Tourism Itinerary Recommender Systems

In the age of personalization and crowdsourcing

LIVIU-ADRIAN COTFAS **CAMELIA DELCEA**

Tourism Itinerary Recommender Systems

In the age of personalization and crowdsourcing

EDITURA UNIVERSITARĂ
Bucureşti, 2017

Colecția ŞTIINȚE EXACTE

Redactor: Gheorghe Iovan
Tehnoredactor: Ameluţa Vişan
Coperta: Monica Balaban

Editură recunoscută de Consiliul Naţional al Cercetării Ştiinţifice (C.N.C.S.) şi inclusă de Consiliul Naţional de Atestare a Titlurilor, Diplomelor şi Certificatelor Universitare (C.N.A.T.D.C.U.) în categoria editurilor de prestigiu recunoscut.

Descrierea CIP a Bibliotecii Naţionale a României
COTFAS, LIVIU-ADRIAN
Tourism itinerary recommender systems : in the age of personalization and crowdsourcing / Liviu-Adrian Cotfas, Camelia Delcea. - Bucureşti : Editura Universitară, 2017
Conţine bibliografie
ISBN 978-606-28-0549-4

I. Delcea, Camelia

338.48

DOI: (Digital Object Identifier): 10.5682/9786062805494

Editura Universitară
Editor: Vasile Muscalu
B-dul. N. Bălcescu nr. 27-33, Sector 1, Bucureşti
Tel.: 021 – 315.32.47 / 319.67.27
www.editurauniversitara.ro
e-mail: redactia@editurauniversitara.ro

Distribuţie: tel.: 021-315.32.47 /319.67.27 / 0744 EDITOR / 07217 CARTE
comenzi@editurauniversitara.ro
O.P. 15, C.P. 35, Bucureşti
www.editurauniversitara.ro

CONTENTS

1. INTRODUCTION

Designing flexible, efficient and user-friendly mobile tour guide applications is important, both from a commercial and a research point of view. Without any support from a system, a user manually building an itinerary should spend a lot of time prior to the trip searching for information regarding the tourist attractions, reading facts about them and designing possible itineraries by taking into account factors such as opening hours, visiting durations, distances and available means of public transport.

The contributions of the present book are two-fold. From the theoretical point of view, it contributes to the design and development of highly personalized itinerary recommender systems, while from the practical point of view, an approach for building mobile applications using the latest web technologies is proposed and validated.

The first chapter, **Introduction**, outlines the evolution, current standards and the application areas of Geographic Information Systems. Based on the possibility to identify the user's location, a new type of applications, called Location Based Services – LBS are becoming increasingly popular. Such applications customize their content based on the user's geographic position. The evolution of GIS architectures from monolithic systems to distributed web service based architectures is also discussed. The main characteristics of an itinerary recommender system are presented afterwards, as well as a comparison between the system described through this book, called MoTripAssistent, and several other state of the art solutions. The proposed approach uses multi-objective genetic algorithms and collaborative filtering recommendation techniques in order to allow users to build highly personalized touristic itineraries.

The second chapter, **Context-Aware Personalized Recommendations**, focuses on the approach required for providing relevant touristic attractions suggestions, in accordance with the user's preferences and context. The context includes aspects, such as the time of the day, the period of the year, weather information and traffic data, that while not directly related to the user, can have a dramatic effect on the perceived utility associated to visiting a particular touristic attraction. A context-aware collaborative filtering recommender algorithm is presented, that takes into consideration both the user's long-term and the short-term preferences.

The third chapter, **Multi-Objective Itinerary Building**, discusses the approach required for building the complete itinerary of touristic attractions. Similar to many other real-world problems, tourists building an itinerary face a multi-objective optimization problem, in which the perceived utility of visiting a particular attraction is only one of the factors. However, most itinerary building approaches only try to maximize the sum of the perceived utilities for the chosen touristic attractions. The algorithm presented in this chapter combines multi-objective genetic algorithms and similarity measures to suggest highly distinct Pareto-optimal itineraries. Advanced heuristics have been used in order to provide solutions in near real-time.

