Towards spatial web personalization

Yanwu Yang

Ecole Nationale Supérieure d'Arts et Métiers

Institut de Recherche de l'Ecole Navale (EA 3634)

THÈSE

Présentée pour obtenir le grade de

DOCTEUR

de

L' Ecole Nationale Supérieure D'Arts et Métiers

Spécialité : Informatique

par

Yanwu Yang

TOWARDS SPATIAL WEB PERSONALIZATION

---

soutenue le 18 juillet 2006 devant le jury composé de

M.	Y. Kermarrec,	HDR Enseignant chercheur à l'ENST Bretagne, Brest,	Président
Mme	M.A. Aufaure-Portier,	HDR, Maître de Conférences à Supelec, Paris,	Rapporteur
M.	T. Tanzi,	Professeur à l'ENST, Paris,	Rapporteur
M.	M. Farine,	Professeur à l'Ecole navale, Lanvéoc-Poulmic	
M.	C. Claramunt,	Professeur à l'Ecole navale, Lanvéoc-Poulmic	

Acknowledgement

I would like to thank my thesis advisor, Professor Christophe Claramunt, for his continuous advice, guidance, and encouragement during this thesis. He proof-read the whole thesis, made numerous discussions on some ideas, cleared up many ambiguities. His profound knowledge and kindness did inspire me a lot.

I want to thank my thesis reviewers for having accepted to evaluate my thesis and their valuable comments for the improvement of the final version.

Also thanks to some colleagues in Naval Academy Research Institute, France. They were kind to me and helped a lot during my PhD research.

Finally, I also wish to gracefully thank my parents for everything they did for me. I dedicate this thesis to my parents.

To my mother and father

with love and gratitude

TOWARDS SPATIAL WEB PERSONALIZATION

ABSTRACT: In the past few years, spatial information and services have proliferated on the Web, due to the fact that most of our daily activities are related to the spatial dimension. The user communities involved in spatial web services are essentially diverse, still in an expansion and transformation with constantly increasing number of user and applications. This opens many research challenges, such as the elicitation of user's interests and preferences and customization of information services on the spatial Web. This PhD research proposes an integrated framework for user modeling and preference elicitation, and personalization services on the spatial Web. The framework identifies personalization services and a semantic user model for spatial web applications. These two components communicate information and knowledge about the user through inter-process communications. The personalization services are based on three mechanisms: the Bi-directional Neural Associative Memory, user-centric spatial proximity and similarity measures, image schemata and affordance concepts. A web-based user interface is integrated with these components, and offers a spectrum of personalized search strategies and a hybrid personalization engine. The user model employs expressive description logics to describe assumptions about the user and to infer implicit user features from user's descriptions as required by an application system. An application scenario in the tourism domain and a Web-based Java prototype provide an experimental validation of the research framework and identified personalization techniques.

Key Words: user preference elicitation, the spatial web, spatial web personalization, semantic user model

VERS LA PERSONNALISATION D'INFORMATION SPATIALE

SUR LE WEB

RESUME: La mise à disposition d'information et de services spatiaux a récemment proliféré sur le web dans la mesure où la plupart de nos activités quotidiennes sont géo-réferencées. Les communautés d'utilisateurs de services spatiaux sur le web sont de plus larges et variées, en constante expansion et transformation avec une augmentation constante des gammes d'applications proposées. Cette profusion d'applications entraîne un nombre important de problématiques de recherche, et notamment celles liées à l'identification des intérêts et des préférences de l'utilisateur, afin d'adapter les services délivrés aux besoins du client. Cette recherche propose une architecture intégrée de modélisation de profils d'utilisateur et d'approximation de leurs préférences, et de mise à disposition de services personnalisés orientés vers l'information spatiale. L'architecture proposée se compose d'un service de personnalisation et d'un modèle sémantique orienté utilisateur. Ces deux composants communiquent des informations sur l'utilisateur par des processus interactifs. Ce service de personalisation est basé sur trois principes : la mémoire associative neurale bi-directionnelle, des mesures contextuelles et spatiales orientées-utilisateur de proximité et de similarité, des schémas d'image et des concepts d'affordance. Ces concepts sont implémentés à partir d'une interface utilisateur qui intègre les différents composants identifiés, et offre un éventail de stratégies personnalisées de recherche, et un moteur hybride de personnalisation. Le modèle d'utilisateur utilise des logiques expressives de description pour caractériser les différentes catégories d'utilisateur, afin d'adapter les besoins d'utilisateur aux exigences d'une application. Un scénario dans le domaine du tourisme et un prototype Java réalisent une validation expérimentale de notre recherche à partir de techniques de personnalisation.

