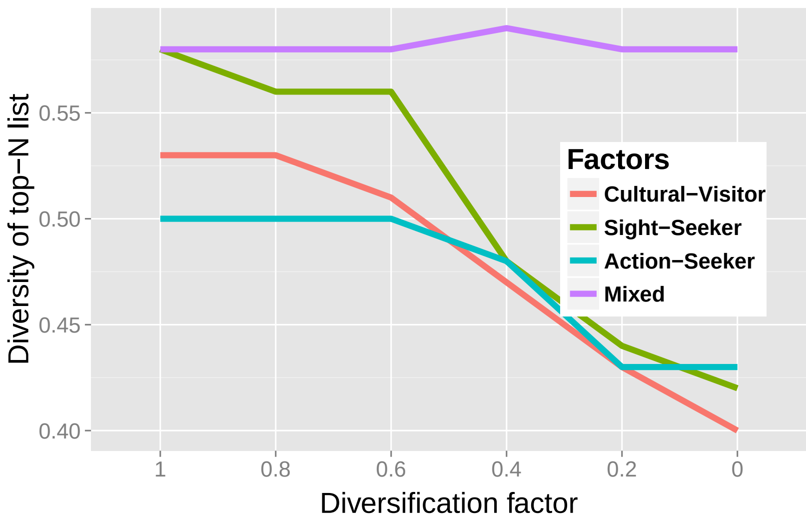


Figure 7.13 *Diversity of top-N list with different levels of the diversification factor.*

(Figure 7.13) shows the diversity of the top-N list depending on the given level of the diversification factor α, which ranges from 1 to 0. A large diversification factor produces a top-N list that favors diversification and outputs a rather different result set to the original ranked list of objects whereas a factor near to 0 has low effects on the original list.

The diversity of the list returned to the user is expressed as the average distance of each tourism object in the top-N list to its centroid. The centroid of the top-N list is defined as the mean vector of all tourism object profile vectors in the top-N list. The distance between each profile vector and the centroid is computed based on the Euclidean distance.

Finally, the diversity of the top-N list is calculated by taking the average of all distances.

In this Figure, the diversification curves are plotted for specific profiles such as the *Cultural Visitor*, *Sight Seeker* and *Action Seeker* in order to represent tourists who are only interested in a single tourist factor. In addition, a mixed profile is depicted which represents tourists who describe their preferences over all factors with equal shares. The curves of the profiles *Cultural Visitor*, *Sight Seeker* and *Action Seeker* are downward sloping, which means that the diversity of the top-N objects decreases when using a lower value for the diversification factor α. In other words, the produced list of top-N objects is closer to the original ranked list. When looking at these curves together, the greatest effects regarding the decrease of the diversity seem to result from using diversification factor values between 0.6 and 0.2.

The curve of the mixed profile is different from the other curves. The reason for this is that when using a mixed profile the top-N objects of the original ranked list of tourism objects are already rather diverse as tourism objects related to all factors are included in the top places, thus producing a rather diverse set even when diversification is not considered (i.e. the diversification factor α set to 0). The results indicate that applying diversification on the original ranked list of objects is especially relevant for users who rather describe their preferences based on a single factor and has lower impact on users with a mixed profile.

In the following, two scenarios describe the behavior of the system with different settings of the diversification factor.

First scenario - Cultural Visitor with interest in museums.

The first scenario describes a user who is of type *Cultural Visitor*.

As shown in (Table 7.3), he or she might receive a large amount of cultural history museums when using no diversification (the diversity factor α set to 0) although he/she might be interested in a variety of different types of museums. When setting the value of the diversification factor α to 1 the list of returned objects is rather diverse, containing not more than 1 object for each type of museum. We assume that users rather prefer such a diverse list at the beginning of a recommendation session, when they declared interest in cultural objects but have not expressed their detailed interests so that they can evaluate and comment on the proposed recommendations more efficiently. However, when a user has declared interest in specific concepts (e.g. concept `music museum`) by positively rating corresponding tourism objects, the diversity of the top-N list should decrease in order to acknowledge the fact that the user might be interested in objects which are related to the positively rated concepts. This seems to be supported by the responses of the users to the questionnaire (see Section 7.3).

Second scenario - Independent Visitor with interest in churches.

The second scenario describes a user of type *Independent Visitor*, who has repeatedly declared interest in objects of type `church`. (Table 7.4) presents the corresponding top-N list when applying diversification (upper part of the table) and without diversification (lower part of the table). When the degree of interest in certain concepts gets higher, the diversity factor should decrease so that the top-N objects reflect the specific interests of the user.

(Figure 7.14) depicts the precision and diversification curves for the scenario stated above. In order to compute the precision, it is assumed

Tourism Object	Type
Diversification factor: 0 (original ranked list)	
Cultural History Museum Vienna	Cultural History Museum
Burial Museum Vienna	Burial Museum
Carnuntum	Cultural History Museum
Roman Ruins	Cultural History Museum
Imperial Furniture Collection	Cultural History Museum
Wagenburg	Cultural History Museum
Clock Museum	Museum
Kunsthalle	Contemporary Art
Military Museum Vienna	Military Museum
Museum of Applied Arts Vienna	Applied Art Museum
Diversification factor: 1	
Cultural History Museum Vienna	Cultural History Museum
Haydn House	Music Museum
Jewish Museum Vienna	Jewish Museum
Leopold Museum	Art Gallery
Museum of Applied Arts Vienna	Applied Art Museum
Kunsthalle	Contemporary Art
Military Museum Vienna	Military Museum
Museum of Modern Art	Modern Art Museum
Museum of Ethnology	Ethnology Museum
Theatre Museum Vienna	Theatre Museum

Table 7.3 *Top-N tourism objects for a Cultural Visitor based on different levels of the diversity factor α.*

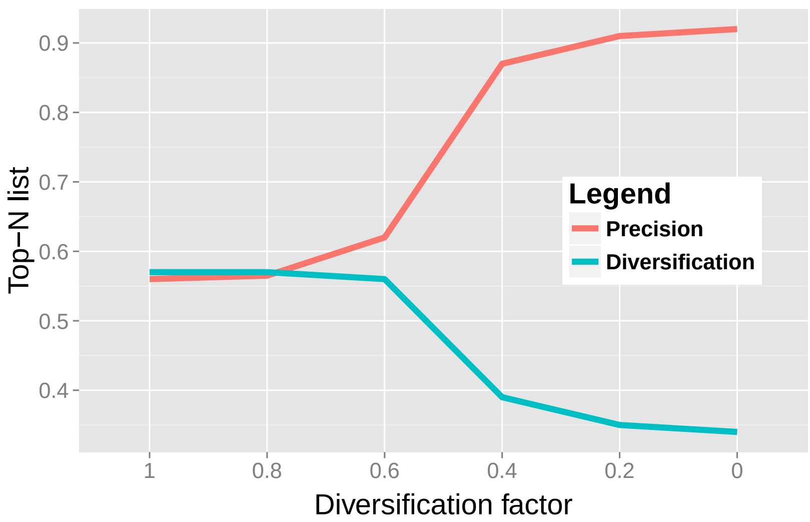


Figure 7.14 *Diversity and precision of a top-N list with different levels of the diversity factor.*

that the recommendation process is correct and returns a ranked list of objects with a 100% accuracy, i.e. its objects are sorted according to their relevance to the user. Precision is defined as the relevance of the tourism objects for the user by computing the similarity between the profiles of the top-N objects and the user profile. Diversification is computed as the

Tourism Object	**Type**
Diversification factor: 1	
St. Peter Church	Church
St. Stephen Cathedral	Cathedral
Stadttempel	Synagogue
Stadtbahn Pavillions	Civil Architecture
Imperial Crypt	Cemetery
City Hall	Civil Architecture
Leopoldsberg	Mountain
Technisches Museum Vienna	Technical Museum
Palais Schwarzenberg	Palace
Museum of Natural History of Vienna	Natural History Museum
Diversification factor: 0 (original ranked list)	
St. Peter Church	Church
St. Charles Church	Church
Scottish Church	Church
Jesuit Church	Church
Dominican Church	Church
Augustinian Church	Church
Capuchin Church	Church
St. Michael Church	Church
Votive Church	Church
St. Mary on the Strand	Church

Table 7.4 *Top-N tourism objects for an Independent Visitor with interest in objects of type church based on different levels of the diversity factor α.*

average distance of each tourism object in the top-N list to its centroid. The precision increases with a lower value of the diversification factor and outputs a top-N list which gradually includes more and more objects of type `church`.

Thus, the degree of diversity that the system should use depends on the expressed level of interest in certain concepts of the tourism ontology through positively rating certain tourism objects. The higher the interest in a certain concept is, the lower the diversification factor α should be so that precision has a greater emphasis on the computation of the top-N list than diversification.

Therefore, the value of α should be set dynamically and we have decided to use a linear function that depends on the maximum number of positive rates the user has implicitly assigned to a concept within the ontology. Based on the data shown in (Figure 7.13) we decided to set the diversification factor α at the beginning of a new user session to 0.6. When the user expresses positive interest in a certain concept, the value decreases and when the number of positive rates of a concept is greater than a constant factor c, α is set to 0 which means that diversification is not anymore considered and the original ranked list of top-N objects is returned. We propose to set the constant factor c to a value of 3, i.e. if a concept is rated 3 or more times, diversification is not applied anymore.