The fourth chapter, **Collecting Data using Crowdsourcing,** focuses on the particularities of map based collaboration and especially on collecting information regarding touristic attractions using crowdsourcing. Although crowdsourcing offers the premises for easily acquiring large amounts of data, special measures have to be taken in order to assure a high degree of accuracy. The proposed approach combines flexible validation rules and artificial neural networks.

The fifth chapter, **Collaborative Solution for Itinerary Recommender Systems,** presents the architecture of the proposed itinerary recommender system and the approach used for building the client implementation. A special communication layer was added with the role of facilitating the communication with the client implementations using HTTP requests, Web Service Calls and Socket Connections. Thanks to the multiple communication methods supported by this layer, client side applications can be implemented using a wide array of technologies. The reference client implementation uses the latest web technologies including HTML 5 and Web SQL in order to support a wide array of mobile devices. By adapting to the capabilities of each device, the approach ensures an optimal and intuitive user experience. The resulting application can be either run directly from the browser or can be installed similar to classic native applications using a thin native wrapper.

The sixth chapter, **Further Improvements**, discusses several recent topics, that could be helpful for readers willing to improve upon the system presented in this book, including how to use semantic-web technologies to provide a common understanding of data between different systems, and how next generation architectures based on automatic and semi-automatic web service composition could be built. A multi-agent framework for web service composition framework that allows users to easily add new functionalities and access external information is described.

The seventh chapter, **Conclusions and Future Work,** summarizes the results included in the book and also presents possible future developments.

In order to validate the theoretic results presented in this thesis, the authors have developed an itinerary recommender system, called MoTripAssistent. The screenshots in this book present the application running either in the emulator or on a mobile device.

This book has been specifically designed and written for a broad audience, including both researchers interested in the field of recommender systems and practitioners willing to either develop a new or to improve an existing itinerary recommender system.

1.1. Geographic Information Systems

According to a broadly accepted definition, a Geographic Information System - GIS is composed of hardware, software and procedures to facilitate the management, manipulation, analysis, modeling, representation and display of georeferenced data to solve complex problems regarding planning and management of resources [NCGI10]. Since the first implementations in 1960, GIS has become an essential tool for both public administration and business due to its ability to store, retrieve, analyze and display huge volumes of spatial data.

Public administration can use GIS for transport and infrastructure planning, utilities design and operation, public safety, tax assessment, sustainable development and disaster management. Business can use GIS to better manage vehicle fleets, to analyze market areas, profile customers or provide advanced Location Based Services - LBS. Many GIS implementations are used in retailing, financial services, media, insurance and real estate sectors. Other application areas include natural resources, education and military [MAGU07]. The last years have seen an important increase in GIS usage by non-expert users.

Geographic Information Systems have been successfully applied for:

- **mapping locations:** digital maps are created using automated mapping, data capture and surveying analysis tools; besides the classic data collection techniques, new collaborative mapping approaches based on crowdsourcing have recently been developed [QIAN09]; [COTF09e] and [COTF11d] present a crowdsourcing approach for building a comprehensive database of tourist attractions;
- **mapping quantities and densities:** by mapping quantities and densities areas that meet specific criteria are identified; moreover, relationships between places are easier to analyze; an additional

level of information is thus provided as compared to simply mapping the locations; by mapping densities, the distribution and the areas with higher or lower concentrations are easier to identify; density maps display the number of features in an area unit, such as km^2; [DARD09] presents a method for creating tri-dimensional geology models applied for the historic center of Bucharest;

- **finding distances and routes:** GIS is used to calculate distances between geographic points taking into consideration the particularities of the relief; finding routes between two or more points, like in case of goods delivery, can also be accomplished [PAN08]; [COTF09e] and [COTF11a] present a hybrid genetic algorithm for route finding in public transport networks;
- **monitoring and predicting changes:** GIS is used to analyze changes over time in a certain area and also to anticipate future conditions; thus, geographic information systems are used for planning purposes and risk assessment [PIJA09].