Mots clés: Elicitation de préférences utilisateur, Web spatial, personnalisation d'information sur Web spatial, modèle sémantique utilisateur

Contents

List of figures

Chapter 1 Introduction

1.1 The World Wide Web

The World Wide Web (Web) is a collection of servers on the Internet accessible through the Hypertext Transfer Protocol (HTTP). Bush's Memex machine (Bush 1945) is regarded as the harbinger of the notion of hypertext (Nyce and Kahn 1991), browsing and trails on the Web as we know today. A Website is a collection of Web pages whose objective is to diffuse data and information on a given subject to a large audience community. Since its advent, the Internet has growed from a small research project of local network to an invaluable and popular information space. Initially, Berners-Lee proposed a project based on the concept of hypertext, to facilitate the sharing and transfer of information among researchers disseminated throughout the world[1]. He took an opportunity to combine the hypertext concept and the Internet to create the World Wide Web, and designed and built the first Web browser and editor, and the first Web server. The Web provides a novel way for information services not limited to researchers, techniques, but turns into an important information medium for a large portion of the population, particularly in developed countries. First, it provides a vast repository to distribute, deliver, and share different kinds of information from different domains, organizations and to various user communities. Secondly, it's a new medium for personal, intra- and inter-organization communication. A notable milestone is the resurgence and the continued popularity of usenets and Web-based emails in the past decade. Thirdly, it's also an efficient environment for many application domains such as e-business and distributed decision-making systems and processes.

The Web has made a major impact on our society and everyday lives (Lesk 1997, Lynch 1997), influenced people's social style, and even the role of space and time in social networks. Conceptually, the Web is considered as an "inherently social network, linking people, organizations, and knowledge" (Wellman 2001). Despite its incredible successful diffusion, the Web still requires novel solutions for online searching and Web information retrieval (Baeza-Yates and Ribeiro-Neto 1999). Web information retrieval is a generic engineering domain that can be considered at different levels, from the development of user interfaces, indexing and optimization mechanisms, to data communication constraints (Jansen and Pooch 2001). With a virtual network of content and hyperlinks instead of a centralized architecture (Kleinberg and Lawrence 2001), information on the Web increases exponentially due to its decentralized and

[1] http://en.wikipedia.org/wiki/Tim_Berners-Lee

distributed nature, and hyperlink component. Nowadays, the Web constitutes a large repository of information where the range and levels of services offered to various user communities are expected to increase and be dramatically enriched in the next few years.

The Web leads to the availability of unprecedented large volumes of data, and novel interactive ways of storing, exchanging, searching and mining multi-dimensional information. Web engineering reveals a specific characteristic that makes the design process difficult to delineate as any new Website is initially connected to the Web as a whole. This is due to the unbounded nature of the Web. Moreover, and due to the extraordinary large dimensions of the Web information space, finding and delivering information to the appropriate people from relevant sources in the most appropriate way are becoming an important challenge that, if not addressed, will make the whole Web information space almost useless. Nowadays, end users have to extract meaningful information from the plethora of accessible Web resources. This leads to many problems such as information overload, irrelevant information supply and “lost-in-navigation”.

In the last few years, a lot of researches have been oriented to the development of search engines (Kleinberg 1999), analysis of Web communities (Greco *et al.* 2001), statistical analysis of Web content, structure and usage (Madria et al. 1999, Tezuka et al. 2001), and Web information retrieval (Belkin and Croft 1992, Baeza-Yates and Ribeiro-Neto 1999). For instance, Web information systems use retrieval models based on word frequency to identify relevant Web documents. However, the degree to which current Web search engines and information retrieval models can interpret and infer user’s information needs is limited. This leads to a frequent inadaptation of search results to user’s requirements. In order to improve the way that information is delivered to the user, many Web mining techniques have been developed to filter out irrelevant information. This also reflects a conceptual difference between Web-authors’ intentions and Web-users’ understandings of hypertext documents (Borgman 1986, Suchman 1987, Perugini and Ramakrishnan 2002), and the complexity of information resources and services on the Web. This outlines two distinct roles on the Web: Web authors and Web consumers (Bodner and Chignell 1999).

Filtering irrelevant information is a key issue, but Web end-users also often suffer from information overload, (Levene and Wheeldon 2003). This is reinforced by the fact that the Web favors uncontrolled navigation, and resulting situations where users are often lost in the “hyperspace” when browsing large volumes of information even if they are to some degree relevant. This leaves the user in a difficult position to find out relevant information, and also to identify the most comprehensive answers to a given information search task. This is due to many factors such as semantic ambiguities and imprecisions in the definition of a search query, inadaptation of keyword-based searching systems and the inherent difficulty of evaluating user’s intentions. Amongst many solutions, personalization of information services on the Web offers a promising solution to alleviate these problems, and to customize the Web environment according to user's information needs, interests and preferences.

1.2 Web personalization

The domain of Web personalization generates many research and technical challenges for the information engineering science. Amongst many challenges still open to the Web engineering research community, there is an urgent need for the design of intelligent interfaces that monitor user interactions with Web systems, in order to approximate as much as possible user profiles and preferences (Shahabi and Chen 2003). This is essential to adapt Web structures and resources to user's expectations and needs. Another issue to explore is the development of unsupervised mechanisms that could implicitly help the users in the manipulation and analysis of information on the Web. This implies to design and develop appropriate mechanisms to approximate user's intentions and preferences in order to guide and constrain information retrieval processes.

Web personalization attracts many research interests from different communities and domains such as user modelling, information retrieval, machine learning, human computer interaction, distance learning, intelligent tutoring systems, and cognitive sciences. Over the past several years, there has been a wide interest of the Web engineering research community to address user preference elicitation and personalization on the Web. Web personalization is a relative young research field started in the early 90s. A lot of proposals, Web agents and personalization systems have been introduced, such as the landmark systems Letizia (Lieberman 1995) and WebWatcher (Armstrong *et al.* 1995, Joachims *et al.* 1997), which can be considered as the first generation of recommendation and Web personalization systems. The first generation systems typically use collaborative filtering and content-based algorithms to infer user's interests and information needs, and to deliver personalized information. These systems are particularly employed in Web browsing and E-Commerce business, to help the user to locate Web pages or products she/he would like to visit or purchase.