7.3 Participants' feedback

At the end of each questionnaire, we collected information from the participants regarding their level of satisfaction, their judgment of the application interface and the recommendation feature as well their opinion whether they find the diversification approach useful (see Figure 7.15 and Appendix for the questions in detail). The participants were asked

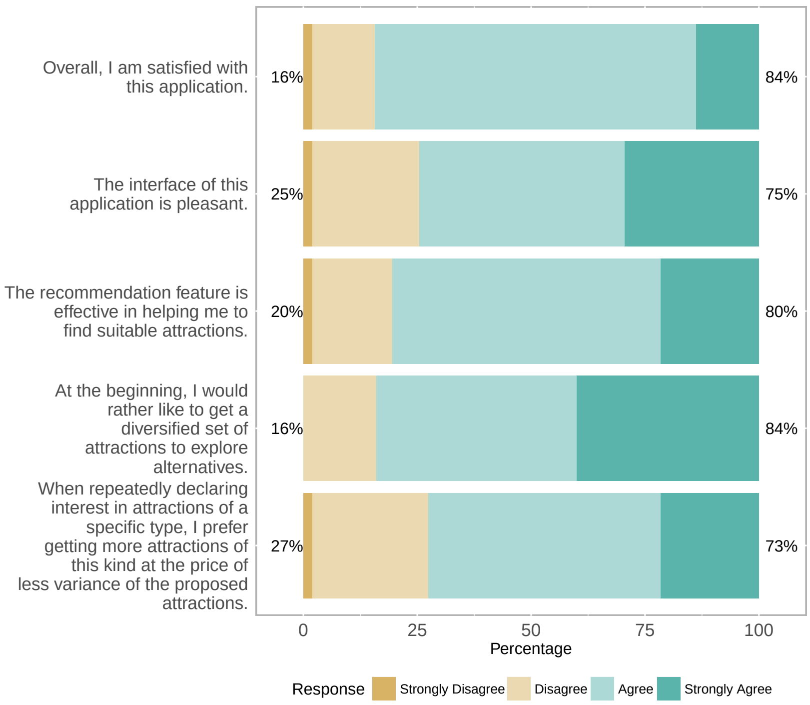


Figure 7.15 *Participants' feedback regarding their satisfaction, their judgment of the application interface and the recommendation as well as diversification feature.*

to rate on a 4-point Likert scale how much they were satisfied with the application. 84% of the participants agreed or strongly agreed to the statement "Overall, I am satisfied with this application." 75% of the participants found the interface pleasant. 80% of the participants confirmed that the recommendation feature is effective to find suitable objects.

We asked the question whether they would rather like to get a diversified set of attractions at the beginning in order to explore alternatives (e.g. the set of recommended attractions should not only contain attractions of a specific type such museums). 84% of the participants agreed or strongly agreed to this question. This might be an indication that users like to get a diversified set of recommendations at the beginning of the recommendation session when no further specific preferences are known to the system. 73% agreed to the follow-up question "When repeatedly declaring interest in attractions of a specific type (e.g. churches) I prefer getting more attractions of this kind at the price of less variance of the proposed attractions.". This question is closely linked to the former one - when the user expresses interest in specific objects it seems that he/she prefers to get results tailored to his/her specific preferences in favor of a rather diversified result set.

We further asked the participants which improvements they would suggest for the application. Some users were concerned with the current size of the favorites list (i.e. 10 objects) and argued that the size should be made configurable so that they can mark more attractions as favorites when they plan a longer trip:

> "The number of recommended attractions should be made configurable. Depending on how long I stay I would be interested in different number of attractions."
> "Maybe there could be a way to add more to the 10 attractions that i have chosen."

Some users pointed out that exploring new tourism objects they are not aware of or trying out different things are important for them. They would prefer to get the possibility to browse through all tourism objects that are available in the system or get a few recommendations that do not closely match their stated preferences so that they can discover new tourism objects. In addition, they would like to receive context-aware recommendations, e.g. to explore tourism objects in the near surroundings of their hotel. Selected quotes are:

> "I would like to see list of all attractions regardless recommender system, to see there is something new that I didn't have idea about."
> "I would prefer to get some attractions of other types even though I haven't specified it in my profile. I do not know the city, and I am not even sure of my traveling preferences, so I would like to try different things with accent on things that I know I like."
> "I'd also like to find out what's around my Hotel."

Receiving explanations why objects are recommended seems to be important for the participants so that they can understand how the recommendation algorithm works and adjust their profile accordingly. Moreover, users would like to get information which tourism objects other users (similar to them) like or have selected so that they get some new inspirations. Selected quotes are:

> "The application can show the reason of each recommended place, like discovery feature Spotify. For example, Prater was recommended due to adventure that you have selected. This way I can tune my profile much better."
> "I would like to see what other people selected, also visited and liked it in order to be able to select the attractions a) that I would prefer, but also b) that are recommended by others like me to visit."

At the moment, the system provides a routing service which calculates a route between all favorite tourism objects by considering constraints such as opening-hours, travel time and available time of the total tour. Some users would like to have additional customization possibilities (e.g., determine location to start or end the tour, change visit order of the objects, determine which objects to see at which day) so that they can plan their stay at the destination in a more detailed way.

> "Feature of create a custom tour and save it somewhere or print it out in a nice format. This way a user could plan a trip and sightseeing route better."

Regarding the interface, the users gave both positive and negative feedback. Some liked the design whereas others argued that they were overwhelmed by the information and that it took some time to understand the way how to interact with the system. Selected quotes are:

"The interface is catchy."
"Overall this could be very useful and handy."
"The system will benefit from an enhanced map-attraction interface. Good, intuitive interface."
"Flexible, easy to handle, fast."
"Yes, the user interface is very helpful and recommendation at the end of the application really gives me a chance to rethink my earlier option. Plus with a brief explanation, it develops my perspective on how the destination would be.
"On the second page (where you have to select the ten favorites) I didn't understand at all what to do! How to choose ten favorites from only ten recommended attractions? I didn't understand where and how to drag and drop the attractions to order the favorites (first, on the third page with the additional recommendations I discovered the icons on the bottom of the page)."
"Please, more simple and clear."
"Interacting with the top 10 recommendations was a bit tedious (I wanted read the description again and this was not straightforward because clicking on them did not work right away)."
"Not intuitive but maybe it was just the survey interface that was irritating. design-wise it is very nice."
"The UI doesn't feel natural, there are many unnecessary choices.. + - .. the best would be you like this or not.. so just the heart is sufficient i think."
"The user-interface seem unclear and a little bit crowded with information. I would rethink adding the tags (does the user really need them?)."
"It is quite complex. Especially I was not sure whether the top 10 should be ranked or not. If this was explained at one point, I missed it - maybe I was a bit overwhelmed with information."
"I couldn't figure out how to say NOT AT ALL AVID ATHLETE. I didn't even want to assign 1 heart to it. A heart means at least somewhat. But I am definitely not."

7.4 Summary

In this chapter we present the results from a user study we have conducted with 54 participants in order to evaluate our prototypical application. Overall, the users seem to be satisfied with the proposed recommendations. Most of them choose at least five tourist factors to generate a high-level profile. The most prominent factors for the city destination Vienna are Sight Seeker, Cultural Visitor and Nature Lover. They agree that diversification should be considered at the beginning of the recommendation session when no specific preferences are known to the system. Both positive and negative feedback is used in order to fine-tune their interests and deliver a new set of recommendations.

8 Conclusion

The overall goal of this thesis was to close the gap between users' needs and suppliers' perspectives by matching their respective views (see Section 1.5). We addressed this issue by developing a matchmaking process and applied this process in the e-tourism domain, thus matching tourist profiles with the characteristics of tourism objects in order to obtain a ranked list of appropriate tourism objects for a particular tourist. We implemented a Web-based prototype that proposes tourists, who would like to visit Vienna, a set of suitable tourism objects. For that, different external services were integrated in the prototype in order to deliver destination-specific information (see Section 6.2). The evaluation took place in form of a user study, where participants were asked to explore the prototype and fill in a questionnaire in the end (see Chapter 7). Although the matchmaking process was applied in the e-tourism domain, it could be adapted for other domains (e.g. retail and consumer products) as well in order to suggest items or services to the users.

In this chapter, we revisit the research questions posed in (Section 1.5) and summarize our findings, while highlighting the main contributions of this thesis. We close by pointing out issues and challenges for future work.

8.1 Answers to research questions

8.1.1 Question 1

Q 1: *Can tourist types existing in scientific tourism literature be used to obtain a high-level user profile?*

In (Chapter 3) we have demonstrated successfully a method to use the tourist factors proposed by Werthner et al. [96] (i.e., *Sun Worshipper*, *Educational Buff*, *Sight Seeker*, *Cultural Visitor*, *Avid Athlete*, *Action Seeker*, and *Nature Lover*) as basis to construct the initial profile of a tourist and to derive a quantitative representation in form of a vector. Our evaluation results (see Section 7.2) reveal that most participants of the user study identified themselves with the factors *Sight Seeker*, *Nature Lover* and *Cultural Visitor*, and that the vast majority (i.e., 75% of the participants) chose at least five factors. These results are consistent with the findings by [57], indicating that individuals rather select more than one type if they have the opportunity.

The basic profiles of the tourism objects are constructed based on the seven tourist factors as well so that they can be represented in the same vector space as the tourist profiles. However, if the profiles of the tourism objects are constructed manually by domain experts, this would put a big burden on them and would not scale if the profiles of tourism objects from many destinations worldwide need to be established.

To counteract this problem, we have proposed a method to derive the basic profiles of the tourism objects in a semi-automatic way (see Section 3.4), whereby domain experts first mark manually for each of the tourist factors a small sample of typical tourism objects that are closely related to these factors. For the second step, we have defined a semantic similarity measure that propagates the scores given by the domain experts to those tourism objects that have not yet been marked by the experts. As a prerequisite, the tourism objects have to be annotated with concepts from an ontology.

We have evaluated existing ontologies in the tourism domain and developed the ontology cDOTT (see Section 3.4.1) which is partly based on the Harmonise ontology [49] as well as on the EON Traveling ontology [43]. cDOTT is defined on a more abstract domain level, so that existing, more fine-grained tourism ontologies such as the QALL-ME [94] or the DERI OnTour ontology [39] can be easily integrated.

In our case, we have annotated the tourism objects manually with concepts from the ontology. However, different approaches exist that can automate this task in order to derive such annotations by parsing text documents which are related to the tourism objects. In [149] an approach is presented that is able to detect relevant features from textual resources and associates them with concepts modeled in an ontology. This approach has been applied in the context of tourism, whereby Wikipedia articles are parsed to detect concepts related to a tourism ontology.

In our evaluation we wanted to assess whether the linkage of the tourism objects to the tourist factors derived in this semi-automatic way is suitable to define the generic profiles of the tourism objects. Therefore, we asked the participants of the user study to quantify for a small set of given tourism objects how much each of them matches the seven factors based on their personal opinion. We compared the participants' assessments with the quantifications we derived semi-automatically. The results (see Section 7.2) indicate that tourism objects which act as representatives for a given factor (e.g., the St. Stephen's Cathedral for the Sight Seeker or the State Opera for the Cultural Visitor) can be clearly assigned to those factors. However, tourism objects that provide the tourists the possibility to perform different activities, such as the Prater or Danube Island in Vienna seem to be adequate for nearly each tourist factor according to the participants' opinions.