First GIS applications were manly targeted for government and military use as they required expensive mainframe computers running UNIX based operating systems. Their use in business was limited due to high cost and need for specialized personnel. The situation gradually changed with the introduction in the late 1990s of desktop GIS applications like PARC Xerox and Map Server that could be run on personal computers.

Later on, starting from 2005, Web GIS applications like MapQuest, Google Maps and Microsoft Bing Maps [MICR11c] allowed regular users to access detailed maps using only a web browser. Figure 1.1 shows the use of Microsoft Bing Maps Web GIS for displaying spatial information in MoTripAssistent, the itinerary recommender system presented in this book. Besides offering access to huge amounts of spatial data using only a web browser, Web GIS also has the advantage that all the changes made on the server hosting the system are immediately available to all users. A disadvantage of Web GIS is that a connection between the users' device and the central server is constantly required. Many implementations heavily relay on raster images generated on the server for displaying geo-referenced data.

OpenLayers [OSGF16] is the best known open source JavaScript framework that allows the development of Web GIS applications. Commercial solutions, usually provided without any cost for non-intensive use, include Google Maps, Microsoft Bing Maps and HERE Maps [HERE16].

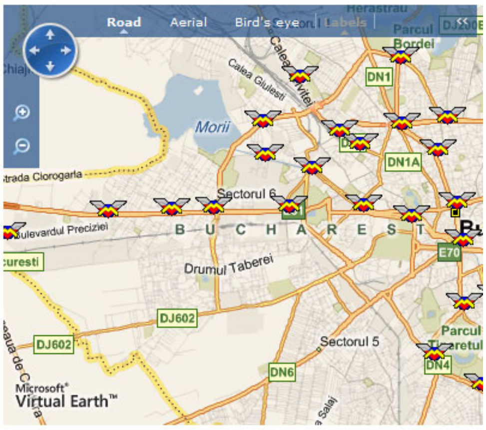

Figure 1.1 Microsoft Bing Maps Web GIS

Currently, Mobile GIS applications are becoming increasingly popular [DOYL10]. Figure 1.2 shows the use of Google Maps in the mobile version of the developed system. MoTripAssistent uses Google Maps API [GOOG11d] to represent spatial information across different mobile platforms.

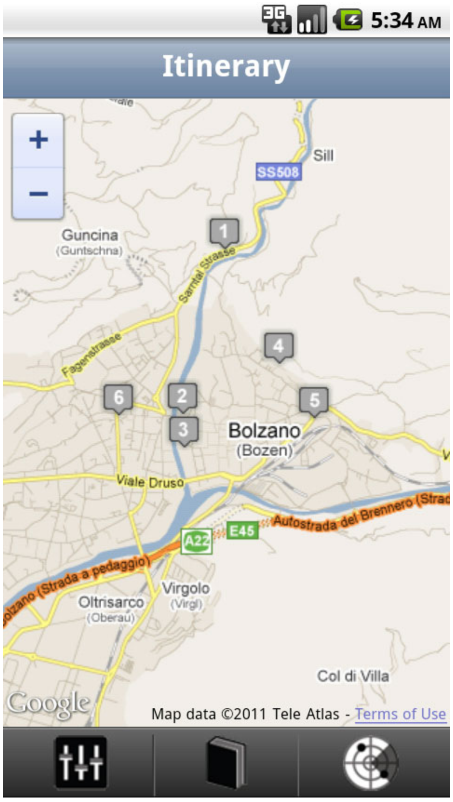

Figure 1.2 Google Maps on a mobile device

Custom markers were used to represent on the map the locations of the Points of Interest included in the generated itinerary.

Mashup GIS has rapidly expanded based on the new possibilities associated with Web 2.0. A map based application is considered a mashup if it either combines functionalities from several sources at a service level or if it aggregates data from multiple sources in a unified representation. Although the technologies used to build mashups are completely different than the ones used to add overlays in traditional GIS, the two approaches share a conceptual similarity as they both target the integration of multiple pieces of information. [POOR11] uses mashups to represent the level of air pollution for different substances in an easy to understand way.

MoTripAssistent, the itinerary recommender system presented in this book, combines information from multiple sources, through mechanisms such as web services, eXtensible Markup Language - XML or Really Simple Syndication - RSS, allowing easy intercommunication to other solutions.