In the late 90s, several research works attempted to extend collaborative filtering and content-based approaches. Personalization algorithms constitute the second generation of hybrid approaches that combine collaborative filtering and content-based filtering (Baudisch 1999, Claypool *et al.* 1999, Melville *et al.* 2002) and those based on Web usage mining (Mobasher *et al.* 2000, Pierrakos *et al.* 2003, Eirinaki and Vazirgianis 2003). The current generation of Web personalization systems applies semantic knowledge in order to improve personalization performance and provide the user more intelligent and refined assistance and information services. Personalization algorithms can be classified into two categories: ontology-based (Dai and Mobasher 2002, Gauch *et al.* 2003 Dai and Mobasher 2005) and knowledge-based systems (Burke 1999, Towle and Quinn 2000). Both refine personalization results with

Artificial Intelligent techniques. The former emphasizes on ontology and knowledge representation, and the latter on intelligent agents and learning.

From an application perspective, main research contributions can be categorized as follows:

- Adaptive Web techniques whose objective is to create and adapt Web sites by learning from user access patterns, and synthesize Web indexes (Perkowitz and Etzioni 1997, 1998, 2000);
- Intelligent agents (Pazzani and Billsus 2002) and Web tour guide agents such as Letizia (Lieberman 1995) and Web watcher (Armstrong *et al.* 1995, Joachims *et al.* 1997) that help the user to explore the Web;
- Online recommender systems that suggest and deliver information and products to customers (Goldberg *et al.* 1992, Resnick and Varian 1997), such as the ones developed and applied in usenet news and E-Commerce;
- Web personalization systems that improve Web content presentation using knowledge extracted from user's personal information and behaviours on the Web (Mulvenna *et al.* 2000, Mobasher 2000).

Agent-based techniques and recommendation systems provide promising research issues on the elicitation and personalization of information services on the Web. They share the fact that they are to a large degree oriented to the most standard form of information delivered on the Web, that is, textual and indexed information on the Web. One can expect from the next generation of Web agents and personalization systems integration of more complex information such as image, spatial, and dynamical data, and more advanced and intuitive services.

1.3 Towards spatial Web personalization

Conventional Web personalization approaches such as collaborative and content-based filtering, and their derivatives are successful in conventional Web applications. However, they are not adapted to the semantics and complexity of spatial information as exhibited on the Web. To the best of our knowledge, a few works are currently oriented to personalization techniques applied to the modelling and manipulation of spatial information available and embedded on the Web. Spatial information in the context of our work denotes geo-referenced information that is implicitly or explicitly related to geo-located objects. As a large proportion of Web resources can be mapped to some degree to geo-referenced entities (Winter and Tomko 2005), Web personalization mechanisms should take into account the spatial dimension, its semantic, and the large data processing opportunities offered. Therefore, personalization of spatial information on the Web should consider the spatial properties and relationships exhibited and embedded in Web documents. Integration of the spatial dimension in personalization processes extends the dimensions to cover the need for a better understanding of user's profiles and information needs, and the range of processing capabilities for Web-based information. Possible services covers a wide range of applications such as, assisting tourists to travel in a given city, or categorizing users according to the place from where they interact with the Internet.

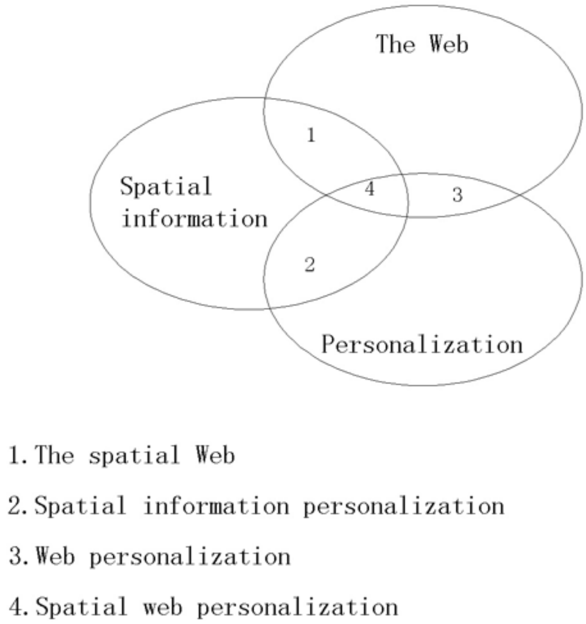


Figure 1.1 The Web, personalization and spatial information

1.4 Research scope

The objective of this research is to explore and study the concept of Web personalization when applied to spatial information, and to evaluate to which degree this should enrich the engineering of a Website with embedded spatial information, and the way information should be delivered to the end users interacting with the Web. Hereafter in the report, Web personalization when applied to spatial information is denoted as "spatial Web personalization".