A separate study with tourism experts would be needed in order to specify the mapping between the tourist factors and certain tourism objects in a quantitative manner and then to run different experiments to approximate these quantifications by adjusting the semantic annotations or certain parameters in the score propagation method.

8.1.2 Question 2

Q 2: *Can user feedback be exploited to improve the matchmaking process?*

In (Section 5.2) we have demonstrated a method how to use explicit user feedback in order to obtain the specific interests of a tourist and adjust her or his profile accordingly. Users have the possibility to rate presented tourism objects either in a positive or negative way (see Section 6.1). We presented an approach to transform these explicit ratings to numeric values and update the tourist's profile accordingly. The tourist's profile is represented as vector of interest scores that are assigned to concepts within the tourism ontology. We assigned the numeric values to those leaf concepts with which the rated tourism object is annotated and used score propagation to update the super concepts as well (see Question 3).

As presented in (Chapter 7) the participants of the user study used both positive and negative ratings in order to get new recommendations that are adjusted to their interests. On average, 8.48 ratings had been given in each user session, comprising about 5.3 positive ratings and about 3.2 negative ratings. This indicates that next to positive ratings also negative ratings seem to be accepted as valid means to inform the recommendation system about likes and dislikes. In fact, negative ratings are helpful to fine-tune the profile more rapidly as this would be the case with positive ratings only.

8.1.3 Question 3

Q 3: *Can we exploit the semantic relations within a tourism ontology to infer the user's interest in objects not having been rated yet?*

In (Section 4.2) we presented an approach to model both the specific profile of tourists and tourism objects as high-dimensional vectors whereby the dimensions represent the concepts and the values the numeric scores assigned to the concepts of the ontology. The ontological structure allows to propagate the numeric scores that are assigned to the leaf concepts of the ontology based on either positive or negative feedback from a particular tourist, further up in the ontology to the super concepts. For example, if the user has positively rated the St. Stephen's Cathedral in Vienna, and we know that this cathedral is annotated with the concept `church` which has as parent the concept `religious architecture`, these concepts receive a certain interest score. Thus, the user profile vector gets more similar to vector representations of other tourism objects which are kind of religious buildings, as the overlap in concepts is greater and the vectors get more similar.

We infer an interest score for concepts by adopting a measure from literature that exploits the ontological structure of the tourism ontology. In (Section 7.2.3) we have demonstrated through a scenario how the impact of score propagation affects the position of tourism objects in the ranked list.

Without utilizing score propagation, only the leaf concepts with which the tourism objects are directly annotated receive a certain (negative or positive) interest score based on the tourist feedback, but not their parent concepts. In this case, only those tourism objects that are annotated with the same leaf concepts get more similar to the user profile as their vector representations are similar.

When score propagation is turned on, the interest scores are propagated upwards within the ontology. This has the positive effect that the user profile gets more similar to other tourism objects that share certain concepts in the ontology as the similarity between those vectors get closer.

The scenario shows that score propagation is essential to infer the user's interest in objects that have not been rated directly by her- or himself but might be highly relevant as they share certain concepts within the ontology.

8.2 Future work

In our work, we have presented a matchmaking process in order to provide personalized offers about tourism objects to tourists in the pre-trip phase of the tourist life cycle. However, as this research is bringing together methods and techniques from the fields of semantic technologies, recommendation systems, context-aware computing and others, there are still important issues remaining. This section proposes future work on the following aspects:

Automation of semantic annotations of tourism objects. Our matchmaking process exploits tourism ontologies to describe the tourism objects of a destination in a qualitative, formalized model (see Chapter 4). However, the tourism objects were annotated with semantic concepts of the tourism ontology in a manual way which does not scale if the matchmaking process is used to provide personalized offers for various destinations. Most tourism information in the Internet is still predominantly published in form of documents containing unstructured text. An important research issue would be the exploitation of this document space with the help of *text mining* methods to *automate the semantic annotation* of tourism objects by linking meaningful terms from the text descriptions to concepts within the ontology.

Delivering explanations for recommendations. In the current prototype, no *explanations* are given to the users, why certain tourism objects have been recommended to them. The only information that is provided to the user are the names of the concepts with whom the tourism objects are directly annotated, such as the concepts `Cathedral`, `ghotic`, `romanesque`, `Anton Pilgram` for the object 'St. Stephen's cathedral' (see Section 6.1, Figure 6.3). Obtaining such explanations seems to be important for the users (see participants' feedback in Section 7.3) so that they can adjust their profile accordingly. One way would be to show the profiles of the tourism objects to the users so that they can see how much the specific tourism objects are associated with the different tourist factors. Another way is to generate explanations in form of "You seem to like objects related to religious architecture. That's why we recommend you to visit the 'St. Stephen's cathedral'" so that the users understand the reason why they have obtained this recommendation [139].

Seamless support between different trip phases. In the current state, there is still a *perceptible gap* between the respective phases of the tourist life cycle, resulting in the need for tourists to use different sources to satisfy the information requirements in each phase, ranging from online travel communities, mobile applications, Internet Websites, destination portals, metasearch and booking engines to traditional guide books. Tourism information systems of the *next generation* should deliver a single point of access that provides relevant services during all phases in a seamless way and facilitates the creation of personalized trip plans. In this way, they would certainly enhance the tourists' satisfaction with the *planning process* and the *travel experience*. In the ideal case, the entire vacation is planned in space and time, which involves decisions including length of trip, primary destinations, activities, attractions, accommodations or trip routes [79].

Plan revision based on contextual information. The prototypical application of the matchmaking process provides the possibility to select up to ten favorite tourism items of a certain destination and to obtain a basic trip plan, comprising descriptions of the selected objects and a route proposal to visit the favorite objects in an optimized order. However, even the best travel plan cannot protect tourists from having to face spontaneous, unexpected situations during the trip. Attractions can be temporarily closed or open-air concerts may be canceled due to bad weather conditions. This issue tackles the ability of gathering such *contextual information* and performing *context-based reasoning*. The process of *plan revision* poses a great dilemma for tourists when they have to do this manually as they are located at an unfamiliar destination and have insufficient travel information regarding existing alternatives available on-site. Trip plans that are adapted to the dynamics of traveling can reduce the risk of mishaps while being on the move. Tourists might be able to experience a more relaxed itinerary since adaptive travel plans are capable of dynamically reschedule planned tourism objects in order to counteract new, spontaneous situations of (mobile) tourists. However, rescheduling of existing activities or including new activities into the trip plan is very complex. Rescheduling has to consider factors such as the duration of visit, the opening hours ob objects as well as user-defined priorities of activities. Moreover, activities may not overlap and might be dependent on each other (e.g., if the flight is canceled, other activities have to be canceled as well).

Appendix

The Appendix contains following sections:

- **Newsletter for participating in the online survey.** This newsletter was used to invite people to participate in the survey.
- **Facebook posting with the favorite tourism objects.** Participants had the possibility to share their favorite tourism objects with their Facebook community.
- **Online survey.** The questionnaire is shown that was used for evaluation of the prototype.

Newsletter for participating in the online survey

This newsletter was used to acquire participants for the online survey.

GO2VIENNA

What are your favorite attractions?

Have you ever visited Vienna or plan to visit in the near future? Participate in research and receive valuable tips the next time you are in Vienna.

We are writing to ask you to get involved in research that is being undertaken by the e-commerce group at the TU Vienna in Austria. Don't click me away and invest about 15 minutes of your time. Please forward me to your colleagues & friends via the social media buttons at the bottom. Thank you!

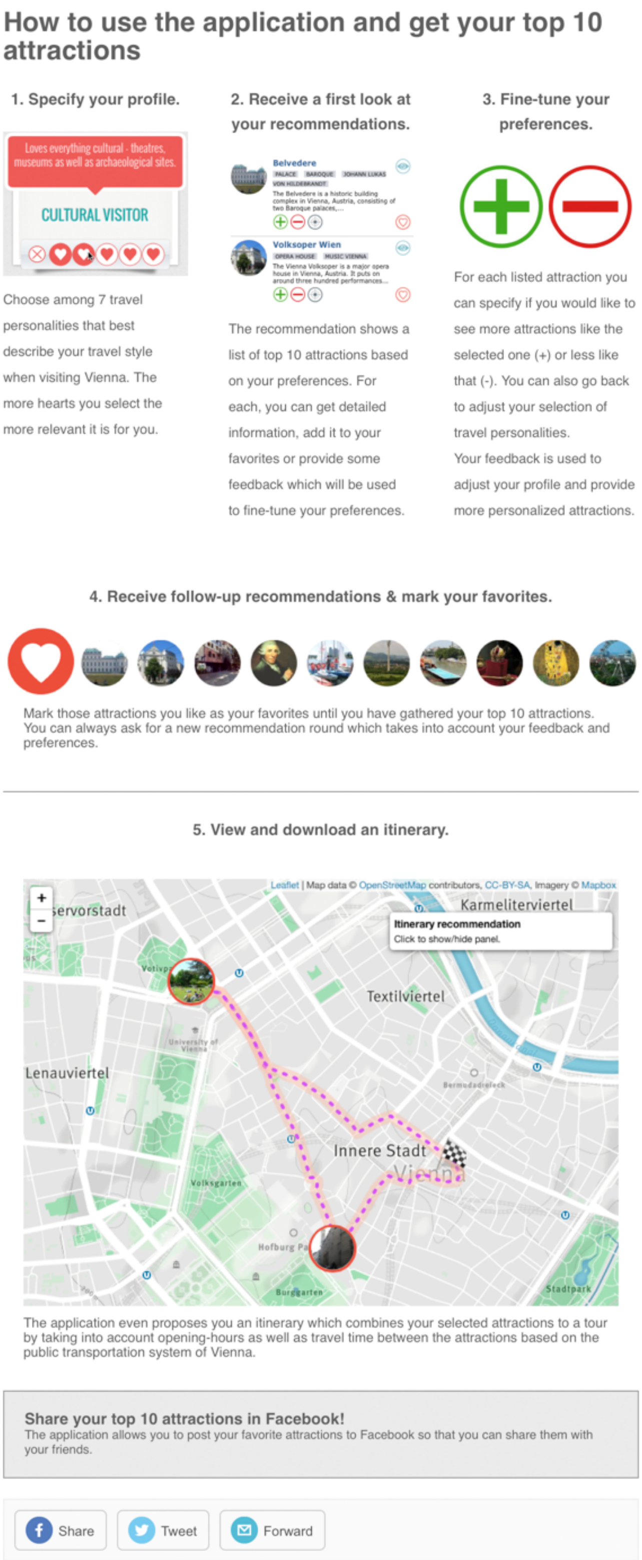
The aim of this project is to evaluate an online recommendation application which helps tourists visiting Vienna to select attractions which are best tailored to their personal needs.