From the architectural point of view, recent trends include integrating GIS with cloud computing, creating easily reconfigurable applications using automatic and semi-automatic web service composition and a rapid expansion of web and mobile GIS. Geographic applications have constantly evolved in the recent years from solutions based on strong integration towards increased flexibility based on the usage of web-services [MILL10]. Although initially restricted to professional use due to high costs and need for specialized personnel, GIS applications have gradually become more and more attractive for personal use.

The most important commercial GIS software vendors a are: ESRI with the ArcGIS suite, Intergraph with the GeoMedia suite, GE Energy, Bentley Systems with Bentley Map and Autodesk with Map 3D, Topobase and MapGuide.

Open source GIS software includes GRASS GIS originally developed by the U.S. Army Corps of Engineers, Integrated Land and Water Information System - ILWIS, MapServer developed by the University of Minnesota and MapGuide.

1.2. Location Based Services

The rapid technology progress characterized by communication improvement, enhanced portability and mobility enables software developers to implement online services targeted towards people on the move. Location Based Services - LBS are defined as applications that customize their content based on user’s geographic position and are considered a key component for building ubiquitous computer environments [YU09b].

An important step for the development of Location Based Services was the introduction of Web and Mobile GIS. Web GIS allows users to view maps and spatial data on any computer connected to the internet using just a web browser. Mobile GIS takes things even farther, allowing users to access geographic information and data while on the move.

Acknowledging the importance of Location Based Services, W3C has recently created the Points of Interest Working Group, having as purpose to develop technical specifications for the representation of "Points of Interest" information on the Web [W3C10b].

An important challenge for the success of any Location Based Services application targeted at regular users is whether it is accessible from the huge diversity of mobile devices owned by the users. This represents an important change in paradigm compared to classic GIS were the developer of the application had a tight control over the platforms on which the developed system was deployed.

The approach used for the development of the reference mobile client implementation of the proposed itinerary recommender system uses the latest web technologies in order to provide both portability among many different mobile platforms and a good user experience. The application is compatible with several platforms and can be installed using a thin native wrapper. Figure 1.3 shows the menu from which the user can select the desired location. Closest location is determined based on the GPS coordinates determined by the mobile device.

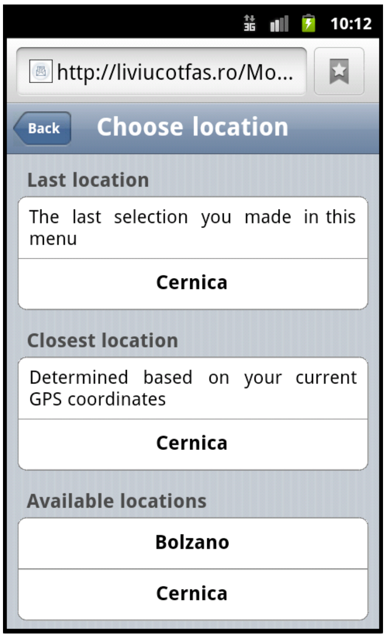

Figure 1.3 Location selection menu

The desktop version is also implemented using advanced web technologies and is compatible with all modern browsers. The proposed system is also implemented in a collaborative manner, allowing users to add or modify information. Maps are used to display data from external sources, user generated content and transport routes in an easy to understand way. This allows the system to incorporate weather information, georeferenced information from sources such as Flickr or Wikipedia, as well as suggestions relating to attraction points or available city services.

Based on the factor that initiates the interaction between the system and the user, existing LBS applications are classified as:

- **Trigger-based** - provide information without a specific request from the user. Such services typically take the form of Location Based Advertising – LBA [RASH08]. [COTF09d] describes such a system and provides an architecture design that uses the geo-spatial functions available in Microsoft SQL Server.
- **User-requested** - provide information on demand. This category of services includes finding nearby Points of Interest – POI [ZHU09] like restaurants, shops and cinemas, finding routes for personal cars or in public transport networks, vehicle tracking applications. In [COTF09e] the author proposes a collaborative LBS system that allows finding routes in multi-modal, multi-operator public transport networks.