Several research components constitute our research domain: Web personalization, the spatial Web, spatial information personalization, and spatial Web personalization (Figure 1.1). We consider an inter-disciplinary research approach as integration of these components that cover a wide range of scientific domain, from knowledge representation to ontology and human-computer interfaces. We also believe that promoting the diffusion of spatial information might favour knowledge sharing and interpersonal relationships (Haffernick *et al.* 1997, Gallagher and Golde 1999). Spatial Web personalization can be viewed as an inter-disciplinary research field among the Web, personalization and spatial information. The scope of this PhD research is to make a preliminary investigation of Web personalization techniques to manage and deliver spatial information services on the Web according to user's interests and preferences.

1.5 Thesis outline

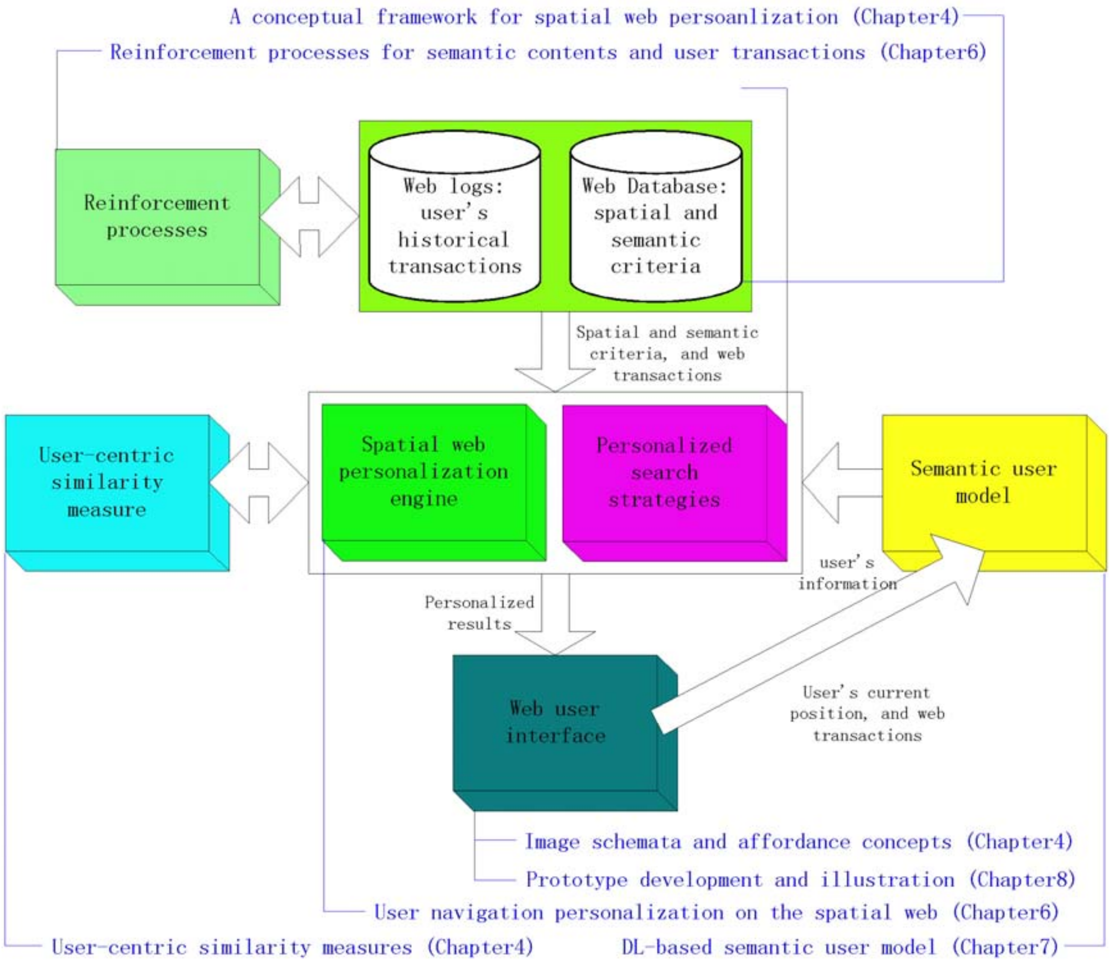


Figure 1.2 Thesis outline

The research work introduces a generic framework, where user models and profiles, inference rules, a personalization engine and similarity measures are developed to personalize spatial information services on the Web. In order to enhance and support Web search and personalization results, ontology-based techniques are used

- to organize and manipulate spatial entities with respect to a given application domain;
- to integrate syntactic and semantic properties when evaluating spatial proximities and similarities between spatial entities;
- to design a user model and infer user features relevant to a given application.

The framework identifies personalization services and a semantic user model. The two components communicate information and knowledge about the user through inter-process communications: "*tell*" and "*ask*" operations. Personalization services are based on three components: the Bi-directional Neural Associative Memory

(BNAM), image schemata and affordance concepts, and user-centric spatial proximity and similarity measures. Within the framework, a spectrum of personalized search strategies based on the BNAM, and a personalization engine supported by Markov chains are introduced. A description logics based semantic user model is developed in order to infer domain-dependent user features, e.g., interests and preferences, from user's personal information. User features are triggered by personalization components to deliver information services tailored to a given user.

The outline of the PhD thesis is presented in Figure 1.2. The remainder of this report is organized as follows:

Chapter II surveys the state of the art in the field of Web personalization.