We would very much welcome your assistance in testing the application by selecting your personal top 10 attractions and completing afterwards the questionnaire which can be found at the application website.

The value of this research lies in your deliberate responses and therefore it will be very helpful if you are able to answer all questions. Your individual responses will be held in the strictest of confidence and only used anonymously to provide overall results.

Go to the application by clicking the link below.

START HERE



Facebook posting with the favorite tourism objects

Following posting demonstrates how the user can share his/her favorite tourism objects with his/her Facebook friends. When the posting is clicked a new tab is opened in the browser and the user is directed to the newsletter.

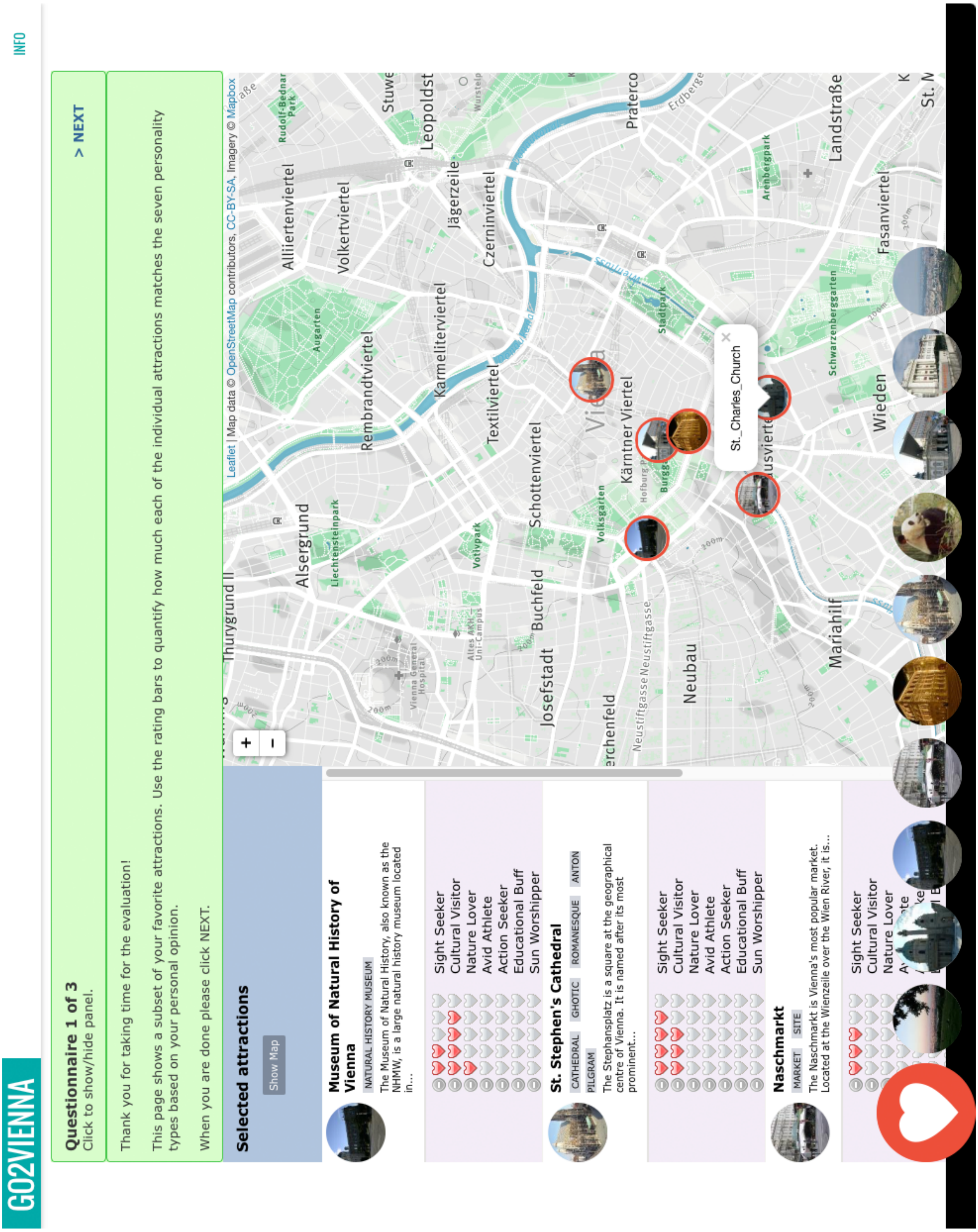
Online survey

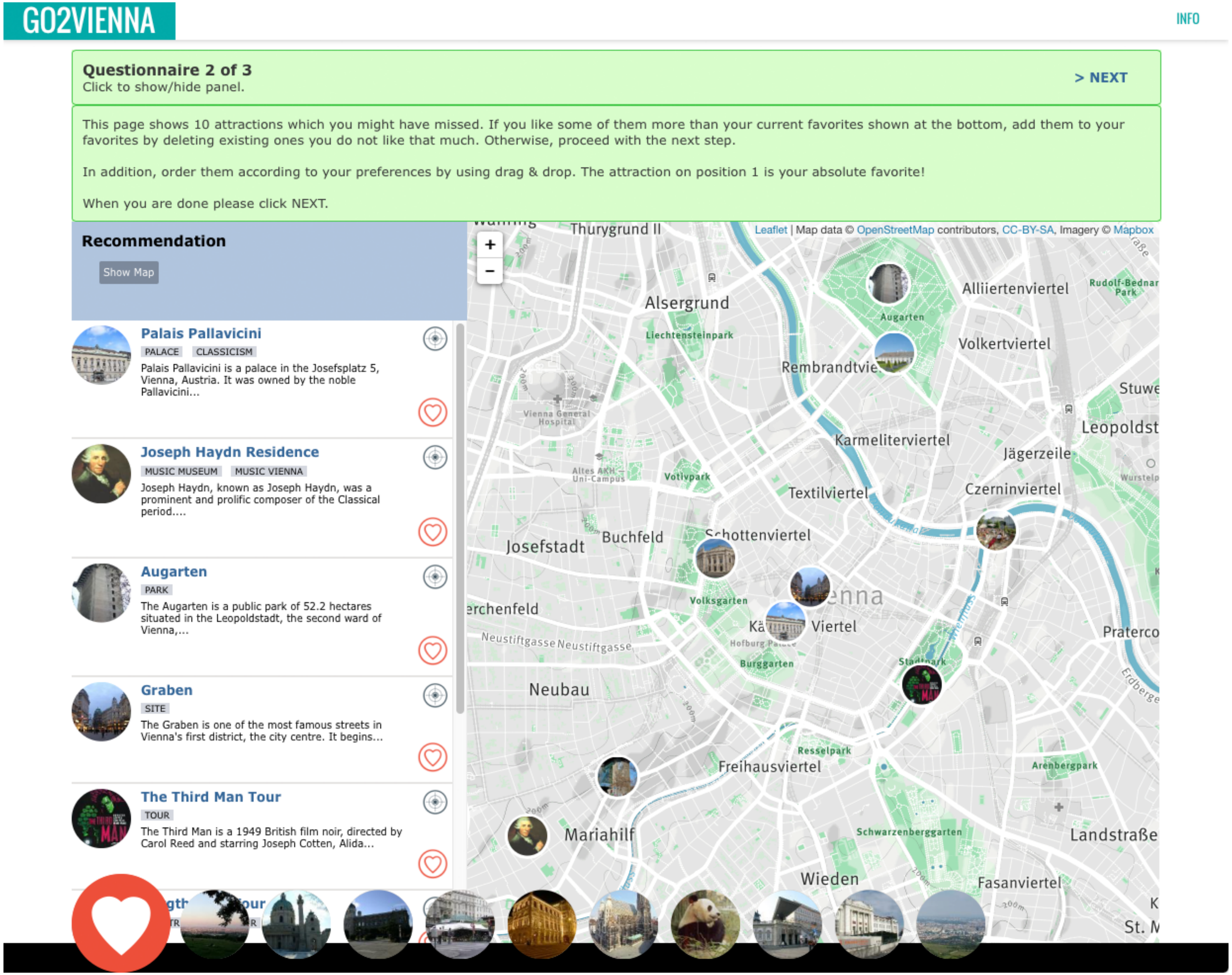
The online survey consisted of three pages which are depicted in the following. The time the user spent on each page and for the total questionnaire was measured in order to identify users who do not take the survey serious and just click through the questionnaire within some seconds.

The first page showed up to five tourism objects which had been selected by the user beforehand. The user was asked to quantify how much each of these objects matches the seven tourist factors in his/her personal opinion.

The second page showed ten tourism objects from the tourism objects database which the user had not selected beforehand. The selection was done randomly. The user had the possibility to replace one or more of his/her favorite objects with these new objects if he/she liked them better than the currently selected favorites.

The third page asked the user to provide some personal information and comment on the usability of the application, its interface as well as the recommendation features.