Given the importance of geo-processing and the extensive usage of maps, most LBS applications are in fact full-featured Geographic Information Systems made available through mobile devices. As opposed to classical GIS applications however, GIS-LBS applications usually need to dynamically handle heterogeneous data from multiple sources and sensors such as weather information, traffic information and pollution information. Moreover, they must be accessible from a wide range of mobile devices, making interoperability an important issue [COTF09c]. A solution for addressing data heterogeneity is the use of ontologies. Concepts such as museums, parks or castles can be represented using an objective ontology. Products and services offered by the various businesses present on the LBS platform can be described using a product ontology. A payment ontology can be used for concepts related to types of payment. [SEBA10] uses an ontology derived from the Art & Architecture Thesaurus Ontology in order to classify the Points of Interest. A comparison of existing ontologies for the tourism domain is presented in [PRAN07].

1.3. Evolution from Monolithic to Distributed Architectures

The first enterprise applications used a monolithic architecture, making them both hard to maintain and difficult to expand with new features. An important step in architecture design was made with the client-server model, which was further enhanced with the appearance of 3 and N tier architectures. The introduction of web service paved the way to creating highly distributed SOA applications. Services replaced the server module in the earlier architectures, by making functionalities available over the internet using a standardized set of protocols and technologies. An important accent is put nowadays on patterns like Semantic Web Service Composition – SWSC [SMEU10], Mashups, SaaS and Cloud Computing [GOVE09]. The main steps in the evolution of software architectures are presented in Figure 1.4.

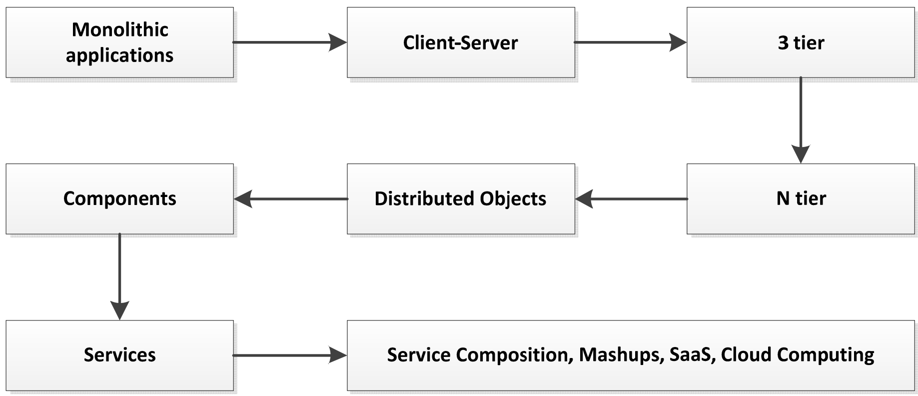

Figure 1.4 The evolution of software system architecture. Modified from: [ADOB08]

The constant demand for better interoperability between systems led to a rapid development of new standards, technologies and paradigms. Interoperability is defined by IEEE as the ability of two or more software systems or components to exchange information and to use the exchanged information. It is becoming a basic requirement for distributed information systems and also critical to GIS and to geospatial web services.

An important step towards better interoperability was made with the introduction of web services, which are considered the building blocks of Service Oriented Architectures - SOA. SOA applications are usually built using web services, although other technologies such as Jini [JINI10] or CORBA [CORB10] can also be used. SOA was defined by OASIS Reference Model for Service Oriented Architecture as a paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains.

Web services are defined by W3C as software system designed to support interoperable machine-to-machine interaction over the network. Also, [W3C11g] defines web service as a software system designed to support interoperable machine-to-machine interaction over a network, described using Web Services Description Language - WSDL. Web service specifications are described using eXtensible Markup Language - XML for describing the message format, the Simple Object Access Protocol - SOAP communication. Available operations are described using Web Service Description Language - WSDL. Publishing and discovering capabilities are provided using Universal Description, Discovery and Integration - UDDI.