Chapter III describes the main principles of spatial Web personalization, motivates and discusses our research issues.

Chapter IV introduces a framework for spatial Web personalization. It is based on three components: the BNAM architecture, image schemata and affordance concepts, and user-centric spatial proximity and similarity measures. The framework is used to manipulate and deliver candidate spatial entities according to user's interests and preferences.

Chapter V: presents a range of personalized search strategies on top of the research framework. These search algorithms combine different user preference indexes, spatial proximities and semantic similarities.

Chapter VI proposes a spatial personalization engine based on Markov chains to provide customized spatial information services to the user.

Chapter VII applies description logics to build an ontology-based User Model Knowledge Base (UMKB). The UMKB infers domain-dependent user features from explicit user information, and communicates with personalization components.

Chapter VIII introduces the implementation achieved so far as experimented in the tourism domain and some experimental evaluations.

Chapter IX concludes the PhD research and discusses research perspectives.

Chapter 2 Web personalization: state of the Art

2.1 Introduction

Web information retrieval collects Web documents that best satisfy user's needs for data and information. In particular, Web information filtering removes irrelevant contents from large amounts of data according to some predefined criteria. These criteria include similarity and relevance parameters, based on descriptions of individual or group information needs and preferences (Belkin and Croft 1992). Applying semantic-based criteria and searching encompass at least two complementary research challenges to address: development of query and search engines, and design of intuitive interfaces that can satisfy user communities that are diverse in terms of expectations and capabilities. This requires novel methodological and conceptual principles for the approximation of user's profiles and intentions, and adaptive information services. This also implies the development of models and algorithms for user modelling and preference elicitation, and personalization mechanisms to provide Web services customized to the user's information needs.

In the past decade, Web personalization researches and applications have provided several achievements on the improvement and optimisation of Web searching and recommendation taking into account user's interests and preferences. An application area that has been particularly investigated is the one of E-Commerce that has been a privileged personalization application domain on the Web, with many successful examples developed such as the well-known Amazon system. Personalization techniques are expected to overcome information overload, remove irrelevant information supply, and increase the utility and user satisfaction by providing the user with accurate and effective services tailored to her/his specific needs (Riecken 2000). Amongst many personalization tools and adaptive services, recommender and Web personalization systems are the most successful application scenarios developed so far (Shahabi and Chen 2003).

The remainder of this chapter is organized as follows. Section 2.2 discusses personalization techniques developed by the Web research and engineering communities. Section 2.3 introduces user modelling and preference elicitation as the kernel for Web personalization. Section 2.4 gives a survey on Web personalization techniques and algorithms, including classification, optimisation strategies and machine learning oriented approaches. Section 2.5 introduces main Web personalization tools and systems developed so far. Section 2.6 presents several emerging research perspectives on Web personalization. Finally, Section 2.7 concludes the chapter.