GO2VIENNA
INFO
Questionnaire 1 of 3
Click to show/hide panel.
> NEXT
Thank you for taking time for the evaluation!
This page shows a subset of your favorite attractions. Use the rating bars to quantify how much each of the individual attractions matches the seven personality types based on your personal opinion.
When you are done please click NEXT.
Selected attractions
Show Map
Museum of Natural History of Vienna
NATURAL HISTORY MUSEUM
The Museum of Natural History, also known as the NHMW, is a large natural history museum located in...
Sight Seeker
Cultural Visitor
Nature Lover
Avid Athlete
Action Seeker
Educational Buff
Sun Worshipper
St. Stephen's Cathedral
CATHEDRAL
GOTHIC
ROMANESQUE
ANTON
PILGRAM
The Stephansplatz is a square at the geographical centre of Vienna. It is named after its most prominent...
Sight Seeker
Cultural Visitor
Nature Lover
Avid Athlete
Action Seeker
Educational Buff
Sun Worshipper
Naschmarkt
MARKET
SITE
The Naschmarkt is Vienna's most popular market. Located at the Wienzeile over the Wien River, it is...
Sight Seeker
Cultural Visitor
Nature Lover
Leaflet | Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © Mapbox
St._Charles_Church
Alsergrund
Josefstadt
Neubau
Mariahilf
Wieden
Landstraße
Leopoldstadt


GO2VIENNA
INFO
Questionnaire 2 of 3
Click to show/hide panel.
> NEXT
This page shows 10 attractions which you might have missed. If you like some of them more than your current favorites shown at the bottom, add them to your favorites by deleting existing ones you do not like that much. Otherwise, proceed with the next step.
In addition, order them according to your preferences by using drag & drop. The attraction on position 1 is your absolute favorite!
When you are done please click NEXT.
Recommendation
Show Map
Palais Pallavicini
PALACE
CLASSICISM
Palais Pallavicini is a palace in the Josefsplatz 5, Vienna, Austria. It was owned by the noble Pallavicini...
Joseph Haydn Residence
MUSIC MUSEUM
MUSIC VIENNA
Joseph Haydn, known as Joseph Haydn, was a prominent and prolific composer of the Classical period....
Augarten
PARK
The Augarten is a public park of 52.2 hectares situated in the Leopoldstadt, the second ward of Vienna,...
Graben
SITE
The Graben is one of the most famous streets in Vienna's first district, the city centre. It begins...
The Third Man Tour
TOUR
The Third Man is a 1949 British film noir, directed by Carol Reed and starring Joseph Cotten, Alida...
Leaflet | Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © Mapbox
Thurygrund II
Alsergrund
Liechtensteinpark
Augarten
Alliiertenviertel
Volkertviertel
Rembrandtviertel
Leopoldstadt
Karmeliterviertel
Jägerzeile
Votivpark
Textilviertel
Czerninviertel
Josefstadt
Buchfeld
Schottenviertel
Volksgarten
Burggarten
Neubau
Resselpark
Freihausviertel
Arenbergpark
Mariahilf
Schwarzenberggarten
Landstraße
Wieden
Fasanviertel

GO2VIENNA INFO

Questionnaire 3 of 3
Click to show/hide panel.

On this page, we ask you to provide some personal information and to comment on the usability of the application, its interface as well as the recommendation features.

When you are done please click on the submit button.

What is your gender?
- Male
- Female

What is your age?
- Under 18
- 18-24
- 25-34
- 35-49
- 50-64
- 65 and over

What is your nationality?
Please choose

Have you already visited Vienna?
- Yes
- No

Overall, I am satisfied with this application.

Strongly disagree | disagree | agree | Strongly Agree

The interface of this application is pleasant.

Strongly disagree | disagree | agree | Strongly Agree

Recommendation feature.
The recommendation feature is effective in helping me to find suitable attractions.

Strongly disagree | disagree | agree | Strongly Agree

Diversification feature.
At the beginning, I would rather like to get a diversified set of attractions to explore alternatives (e.g. the set of recommended attractions should not only contain attractions of a specific type such as museum).

Strongly disagree | disagree | agree | Strongly Agree

When repeatedly declaring interest in attractions of a specific type (e.g. churches) I prefer getting more attractions of this kind at the price of less variance of the proposed attractions.

Strongly disagree | disagree | agree | Strongly Agree

I have following feedback regarding the user interface.

I have following feedback regarding the recommendation feature.

I suggest to add following features to improve the application.

Submit

List of figures

List of tables

Bibliography

[1] *E-commerce statistics for individuals.* http://ec.europa.eu/eurostat/statistics-explained/index.php/E-commerce_statistics_for_individuals.

[2] *The National Monuments Record Thesauri.* http://thesaurus.english-heritage.org.uk/.

[3] *Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships. The Knowledge Engineering Review*, 16:111–155, 2001.

[4] *Heracles II: Conditional constraint networks for interleaved planning and information gathering. IEEE Intelligent Systems*, 20:25–33, 2005.

[5] *The Impact of Semantic Web Technologies on Job Recruitment Processes.* In *In Proceedings of the 7th International Conference Wirtschaftsinformatik*, pages 1367–1383, 2005.

[6] *SimPack: A Generic Java Library for Similarity Measures in Ontologies.* , 2005.

[7] *User-centred versus system-centred evaluation of a personalization system. Information Processing & Management*, 44(3):1293 – 1307, 2008.

[8] *Semantic Similarity in Biomedical Ontologies. PLoS Computational Biology*, 2009.

[9] *CRUZAR: An Application of Semantic Matchmaking to e-Tourism.* In *Cases on Semantic Interoperability for Information Systems Integration: Practices and Applications.* IGI GLOBAL, 2010.

[10] Gregory D. Abowd, Christopher G. Atkeson, Jason Hong, Sue Long, Rob Kooper, and Mike Pinkerton. *Cyberguide: a mobile context-aware tour guide. Wirel. Netw.*, 3(5):421–433, 1997.

[11] Akiko Aizawa. *An information-theoretic perspective of tf-idf measures. Information Processing & Management*, 39(1):45 – 65, 2003.

[12] Liliana Ardissono, Anna Goy, Giovanna Petrone, Marino Segnan, and Pietro Torasso. *Intrigue: Personalized Recommendation Of Tourist Attractions For Desktop And Handset Devices.* In *Applied Artificial Intelligence*, pages 687–714. Taylor and Francis, 2003.

[13] Reyhan Aydoğan and Pinar Yolum. *Learning consumer preferences using semantic similarity.* In *AAMAS '07: Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–8, New York, NY, USA, 2007. ACM.

[14] N. Mirzadeh A. Venturini B. Arslan, F. Ricci. *A dynamic approach to feature weighting.* In *In Proceedings of Data Mining 2002 Conference, Bologna, Italy, September 25-27*, 2002.

[15] Elena García Barriocanal and Miguel-Ángel Sicilia. *On Linking Cultural Spaces and e-Tourism: An Ontology-Based Approach.* In Miltiadis D. Lytras, John M. Carroll, Ernesto Damiani, Robert D. Tennyson, David E. Avison, Gottfried Vossen, and Patricia Ordóñez de Pablos,

editors, *WSKS (2)*, volume 19 of *Communications in Computer and Information Science*, pages 694–701. Springer, 2008.

[16] Robert Barta, Christina Feilmayr, Birgit Pröll, Christoph Grün, and Hannes Werthner. *Covering the Semantic Space of Tourism: An Approach Based on Modularized Ontologies.* In *Proceedings of the 1st Workshop on Context, Information and Ontologies*, CIAO '09, pages 1:1–1:8, New York, NY, USA, 2009. ACM.

[17] Tim Berners-Lee and Mark Fischetti. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor.* Harper San Francisco, 1st. Auflage, 1999.

[18] Andreas Billig, Eva Blomqvist, and Feiyu Lin. *Semantic Matching Based on Enterprise Ontologies.* In Robert Meersman and Zahir Tari, editors, *OTM Conferences (1)*, volume 4803 of *Lecture Notes in Computer Science*, pages 1161–1168. Springer, 2007.

[19] Emmanuel Blanchard, Mounira Harzallah, Henri Bri, Pascale Kuntz, and Rue Christian Pauc. *A Typology Of Ontology-Based Semantic Measures.* In *in proc of EMOI-INTEROP'05 workshop, at CAiSE'05*, pages 13–14, 2005.

[20] Keith Bradley and Barry Smyth. *Improving Recommendation Diversity*, 2001.

[21] John S. Breese, David Heckerman, and Carl Kadie. *Empirical Analysis of Predictive Algorithms for Collaborative Filtering.* pages 43–52. Morgan Kaufmann, 1998.

[22] Jos de Bruijn, Dieter Fensel, Mick Kerrigan, Uwe Keller, Holger Lausen, and James Scicluna. *Modeling Semantic Web Services: The Web Service Modeling Language.* Springer Publishing Company, Incorporated, 1. Auflage, 2008.

[23] Peter Brusilovsky and Eva Millán. *User Models for Adaptive Hypermedia and Adaptive Educational Systems.* In Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, editors, *The Adaptive Web*, volume 4321 of *Lecture Notes in Computer Science*, chapter 1, pages 3–53. Springer Berlin Heidelberg, 2007.

[24] Robin Burke. *Knowledge-based Recommender Systems.* In *Encyclopedia of Library and Information Systems*, volume 69, 2000.

[25] Robin Burke. *Hybrid web recommender systems.* pages 377–408, 2007.

[26] Victor Codina Busquet. *Design, Development and Deployment of an Intelligent, Personalized Recommendation System.* Master's thesis, 2009.

[27] Jorge Cardoso. *Developing An Owl Ontology For e-Tourism.* In Jorge Cardoso and Amit P. Sheth, editors, *Semantic Web Services, Processes and Applications*, volume 3 of *Semantic Web And Beyond Computing for Human Experience*, pages 247–282. Springer, 2006.

[28] CEN Workshop on Harmonization of data interchange in tourism - WS/eTOUR. *CEN Workshop Agreement, Draft version.* , 2009.

[29] Guanling Chen and David Kotz. *A Survey of Context-Aware Mobile Computing Research.* , Hanover, NH, USA, 2000.

[30] Keith Cheverst, Keith Mitchell, and Nigel Davies. *Design of an object model for a context sensitive tourist GUIDE.* In *Computers and Graphics*, pages 883–891, 1999.

[31] Mark Claypool, Phong Le, Makoto Waseda, and David Brown. *Implicit Interest Indicators.* In *In Intelligent User Interfaces*, pages 33–40. ACM Press, 2001.

[32] E. Cohen. *Towards a Sociology of International Tourism. Social Research*, 39:164–182, 1972.

[33] Erik. Cohen. *Contemporary tourism : diversity and change / Erik Cohen*. Elsevier, Amsterdam ; Boston :, 1st ed.. Auflage, 2004.

[34] Francisco M. Couto, Mário J. Silva, and Pedro M. Coutinho. *Semantic similarity over the gene ontology: family correlation and selecting disjunctive ancestors*. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 343–344, New York, NY, USA, 2005. ACM.

[35] Graham M. S. Dann. *Anomie, ego-enhancement and tourism. Annals of Tourism Research*, 4(4):184 – 194, 1977.