WSDL provides an XML model for describing the public interface of web services. The description is composed from an abstract and a concrete section. The abstract definitions of ports and messages are separated from their concrete use or instance, in order to facilitate reuses. It provides the necessary information so any potential consumer can evaluate if the service is appropriate and configure itself to consume the service. WSDL offers only a syntactic description and cannot therefore be used directly for automatic Semantic Web Service Composition - SWSC OWL-S is a standard proposed by W3C, built on top of OWL. It is supposed to replaces the former DAML-S and adds a semantic description to web services, besides the syntactic one provided by WSDL [TALA09].

UDDI allows service providers to publish service descriptions in a service registry and service consumers to discover services in a service registry according to their service descriptions, usually described in WSDL. As it was defined before SWSC become an important research issue, UDDI does not allow semantic discovery. UDDI is based on XML and provides three types of information for each published web service:

- white pages - contains address, contact and other identifiers;
- yellow pages - include business categorization based on taxonomies;
- green pages - comprise technical information of the web service along with a link to its WSDL description.

SOAP is a protocol that allows XML-based messages to be exchanged between web services over a computer network. The SOAP message can be seen as an envelope that contains:

- a header with useful information to interpret the data;
- a body containing the actually exchanged data.

The web service interaction model is composed from the information model and the behavior model. The behavior model illustrates the message exchange and invocation patterns. The Information contains the details of the input/output data and can be described in terms of both structure and semantics.

Given the importance of interoperability between various systems, Open Geospatial Consortium - OGC has defined a sub-set of standardized interfaces for geospatial web services shown in Table 1.1. The specifications have been adopted by all important GIS vendors including ESRI [ESRI11d] as an important part of their interoperability strategy.

Table 1.1 OGC Geospatial Web Services. Source: [GRAN09]

Service name	Service Description
Web Map Service - WMS	Dynamically produces spatially referenced maps of client-specified criteria from one or more geographic datasets, returning pre-defined maps in an image or graphics format like png, jpeg and gif.
Web Feature Service - WFS	Allows clients to filter and retrieves vector representation of geospatial features and feature collections.
Web Coverage Service - WCS	Retrieves client-specified coverage including satellite images, digital aerial photos and digital elevation data.
Catalog Service - CSW	Retrieves object metadata stored that meets specified query criteria.
Web Terrain Service - WTS	Dynamically produces perspective views from geographic feature and/or coverage data, returning pictorial renderings of data in an image or graphics format.
Web Coordinate Transformation Service - WCTS	Transforms the coordinates of feature or coverage data from one Coordinate Reference System - CRS to another, including conversions and rectification.
Coverage Portrayal Service - CPS	Dynamically produces pictorial renderings in an image or graphics format of a coverage subset dynamically retrieved from a Web Coverage Service - WCS.

By relying on the standard interfaces defined by OGC, geospatial web service chains can be defined in which each service provides a clearly defined functionality for information manipulation, extraction, computation or representation. However, automatically defining and compositing such flows to satisfy the demands of users presents a string of unsolved problems in terms of both technical matters like identification and networking of services, but also practical matters such as management of significant data flows, handling of large data volumes and usage of different, potentially incompatible file formats. Semantic tagging, using OWL-S and WSMO is thought to be able to overcome such issues, but research in the field is still in rather early stages.

In order to provide a better syntactic interoperability OGC has also introduced a new geographic format called GML that uses an XML Schema file to formally constrain instances of GML data to the set structure of XML items. The GML format is considered an ideal candidate to be used with other web service elements based on XML such as SOAP and WSDL [ADOB08]. The KML format uses an XML representation to store geographic annotations and visualizations for two-dimensional maps and three-dimensional geographic representations. KML is an international standard of the Open Geospatial Consortium.

1.4. Itinerary Recommender Systems

Without any support from a system, a user manually building an itinerary, should spend a lot of time prior to the trip, searching for information regarding the tourist attractions, reading facts about them and designing possible itineraries by taking into account factors such as opening hours, visiting durations, distances and available means of public transport. Conversely, itinerary recommender systems not only assist users to schedule the initial itinerary, but also to easily revise it during the trip. This is especially useful, as in many situations, either the visiting durations will be different than the predicted ones, or the tourist might decide to change some of the previously selected Points of Interest - POI. Moreover, such applications can also provide rich multimedia content for the visited locations like shown in [BIUK08].