2.2 Web personalization

2.2.1 Personalization components

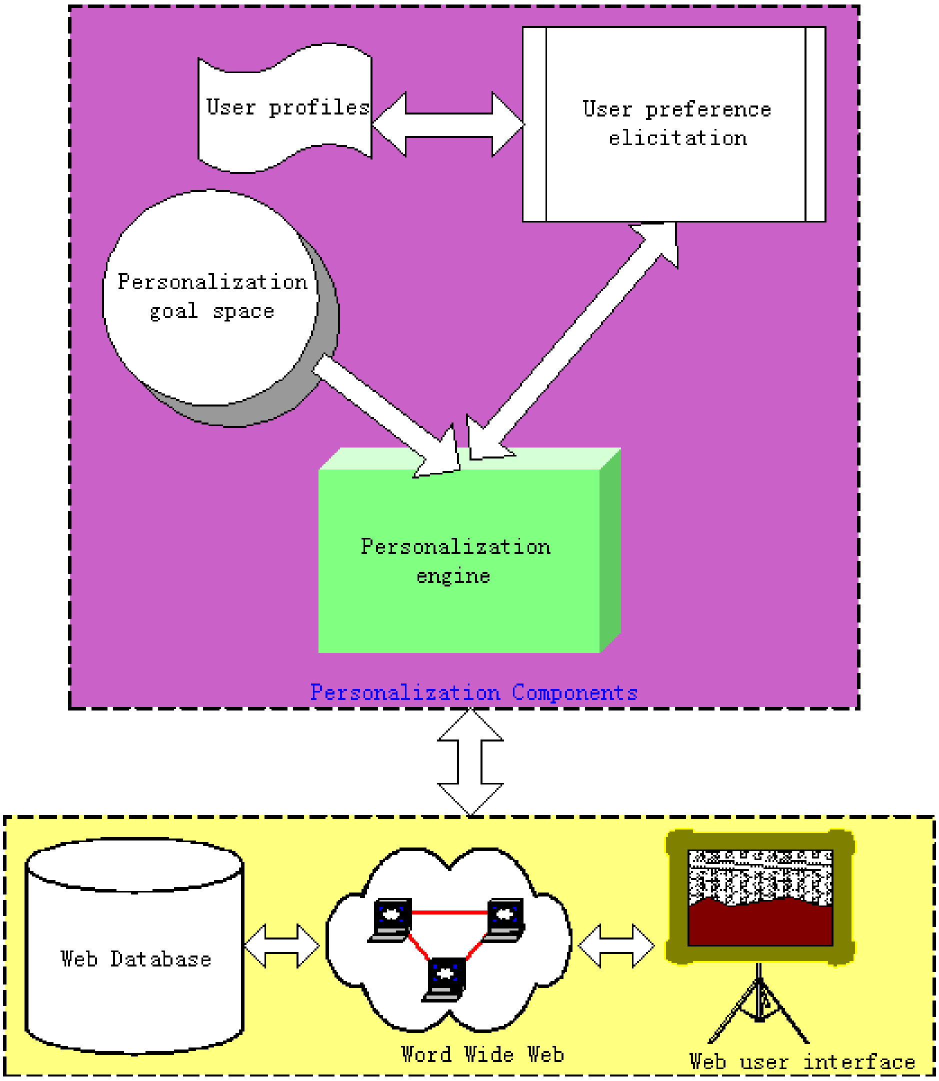


Figure 2.1 Personalization components

A personalization technique usually consists of three elementary components: a *personalization goal*, a *user preference elicitation process* and a *personalization engine* (Figure 2.1).

A *personalization goal* is generally considered as positive, and its objective is to increase the system utility and user satisfaction. There usually exists one specific goal, or several goals that together constitute a goal space. Conceptually, a goal space can be considered as an independent or interdependent n-dimensional space. Resolving qualitatively and numerically such a goal space implies to apply analytical strategies such as multi-criteria analysis in order to assist decision-making processes (Tan and Pearl 1994). For example, tourism services aims to provide attraction and travelling information, e.g., "at the right place, at the right time, to the right user".

User preference elicitation over a given domain knowledge requires either observing user's choice behaviors, or directly interacting with the user with pre-defined questions. The range of techniques used varies from the implicit tracking of user actions to explicit user feedbacks on the information provided. Evaluating user preferences is either derived by explicit information such as direct user feedbacks, keyword-based evaluation of user's interests, or implicit user feedbacks such as analysis of reading times, frequency of document downloads and page browsing (Oard and Kim 2001, Kelly and Teevan 2003).

As far as a personalization goal is explicitly quantified and a user preference elicitation process defined, a *personalization engine* encompasses a series of personalizing activities that provides services to a given end user. Personalization engines are commonly made of a set of coordinated computational components, taking user profiles and logs as inputs, and as outputs a series of information items (e.g. Web pages) that might be of interest to the user.

2.2.2 Multi-disciplinary issue

Web personalization can be viewed as an inter-disciplinary issue related to several research domains including user modelling and preference elicitation, social networks, Web usage mining, and human computer interactions (Figure 2.2). The design of Web personalization systems has been influenced by these disciplines:

- **Social networks:** Perugini and his colleagues made a survey on Web personalization from a *connection-oriented* viewpoint (Perugini *et al.* 2002). They argue that recommendation processes and engines should not be delivered within a vacuum, but that they own a social component, and intend to connect people directly or indirectly. Interactions and associations on the Web could be modelled using social network navigation metaphors and structures from which emergent social properties can be studied (Kleinberg and Lawrence 2001, Wellman 2001).
- **Web data mining:** Another aspect of Web personalization emphasizes the Web usage mining component (Pierrakos *et al.* 2003, Eirinaki and Vazirgianis 2003, Mobasher 2004). These approaches extract information from Web logs recording user's behaviours on the Web. They support inference and categorization of behavioural trends through data mining techniques.

- **Human and Computer Interaction:** (Perugini and Ramakrishnan 2002) studied personalization at the *interaction* level, in order to personalize the way users access an information system, and how such a system can encourage and foster interaction. Personalizing Human Computer Interactions leads to address the mismatch problem between Web authors and end-users. This implies to consider a conceptual dimension in the understanding of information flows and human mental models, in order to customize the way that Web interfaces deliver and present information to the user.

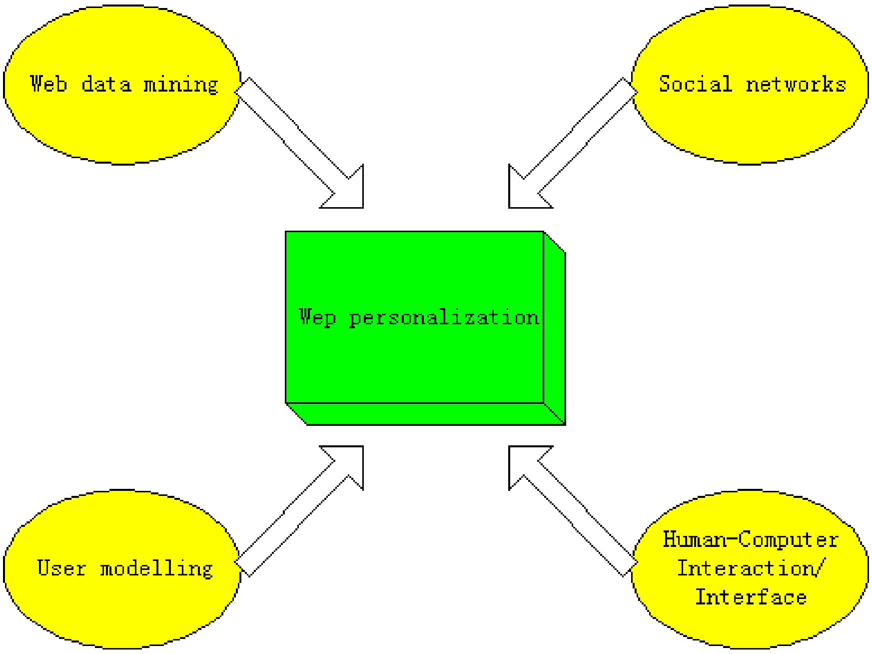


Figure 2.2 Multi-disciplinary aspects

These disciplines are complementary per nature, although each focuses on some specific aspects and dimensions of Web personalization, this being to some degree dependent on the research fields from where they originate. Together, they provide a systematic treatment of Web personalization from an inter-disciplinary viewpoint.