[36] J. Olawande Daramola, Mathew Adigun, and Charles Ayo. *Building an Ontology-Based Framework for Tourism Recommendation Services*. In Höpken et al. [71], pages 135–147.

[37] M. Dell'Erba, Oliver Fodor, W. Höpken, and Hannes Werthner. *Exploiting Semantic Web Technologies for Harmonizing e-markets. Information Technology & Tourism*, 7/4, 2005.

[38] T Dema. *eTourPlan: A Knowledge-Based Tourist Route and Activity Planner*. Master's thesis, The University Of new Brunswick, 2007.

[39] DERI. *OnTour Ontology*. http://e-tourism.deri.at/ont/index.html.

[40] Anind K. Dey. *Understanding and Using Context. Personal Ubiquitous Comput.*, 5(1):4–7, 2001.

[41] Birgit Dippelreiter, Christoph Grün, Michael Pöttler, Ingo Seidel, Helmut Berger, Michael Dittenbach, and Andreas Pesenhofer. *Online Tourism Communities on the Path to Web 2.0: An Evaluation. J. of IT & Tourism*, 10(4):329–353, 2008.

[42] John Domingue, Dieter Fensel, and James A. Hendler. *Handbook of Semantic Web Technologies*. Springer Publishing Company, Incorporated, 1st. Auflage, 2011.

[43] EON. *EON Traveling Ontology*. http://homepages.cwi.nl/ troncy/DOE/ontologies/EON-TravellingOntology-v0.2.daml.

[44] Jérôme Euzenat and Pavel Shvaiko. *Ontology matching*. Springer-Verlag, Heidelberg (DE), 2007.

[45] Ricci F. and Werthner H. *Case Base Querying for Travel Planning Recommendation. Information Technology & Tourism*, 4, 2001.

[46] Alexander Felfernig, Gerhard Friedrich, Dietmar Jannach, and Markus Zanker. *An Integrated Environment for the Development of Knowledge-Based Recommender Applications. Int. J. Electron. Commerce*, 11(2):11–34, 06-7.

[47] D. R. Fesenmaier, H. Werthner, and K. W. Wöber. *Destination Recommendation Systems: Behavioural Foundations and Applications*, chapter Travel Destination Choice Models, pages 3–16. CAB International, 2006.

[48] Josef Fink and Alfred Kobsa. *User Modeling for Personalized City Tours. Artif. Intell. Rev.*, 18(1):33–74, 2002.

[49] Oliver Fodor and Hannes Werthner. *Harmonise: A Step Toward an Interoperable E-Tourism Marketplace. Int. J. Electron. Commerce*, 9(2):11–39, 2005.

[50] Jo-Ann Foo, Robyn McGuiggan, and Andrew Yiannakis. *ROLES TOURISTS PLAY: An Australian Perspective. Annals of Tourism Research*, 31(2):408 – 427, 2004.

[51] Joseph D. Fridgen. *Environmental psychology and tourism. Annals of Tourism Research*, 11(1):19 – 39, 1984.

[52] Angel García-Crespo, Javier Chamizo, Ismael Rivera, Myriam Mencke, Ricardo Colomo-Palacios, and Juan Miguel Gómez-Berbís. *SPETA: Social pervasive e-Tourism advisor. Telemat. Inf.*, 26(3):306–315, 2009.

[53] Susan Gauch, Mirco Speretta, Aravind Chandramouli, and Alessandro Micarelli. *User profiles for personalized information access.* 2007.

[54] Heather Gibson and Andrew Yiannakis. *Tourist roles: Needs and the Lifecourse. Annals of Tourism Research*, 29(2):358 – 383, 2002.

[55] Asuncion Gomez-Perez, Oscar Corcho, and Mariano Fernandez-Lopez. *Ontological Engineering : with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. First Edition (Advanced Information and Knowledge Processing).* Springer, July 2004.

[56] Dina Goren-Bar, Ilenia Graziola, Fabio Pianesi, and Massimo Zancanaro. *The influence of personality factors on visitor attitudes towards adaptivity dimensions for mobile museum guides. User Modeling and User-Adapted Interaction*, 16(1):31–62, 2006.

[57] Ulrike Gretzel, Nicole Mitsche, Yeong-Hyeon Hwang, and Daniel R. Fesenmaier. *Tell me who you are and I will tell you where to go: Use of Travel Personalities in Destination Recommendation Systems. Information Technology & Tourism*, 7:3– 12, 2004.

[58] Ulrike Gretzel and Kyung Hyan Yoo. *Use and Impact of Online Travel Reviews.* In Peter O'Connor, Wolfram Höpken, and Ulrike Gretzel, editors, *ENTER*, pages 35–46. Springer, 2008.

[59] Thomas R. Gruber. *A translation approach to portable ontology specifications. Knowl. Acquis.*, 5(2):199–220, 1993.

[60] Christoph Grün. *Making pre-trip services context-aware.* In *Proceedingsof the 2nd International Workshop on the Semantic Web (SemWeb'2001), Hong Kong, China.* Citeseer, 2001.

[61] Christoph Grün, Wieland Schwinger, Birgit Pröll, and Werner Retschitzegger. *Context-awareness in Mobile Tourism Guides.* In *Handbook of Research on Mobile Multimedia, Second Edition*, pages 298–314. Information Science Reference, 2, 2008.

[62] Christoph Grün, Hannes Werthner, Birgit Pröll, Werner Retschitzegger, and Wieland Schwinger. *Assisting tourists on the move-an evaluation of mobile tourist guides.* In *2008 7th International Conference on Mobile Business.* IEEE, 2008.

[63] N. Guarino. *Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy.* IOS Press, Amsterdam, The Netherlands, The Netherlands, 1998.

[64] C. Michael Hall and Alan A. Lew. *Understanding and Managing Tourism Impacts: An Integrated Approach.* Routledge, 2009.

[65] Patrick A. V. Hall and Geoff R. Dowling. *Approximate String Matching. ACM Comput. Surv.*, 12(4):381–402, 1980.

[66] Martin Hepp and Jos de Bruijn. *GenTax: A Generic Methodology for Deriving OWL and RDF-S Ontologies from Hierarchical Classifications, Thesauri, and Inconsistent Taxonomies.* In Enrico Franconi, Michael Kifer, and Wolfgang May, editors, *ESWC*, volume 4519 of *Lecture Notes in Computer Science*, pages 129–144. Springer, 2007.

[67] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. *Evaluating Collaborative Filtering Recommender Systems. ACM Trans. Inf. Syst.*, 22(1):5–53, January 2004.

[68] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. *Design Science in Information Systems Research. MIS Quarterly*, 28(1):75–105, 2004.

[69] G. Hirst and D. St-Onge. *Lexical Chains as representation of context for the detection and correction malapropisms*, 1997.

[70] Wolfram Höpken, Matthias Fuchs, Markus Zanker, Thomas Beer, Alexander Eybl, Stefan Flores, Sergiu Gordea, Markus Jessenitschnig, Thomas Kerner, Dirk Linke, Jörg Rasinger, and Michael Schnabl. *etPlanner: An IT Framework for Comprehensive and Integrative Travel Guidance.* In *Information and Communication Technologies in Tourism 2006*, chapter 20, pages 125–134. 2006.

[71] Wolfram Höpken, Ulrike Gretzel, and Rob Law, editors. *Information and Communication Technologies in Tourism, ENTER 2009, Proceedings of the International Conference in Amsterdam, The Netherlands, 2009.* Springer, 2009.

[72] Yuxia Huang and Ling Bian. *A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet. Expert Syst. Appl.*, 36(1):933–943, 2009.

[73] Yuxia Huang and Ling Bian. *Ontology-driven tour-planning systems: a conceptual framework. Environment and Planning B: Planning and Design*, 37(3):483–499, May 2010.

[74] Salvatore Maria Ielpa, Salvatore Iiritano, Nicola Leone, and Francesco Ricca. *An ASP-Based System for e-Tourism.* In *LPNMR '09: Proceedings of the 10th International Conference on Logic Programming and Nonmonotonic Reasoning*, pages 368–381, Berlin, Heidelberg, 2009. Springer-Verlag.

[75] Roopa Jakkilinki, Mladen Georgievski, and Nalin Sharda. *Connecting Destinations with an Ontology-Based e-Tourism Planner.* In Marianna Sigala, Luisa Mich, and Jamie Murphy, editors, *ENTER*, pages 21–32. Springer, 2007.

[76] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender Systems: An Introduction.* Cambridge University Press, New York, NY, USA, 1st. Auflage, 2010.

[77] J.J. Jiang and D.W. Conrath. *Semantic similarity based on corpus statistics and lexical taxonomy.* In *Proc. of the Int'l. Conf. on Research in Computational Linguistics*, pages 19–33, 1997.

[78] Terry Stevens John Walsh-Heron. *The Management of Visitor Attractions and Events.* Prentice Hall, 1990.

[79] Soo Hyun Jun, Christine A. Vogt, and Kelly J. MacKay. *Relationships between Travel Information Search and Travel Product Purchase in Pretrip Contexts. Journal of Travel Research*, 45(3):266–274, 2007.

[80] Dong Jin Kim, Woo Gon Kim, and Jin Soo Han. *A perceptual mapping of online travel agencies and preference attributes. Tourism Management*, 28(2):591 – 603, 2007.

[81] Rasmus Knappe, Henrik Bulskov, and Troels Andreasen. *Perspectives on ontology-based querying: Research Articles. Int. J. Intell. Syst.*, 22(7):739–761, 2007.

[82] Danielle H. Lee and Peter Brusilovsky. *Reinforcing Recommendation Using Implicit Negative Feedback.* In *UMAP '09: Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization*, pages 422–427, Berlin, Heidelberg, 2009. Springer-Verlag.

[83] B. Legrand. *Semantic Web Methodologies and Tools for Intra-European Sustainable Tourism.* , Mondeca, 2004.

[84] Qing Li and Byeong Man Kim. *Clustering Approach for Hybrid Recommender System.* In *WI '03: Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence*, page 33, Washington, DC, USA, 2003. IEEE Computer Society.

[85] Dekang Lin. *An Information-Theoretic Definition of Similarity.* In *ICML '98: Proceedings of the Fifteenth International Conference on Machine Learning*, pages 296–304, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

[86] Stephen W. Litvin. *Revisiting Plog's Model of Allocentricity and Psychocentricity... One More Time. Cornell Hotel and Restaurant Administration Quarterly*, 47(3):245–253, 2006.