Prior to their application in tourism, recommender systems have been successfully used in e-commerce for suggesting products and services to users [RICCI11]. One of the first attempts of applying recommender systems in tourism is represented by the Cyberguide [ABOW97] project, in which the authors develop a mobile application targeted to model the tour guide activity. The paper also acknowledges the importance of context, but limited only to location awareness. The Guide system presented in [KEIT02] provides POI recommendations based on the user's profile and context. The authors also present an evaluation of the visitor's experience with the system. A drawback of the approach is represented by the amount of information required from the user in order to customize the tour. Guide offers users the possibility of creating a tour by selecting objectives, presenting their domains of interest and preferences. The restrictions that the system takes into account refer to the schedule of objectives, the best visiting intervals. The tour can be changed dynamically, based on the actual time the tourist already spent at some attractions or on weather conditions. The user profile is also continuously updated, using information from previous trips. [SEBA10] proposes an approach based on intelligent agent interaction and ontologies in which the categories of Points of Interest are described using an ontology derived from the Art & Architecture Thesaurus Ontology. The system expands upon a previous implementation presented in [SEBA08] that uses a taxonomy to represent the users' preferences for different types of tourist attractions. The concepts in the taxonomy are arranged in hierarchical structure connected using "is-a" relationships. The leaf nodes represent the locations that can be recommended. Edges that connect locations and concepts are associated a value in order to reflect the degree of interest, which is updated based on the user's choices. Items are also associated an acceptance indicator, that shows how many times the recommendation was accepted.

Figure 1.5 The main functionalities of itinerary recommender systems

Several reviews of existing solutions are found in [GARC09b] and [KRAM07]. In most approaches, the process of generating itineraries and guiding users is divided in three distinct steps, which also represent the main functionalities of itinerary recommender systems (Figure 1.5):

- **Building the list of tourist attractions** represents the first step in which the POIs that will be included in the itinerary are selected; in the simpler approaches like the one presented in P-Tour [MARU04], the user is requested to select the desired POI from a list without receiving any recommendations from the system; on the opposite side, other systems like [VANS09] opt for a completely automatic approach in which the user has no control over the selected Points of Interest – POI; **MoTripAssistent** implements an advanced recommender algorithm that takes into consideration the user's previous ratings, the ratings of other people with similar profiles as well as context in order to suggest interesting POI; context is taken into consideration as it can have a big impact on the perceived value of the recommendations; in order to offer a high degree of flexibility, the proposed system combines the manual and automatic approaches for building the list of POI, by both allowing the user to manually select locations and using an algorithm to choose the best locations for the remaining time;

- **Building the tourist itinerary** is an extension of the Traveling Salesmen Problem which is known to be already NP-hard; existing approaches either only try to maximize the sum of scores associated to the visited POIs [VANS11], or if multiple criteria are taken into consideration, they are combined using a simple weighted sum; [MARU04] presents an approach based on using weighted sum that takes into consideration the preference expressed by the user and the distance between the Points of Interest - POI; **MoTripAssistent** takes into consideration that similar to many other real world problems, itinerary building also implies several optimization criteria; instead of choosing a simple, but less relevant approach like weighted sum, a multi-objective algorithm that offers several Pareto-optimal solutions was developed;
- **The guidance during the tour** is important as it was shown in the user studies performed by [KRAM07], that tourists typically modify the list of POI during the itinerary. Therefore **MoTripAssistent** allows the user to easily change the itinerary at any moment during the tour. Moreover, the system constantly monitors the position of the users during the itinerary and allows them to rate visited POI.

Table 1.2 presents a comparison between the proposed solution and existing Itinerary Recommender Systems. As shown in the table, the proposed system:

- allows the user to choose between multiple Pareto-optimal itineraries generated in near real-time;
- takes into consideration several optimization criteria in order to better match the user's preferences;
- provides advanced recommendations that take both context and temporal preferences into consideration.