2.3 User modelling and preference elicitation

In order to acquire accurate, relevant user information, personalization techniques take the construction of user model and preference elicitation as a prerequisite for information retrieval and filtering. User modelling and preference elicitation are a key factor in personalization systems (Kobsa 2000, Fink and Kobsa 2000) in that they act as main inputs to personalization components. User profiles are approximated for the evaluation of system reliability and retrieval effectiveness (Yao 1995), optimisation of search engines (Joachims 2002), refinement of search results (Sugiyama *et al.* 2004) and smart back navigation (Milic-Frayling *et al.* 2004). Smart back navigation takes into account user profiles to provide the user intelligent "back" navigation to retrospect to her/his historical Web pages.

Current Web search engines provide searching results without taking into account user's individual information needs, interests and preferences. Recently t (Sugiyama *et al.* 2004) derived user profiles to adapt retrieval results. However, this approach still relies on a keyword-based search engine with a weak assumption on the performance of retrieval algorithms. Moreover, user's interests and preferences should be an additional and essential criterion to refine Web search engines, but still this implies to understand and model the contexts of search requests (Lawrence 2000).

Eliciting user's preferences is a non-deterministic process that involves many intuitive and non well-defined criteria that are difficult to model. Identifying user's preferences over a given domain knowledge requires either observing user's choice behaviours or directly interacting with the user with pre-defined questions. A key issue in eliciting user's preferences is the problem of creating a valid approximation of the user's intentions with a few information inputs. The measurement process for modelling user's preferences consists in the transformation of user's intentions towards a classifier or regression model that rank different alternatives. Several knowledge-based algorithms have been developed for eliciting user's preferences from pairwise-based comparisons to value functions. An early example is the pairwise algorithm comparison applied on the basis of ratio-scale measurements that evaluate alternative performances (Saaty 1980). Artificial neural networks approximate people preferences under certainty or uncertainty conditions using several attributes as an input and a mapping towards an evaluation function (Shavlik and Towell 1989, Haddawy *et al.* 2003). Fuzzy majority is a soft computing concept that provides an ordered weighted aggregation where a consensus is obtained by identifying the majority of user's preferences (Kacprzyk 1986, Chiclana *et al.* 1998). Preference elicitation is already used in E-Commerce evaluation of client profiles and habits (Riecken 2000, Schafer *et al.* 1999), flight selection using value functions (Linden *et al.* 1997), apartment finding using a combination of value function and user's feedbacks (Shearin and Lieberman 2001).

Elicitation of user preferences on the Web is still a non-straightforward task as the amount of information available on the user profile, and the extent to which the system might interact with a given user is limited. In spite of the heterogeneity of the user communities that interact with Web applications, there is a lack of a-priori knowledge on the user profiles, cultural and knowledge backgrounds. Meanwhile it's a non-straightforward task to derive user's information needs during Web interactions. User preference elicitation techniques should consider minimal user inputs at the beginning stage, and collect as much user information as possible through observing user's behaviours. This implies to explore and develop unsupervised mechanisms that facilitate manipulation and analysis of information on the Web, that is, to approximate user preferences and intentions in order to guide and constrain information retrieval processes. Last but not the least, a preference elicitation process should also be flexible in order to favour interactions and refinement between the user initial intentions, and the preference elicited by the system.

2.4 Web personalization techniques and algorithms

2.4.1 Algorithm classification

A lot of searching engines, tour guide or recommendation agents and intelligent user interfaces are designed to provide Web pages and information content to the user according to its intentions and preferences. Common personalization operations on the Web include annotating Web links, creating adaptive Web site by learning from user access patterns, and synthesize index pages (Perkowitz and Etzioni 1997, 1998). Intelligent agents are also a flexible way to help the visitors to explore the Web (Pazzani and Billsus 2002).

There are a number of possible classification schemas for Web personalization algorithms (Resnick and Varian 1997, Schafer *et al.* 1999, Terveen and Hill 2001, Mobasher 2004). Several kinds of Web personalization algorithms can be distinguished: collaborative filtering, content-based filtering, demographic-based personalization, and knowledge-based personalization. Content-based filtering (e.g. Lietiza, Webwatcher) and collaborative filtering (e.g. Grouplens, Amazon) are two orthogonal paradigms widely applied to E-Commerce and Web-based interactive systems.