[87] Bing Liu, Siew-Hwee Choo, Shee-Ling Lok, Sing-Meng Leong, Soo-Chee Lee, Foong-Ping Poon, and Hwee-Har Tan. *Integrating case-based reasoning, knowledge-based approach and Dijkstra algorithm for route finding.* In *Artificial Intelligence for Applications, 1994., Proceedings of the Tenth Conference on*, pages 149–155, Mar 1994.

[88] Nilgun Avci Nurten Cifter Luisa Andreu, Metin Kozak. *Market Segmentation by Motivations to Travel – British Tourists Visiting Turkey. Journal of Travel & Tourism Marketing*, 19:1–14, 2006.

[89] Marko Luther, Yusuke Fukazawa, Matthias Wagner, and Shoji Kurakake. *Situational reasoning for task-oriented mobile service recommendation. Knowl. Eng. Rev.*, 23(1):7–19, 2008.

[90] Laurent Mazuel and Nicolas Sabouret. *Semantic Relatedness Measure Using Object Properties in an Ontology.* In *ISWC '08: Proceedings of the 7th International Conference on The Semantic Web*, pages 681–694, Berlin, Heidelberg, 2008. Springer-Verlag.

[91] Bob McKercher. *Are Psychographics Predictors of Destination Life Cycles? Journal of Travel & Tourism Marketing*, 19:49–55, 2006.

[92] Zbigniew Michalewicz and David B. Fogel. *How to solve it - modern heuristics: second, revised and extended edition (2. ed.).* Springer, 2004.

[93] Gianna Moscardo, Alastair M. Morrison, Philip L. Pearce, Cheng-Te Lang, and Joseph T. O'Leary. *Understanding vacation destination choice through travel motivation and activities. Journal of Vacation Marketing*, 2(2):109–122, 1996.

[94] Shiyan Ou; Viktor Pekar; Constantin Orasan; Christian Spurk; Matteo Negri. *Development and Alignment of a Domain-Specific Ontology for Question Answering.* In European Language Resources Association (ELRA), editor, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08). Marrakech, Morocco*. o.A., 5 2008.

[95] Julia Neidhardt, Rainer Schuster, Leonhard Seyfang, and Hannes Werthner. *Eliciting the Users' Unknown Preferences.* In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 309–312, New York, NY, USA, 2014. ACM.

[96] Julia Neidhardt, Leonhard Seyfang, Rainer Schuster, and Hannes Werthner. *A picture-based approach to recommender systems. Information Technology & Tourism*, 15(1):49–69, 2014.

[97] Magnus Niemann, Malgorzata Mochol, and Robert Tolksdorf. *Enhancing hotel search with semantic web technologies. J. Theor. Appl. Electron. Commer. Res.*, 3(2):82–96, 2008.

[98] Natalya F. Noy and Deborah L. McGuinness. *Ontology Development 101: A Guide to Creating Your First Ontology.* Online, 2001.

[99] World Tourism Organisation-WTO. *Thesaurus on tourism and leisure activities of the World Tourism Organization*, 2002. http://www.world-tourism.org.

[100] Razib M. Othman, Safaai Deris, and Rosli M. Illias. *A genetic similarity algorithm for searching the Gene Ontology terms and annotating anonymous protein sequences. J. of Biomedical Informatics*, 41(1):65–81, 2008.

[101] Bing Pan and Daniel R. Fesenmaier. *Online Information Search: Vacation Planning Process. Annals of Tourism Research*, 33(3), 2006.

[102] Michael Pazzani and Daniel Billsus. *Content-Based Recommendation Systems.* pages 325–341. 2007.

[103] P. L. Pearce. *Fundamentals of tourist motivation*, chapter 7, pages 85–105. Routledge, 1993.

[104] Philip L. Pearce and Uk-Il Lee. *Developing the Travel Career Approach to Tourist Motivation. Journal of Travel Research 43*, pages 226–237, 2005.

[105] Peter O'Connor. *Electronic Information Distribution in Tourism and Hospitality.* CABI, 1999.

[106] Stanley Plog. *Why destination areas rise and fall in popularity: an update of a Cornell Quarterly classic. The Cornell Hotel and Restaurant Administration Quarterly*, 42(3):13 – 24, 2001.

[107] Yaniv Poria, Richard Butler, and David Airey. *Links between Tourists, Heritage, and Reasons for Visiting Heritage Sites. Journal of Travel Research*, 43(1):19–28, 2004.

[108] Richard Prentice. *Tourism and heritage attractions / Richard Prentice.* Routledge, London ; New York :, 1993.

[109] Roy Rada, Hafedh Mili, Ellen Bicknell, and Maria Blettner. *Development and Application of a Metric on Semantic Nets. IEEE Transactions on Systems Management and Cybernetics*, 19(1), 1989.

[110] Jörg Rasinger, Matthias Fuchs, and Wolfram Höpken. *Information Search with Mobile Tourist Guides: A Survey of Usage Intention. Information Technology & Tourism*, 9:177–194(18), September 2007.

[111] Reisewissen. http://reisewissen.ag-nbi.de/.

[112] Philip Resnik. *Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. Journal of Artificial Intelligence Research*, 11:95–130, 1998.

[113] Francesco Ricci. *Travel recommender systems. IEEE Intelligent Systems*, 17 (6):55–57, 2002.

[114] Francesco Ricci. *Mobile Recommender Systems.* , Free University of Bolzano, Italy, 2010. Submitted to the International Journal of Information Technology and Tourism.

[115] Francesco Ricci, Bora Arslan, Nader Mirzadeh, and Adriano Venturini. *ITR: A Case-Based Travel Advisory System.* In *ECCBR '02: Proceedings of the 6th European Conference on Advances in Case-Based Reasoning*, pages 613–627, London, UK, 2002. Springer-Verlag.

[116] Elaine Rich. *User modeling via stereotypes.* pages 329–342, 1998.

[117] G Richards. *Cultural Attractions and European Tourism.* CABI, 2001.

[118] G. Richards. *Tourism attraction systems - Exploring Cultural Behavior. Annals of Tourism Research*, 29:1048–1064(17), October 2002.

[119] R. Richardson, A. F. Smeaton, A. F. Smeaton, J. Murphy, and J. Murphy. *Using WordNet as a Knowledge Base for Measuring Semantic Similarity between Words.* , In Proceedings of AICS Conference, 1994.

[120] Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval.* McGraw-Hill, Inc., New York, NY, USA, 1986.

[121] Badrul Sarwar, George Karypis, Joseph Konstan, and John Reidl. *Item-based collaborative filtering recommendation algorithms.* In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 285–295, New York, NY, USA, 2001. ACM.

[122] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. *Analysis of recommendation algorithms for e-commerce.* In *EC '00: Proceedings of the 2nd ACM conference on Electronic commerce*, pages 158–167, New York, NY, USA, 2000. ACM.

[123] J. Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. *Collaborative Filtering Recommender Systems.* In Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, editors, *The Adaptive Web*, volume 4321 of *Lecture Notes in Computer Science*, chapter 9, pages 291–324. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[124] Silvia Schiaffino and Analía Amandi. *Building an expert travel agent as a software agent. Expert Syst. Appl.*, 36(2):1291–1299, 2009.

[125] Bill Schilit, Norman Adams, and Roy Want. *Context-Aware Computing Applications.* In *In Proceedings of the Workshop on Mobile Computing Systems and Applications*, pages 85–90. IEEE Computer Society, 1994.

[126] Nuno Seco, Tony Veale, and Jer Hayes. *An Intrinsic Information Content Metric for Semantic Similarity in WordNet.* In *ECAI 2004*, 2004.

[127] Sinan Sen and Jun Ma. *Contextualised Event-driven Prediction with Ontology-based Similarity.* In *Intelligent Event Processing, Papers from the 2009 AAAI Spring Symposium*, pages 73–79, 2009.

[128] Ahu Sieg, Bamshad Mobasher, and Robin Burke. *Web search personalization with ontological user profiles.* In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, New York, NY, USA, 2007. ACM.

[129] Smart Information Systems GmbH. *ebSemantics Ontology.*

[130] Barry Smyth. *Case-Based Recommendation.* In *The Adaptive Web: Methods and Strategies of Web Personalization*, chapter 11, pages 342–376. 2007.

[131] Cathy H.C. Hsu Songshan (Sam) Huang. *Travel motivation: linking theory to practice. International Journal of Culture, Tourism and Hospitality Research*, 3 (4):287–295, 2009.

[132] Steffen Staab, Christian Braun, Ilvio Bruder, Antje Düsterhöft, Andreas Heuer, Meike Klettke, Günter Neumann, Bernd Prager, Jan Pretzel, Hans-Peter Schnurr, Rudi Studer, Hans Uszkoreit, and Burkhard Wrenger. *GETESS - Searching the Web Exploiting German Texts.* In *CIA '99: Proceedings of the Third International Workshop on Cooperative Information Agents III*, pages 113–124, London, UK, 1999. Springer-Verlag.

[133] Steffen Staab, Hannes Werthner, Francesco Ricci, Alexander Zipf, Ulrike Gretzel, Daniel R. Fesenmaier, Cécile Paris, and Craig A. Knoblock. *Intelligent Systems for Tourism. IEEE Intelligent Systems*, 17(6), 2002.

[134] Thomas Strang and Claudia Linnhoff-Popien. *A Context Modeling Survey.* In *In: Workshop on Advanced Context Modelling, Reasoning and Management, UbiComp 2004 - The Sixth International Conference on Ubiquitous Computing, Nottingham/England*, 2004.

[135] Heiner Stuckenschmidt and Frank van Harmelen. *Information Sharing on the Semantic Web.* SpringerVerlag, 2004.

[136] Michael Sussna. *Word sense disambiguation for free-text indexing using a massive semantic network.* In *CIKM '93: Proceedings of the second international conference on Information and knowledge management*, pages 67–74, New York, NY, USA, 1993. ACM.

[137] TAGA Ontology. http://taga.sourceforge.net/owl/index.html.

[138] Francisco Tanudjaja and Lik Mui. *Persona: A Contextualized and Personalized Web Search.* In *In Proc. of the 35th Annual Hawaii International Conference on System Sciences*, page 67, 2001.

[139] N. Tintarev and J. Masthoff. *A Survey of Explanations in Recommender Systems.* In *Data Engineering Workshop, 2007 IEEE 23rd International Conference on*, pages 801–810, April 2007.

[140] Eleni Tomai, Maria Spanaki, Poulicos Prastacos, and Marinos Kavouras. *Ontology Assisted Decision Making - A Case Study in Trip Planning for Tourism.* In Robert Meersman, Zahir Tari, Pilar Herrero, Gonzalo Méndez, Lawrence Cavedon, David Martin, Annika Hinze, George Buchanan, María S. Pérez, Víctor Robles, Jan Humble, Antonia Albani, Jan L. G. Dietz, Hervé Panetto, Monica Scannapieco, Terry A. Halpin, Peter Spyns, Johannes Maria Zaha, Esteban Zimányi, Emmanuel Stefanakis, Tharam S. Dillon, Ling Feng, Mustafa Jarrar, Jos Lehmann, Aldo de Moor, Erik Duval, and Lora Aroyo, editors, *OTM Workshops*, volume 3762 of *Lecture Notes in Computer Science*, pages 1137–1146. Springer, 2005.

[141] TRUSTYOU. http://www.trustyou.com/.

[142] A. Tversky. *Features of similarity. Psychological Review*, 84, 1977.

[143] C. Tweed. *Taxonomy of cultural attractors, Deliverable 8.* , PICTURE: Pro-active management of the Impact of Cultural Tourism Upon Urban Resources and Economies, 2005.

[144] Daniel R. Fesenmaier Ulrike Gretzel and Joseph T. O'Leary. *Tourism Business Frontiers*, chapter The transformation of Consumer Behaviour, pages 9–18. Butterworth-Heinemann, 2005.

[145] The World Tourism Organization (UNWTO). *International Recommendations for Tourism Statistics*, 2007. http://unstats.un.org/unsd/statcom/doc08/BG-TourismStats.pdf.

[146] Pieter Vansteenwegen, Wouter Souffriau, Greet Vanden Berghe, and Dirk Van Oudheusden. *Iterated local search for the team orienteering problem with time windows. Computers & OR*, 36(12):3281–3290, 2009.

[147] Giannis Varelas, Epimenidis Voutsakis, Paraskevi Raftopoulou, Euripides G.M. Petrakis, and Evangelos E. Milios. *Semantic similarity methods in wordNet and their application to information retrieval on the web.* In *WIDM '05: Proceedings of the 7th annual ACM international workshop on Web information and data management*, pages 10–16, New York, NY, USA, 2005. ACM.

[148] Manuela Veloso and Agnar Aamodt, editors. *Case-Based Reasoning Research and Development.* Springer Verlag, October 1995.

[149] Carlos Vicient, David Sánchez, and Antonio Moreno. *An automatic approach for ontology-based feature extraction from heterogeneous textualresources. Engineering Applications of Artificial Intelligence*, 26(3):1092 – 1106, 2013.

[150] J. T. L. Wang, K. Zhang, K. Jeong, and D. Shasha. *A System for Approximate Tree Matching. IEEE Trans. on Knowl. and Data Eng.*, 6(4):559–571, 1994.

[151] Mark Weiser. *Some computer science issues in ubiquitous computing. SIGMOBILE Mob. Comput. Commun. Rev.*, 3(3):12, 1999.

[152] H. Werthner and S. Klein. *Information Technology and Tourism - A Challenging Relationship*. Springer, 1999.

[153] Hannes Werthner and Francesco Ricci. *E-commerce and tourism. Commun. ACM*, 47(12):101–105, 2004.

[154] Zhibiao Wu and Martha Stone Palmer. *Verb Semantics and Lexical Selection.* In *Proceedings of the 32nd. Annual Meeting of the Association for Computational Linguistics (ACL 1994)*, pages 133–138, 1994.

[155] Zheng Xiang, Ulrike Gretzel, and Daniel R. Fesenmaier. *Semantic Representation of Tourism on the Internet. Journal of Travel Research*, page 0047287508326650, 2008.

[156] Andrew Yiannakis and Heather Gibson. *Roles tourists play. Annals of Tourism Research*, 19(2):287 – 303, 1992.

[157] Jin Young Young and Dimitrios Buhalis. *Information Needs in Online Social Networks. Information Technology & Tourism*, 10(4), 2008.

[158] Shijun Yu, Lina Al-Jadir, and Stefano Spaccapietra. *Matching User's Semantics with Data Semantics in Location-Based Services.* In *1st Workshop on Semantics in Mobile Environments (SME 05)*, 2005. in conjunction with MDM 2005.

[159] Markus Zanker, Matthias Fuchs, Alexander Seebacher, Markus Jessenitschnig, and Martin Stromberger. *Information and Communication Technologies in Tourism 2009: Proceedings of the International Conference in Amsterdam, The Netherlands, 2009*, chapter An Automated Approach for Deriving Semantic Annotations of Tourism Products based on Geospatial Information, pages 211–221. Springer Vienna, Vienna, 2009.

[160] Markus Zanker and Markus Jessenitschnig. *Case-studies on exploiting explicit customer requirements in recommender systems. User Modeling and User-Adapted Interaction*, 19(1):133–166, 2009.

[161] Xiaodan Zhang, Liping Jing, Xiaohua Hu, Michael Ng, and Xiaohua Zhou. *A comparative study of ontology based term similarity measures on PubMed document clustering.* In *DASFAA'07: Proceedings of the 12th international conference on Database systems for advanced applications*, pages 115–126, Berlin, Heidelberg, 2007. Springer-Verlag.

[162] Jiwei Zhong, Haiping Zhu, Jianming Li, and Yong Yu. *Conceptual Graph Matching for Semantic Search.* In *ICCS '02: Proceedings of the 10th International Conference on Conceptual Structures*, pages 92–196, London, UK, 2002. Springer-Verlag.

[163] Cai-Nicolas Ziegler, Georg Lausen, and Lars Schmidt-Thieme. *Taxonomy-driven computation of product recommendations.* In *CIKM '04: Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pages 406–415, New York, NY, USA, 2004. ACM.

[164] Andreas H. Zins. *Deconstructing Travel Decision Making and Information Search Activities.* In Höpken et al. [71], pages 467–479.

Curriculum Vitae

Christoph Grün holds a Master degree in Business Informatics from the University of Linz. From September 2005 to April 2006, he was employed at the University of Innsbruck, and from May 2006 to August 2010 he was a researcher at the E-commerce Group (Institute for Software Technology and Interactive Systems) at the TU Wien in Austria. His research focused on recommendation systems in the e-tourism domain, semantic Web technologies and mobile, context-aware systems. Since September 2010 he works at a consulting company in the area of systems integration and project management.

Book chapters

- *E-business and Semantic Technologies.* With: C. Huemer, P. Liegl, D. Mayrhofer, T. Motal, R. Schuster, H. Werthner, M. Zapletal. In: Handbook of Semantic Web Technologies, Springer, 2010
- *Context-awareness in Mobile Tourism Guides.* With: W. Schwinger, B. Pröll, W. Retschitzegger. In: Handbook of Research on Mobile Multimedia, Second Edition. Information Science Reference, 2008

Journal Papers

- *Context-based Matchmaking to Enhance Tourists' Experiences.* With: A. Sorzabal, C. Lamsfus, H. Werthner. In: Information Technology in the Tourism Industry. The European Journal for the Informatics Professional (CEPIS), 2010
- *Online Tourism Communities on the Path to Web 2.0 - An Evaluation.* With: H. Berger, B. Dippelreiter, M. Dittenbach, A. Pesenhofer, M. Pöttler. In: Information Technology & Tourism (10), 2008
- *From Business to Software: A B2B Survey.* With: Jürgen Dorn, Christoph Grün, and Hannes Werthner. In: Information Systems and E-Business Management - Special Issue on Design and management of business models and processes in services science, Springer, 2008

Conference and Workshop Papers

- *Ontology-based Matchmaking to Provide Personalized Recommendations for Tourists.* With: J. Neidhardt, H. Werthner. Submitted to: ENTER Conference 2017

- *A 'State of the Art' Evaluation of PM-Systems Exploring their missing Functionalities.* With: B. Dippelreiter, M. Pöttler. In: Proceedings of the 5th International Conference on Project Management, 2010
- *Covering the Semantic Space of Tourism - An Approach based on Modularized Ontologies.* With: R. Barta, C. Feilmayr, B. Pröll, H. Werthner. In: Proceedings of the 1st Workshop on Context, Information and Ontologies, ACM, 2009
- *Assisting Tourists on the Move - An Evaluation of Mobile Tourist Guides.* With: B. Pröll, W. Retschitzegger, W. Schwinger, H. Werthner. In: Proc. of 7th International Conference on Mobile Business, IEEE Computer Society, 2008
- *A Survey of B2B Methodologies and Technologies: From Business Models towards Deployment Artifacts.* With: J. Dorn, H. Werthner, M. Zapletal. In: 40th Hawaii International International Conference on Systems Science, IEEE Computer Society, 2007
- *Exploring Information Services for mobile Tourist Guides: Results from an Expert Survey.* With: M. Fuchs, J. Rasinger, W. Höpken. In: Proceedings of the Travel and Tourism Research Association Europe, CERAM Business School, 2007
- *Pushing location based games further: How to gain end user suitability.* With: S. Drab, J. Krösche, A. Jakl. In: Multimedia and E-Content Trends, Springer, 2007
- *Pinpointing Tourism Information onto Mobile Maps - A Leight-Weight Approach.* With: B. Pröll, W. Retschitzegger, W. Schwinger, H. Werthner. In: Proceedings of ENTER Conference 2006, Springer, 2006
- *A light-weight Framework for Location-based Services.* With: B. Pröll, W. Retschitzegger, W. Schwinger. In: Proceedings of the OTM 2005 Workshop on Context-Aware Mobile Systems (CAMS'05), 2005

Summer and Winter Schools

- *Second International Workshop on Information Technology and Tourism and Open Mobile Networks.* San Sebastian, Spain, 2009
- *Seventh European Summer School on Ontological Engineering and the Semantic Web (SSSW 2009).* Poster presentation. Cercedilla, Spain, 2009
- *1st Semantic Web Services Winter Retreat 2009.* Innsbruck, Austria, 2009
- *The Present and Future of Recommender Systems.* Bilbao, Spain, 2006