The term collaborative filtering has been first introduced by the Tapestry system of electronic mailing lists (Goldberg *et al.* 1992). Collaborative filtering takes explicit user’s interests and preferencess from user’s ratings on entities such as documents and products in E-Commerce, and implicit information derived from user’s behaviours. It generates recommendations according to user profile similarities using proximity measure methods or correlation engines. These approaches are based on the assumption that users might share interests and preferences as far as they belong to a same category. User categories are implicitly derived and organized according to user’s ratings and behaviours. With the example of an Amazon session, when the user buys a specific book, then the system present books also bought by the readers who bought this book (Figure 2.3). Collaborative filtering suffers from a lack of scalability, data scarcity, and the “cold start” problems (Breese *et al.* 1998, Sarwar *et al.* 2000a). Lack of scalability and data scarcity result from the fact that getting sufficient user’s information such as ratings on personalization services in large applications is a non straightforward task. The “cold start” problem arises from a lack of information about a user who logs on for a first time, and about a new item. For personalization systems, there is usually no direct reward for providing examples since they only help other users. This leads to many difficulties in obtaining a sufficient number of ratings.

Figure 2.3 A collaborative filtering application example from Amazon

Content-based filtering personalizes Web services and information retrievals on the basis of content similarity of Web documents and user personal profiles. This approach is based on the assumption that a given user might like entities similar to the one she/he is interested in. Content-based filtering has several identified drawbacks such as biases of input, and static profiles (Mobasher 2004). User inputs are often subjective to the user and thus prone to biases. Static profiles obtained from user registration may also degrade over time.

Demographic-based personalization refers to personalization rules specified by Web designers, based on static user profiles that contain basic information such as gender and age (Mobasher 2004). Demographic-based approaches heavily rely on user's inputs, which are difficult to gather since users usually lose patience or simply give incorrect information when facing with Web registration forms.

Knowledge-based personalization algorithms (Burke 1999, 2000) infer user's information needs, interests and preferences through an original example the user is familiar with and a series of tweaks. It employs techniques derived from case-based reasoning (Hammond 1989, Riesbeck and Schank 1989, Kolodner 1993) whose objective is to solve a new problem by retrieving the olds that are likely to have similar solutions. Personalization executes a series of tweaking actions alter characteristics of the original example solution to make it more closely match the problem situation according to either explicit or implicit user relevance feedbacks. The process is repeated until the system reaches a satisfiable solution. User profiles are derived either from a Web query, an original example entity of interest, or from an explicit user model (Towle and Quinn 2000). Knowledge-based personalization requires the user to input sufficient information, e.g., an original example and refining options at each refinement steps. Therefore, there is still a need to explore flexible inference rules and personalization engines with minimum user inputs.

2.4.2 Optimisation strategies for personalization

Various optimization strategies for Web personalization processes have been applied to address some of the problems unresolved by collaborative and content-based approaches. In particular, several hybrid approaches unify content-based filtering and collaborative filtering in order to form an integrated Web personalization approach (Balabanovic and Shoham 1997, Baudisch 1999, Claypool *et al.* 1999, Melville *et al.* 2002).

Many research proposals enhance Web personalization using similarity indexing, dimension reduction, and clustering techniques. (Aggarwal *et al.* 1999) proposed a similarity indexing method to index market data. This method mainly amounts to an index, so-called signature table, based on the single-linkage clustering algorithm to flexibly support a wide range of similarity functions. A signature table contains a set of items (e.g. products in E-Commerce) that are partitioned from those available in the original data. A transaction activates a signature if and only if the disjunction of the transaction and the signature exceeds an activation threshold at a specific level. Partitioning an information space in collaborative filtering systems can reduce the amount of computation time, increase scalability, and improve prediction quality (O'Conner and Herlocker 1999). Dimensionality of application spaces for generating recommendations can be reduced through applying Latent Semantic Indexing (LSI) techniques (Sarwar *et al.* 2000b). The Latent Semantic Indexing is a dimensionality reduction technique based on the underlying matrix factorization algorithm. It's complementary to the collaborative filtering algorithm challenges such as sparsity and scalablity. In (Ungar and Foster 1998) a statistical model of collaborative filtering addressed the sparsity problem and multi-dimensional characteristics of user preferences. The basic rationale of their statistical model is to assign items of interest (e.g. movies) and to categorize users into classes using clustering mechanisms.

2.4.3 Data mining oriented approaches

Web usage mining applies data mining techniques to analyze and extract information and knowledge from user's historical trails recorded by Web logs (Pierrakos *et al.* 2003, Eirinaki and Vazirgianis 2003). A Web usage mining process is a three-phase process (Figure 2.4) that consists of data preparation, knowledge discovery and pattern analysis (Cooley *et al.* 1997).

- Data preparation process includes several domain dependent tasks such as data cleaning, user identification, session identification and pattern completion.

- Knowledge discovery applies statistical methods and data mining techniques to generate knowledge on navigational rules and patterns.
- Pattern analysis identifies interesting rules and patterns by comparing the discovered knowledge with the Website designer's views.

(Mobasher *et al.* 2000a) proposed a Web personalization architecture based Web usage mining techniques and relevant solutions for usage data preprocessing, usage knowledge extraction and recommendations. Web personalization approaches based usage mining still suffer from a lack of usage data, and from the fact that the contents of Web sites evolve over time (Mobasher *et al.* 2000b). Therefore, these approaches should consider the semantics of Web contents to improve Web personalization results (Eirinaki *et al.* 2004).

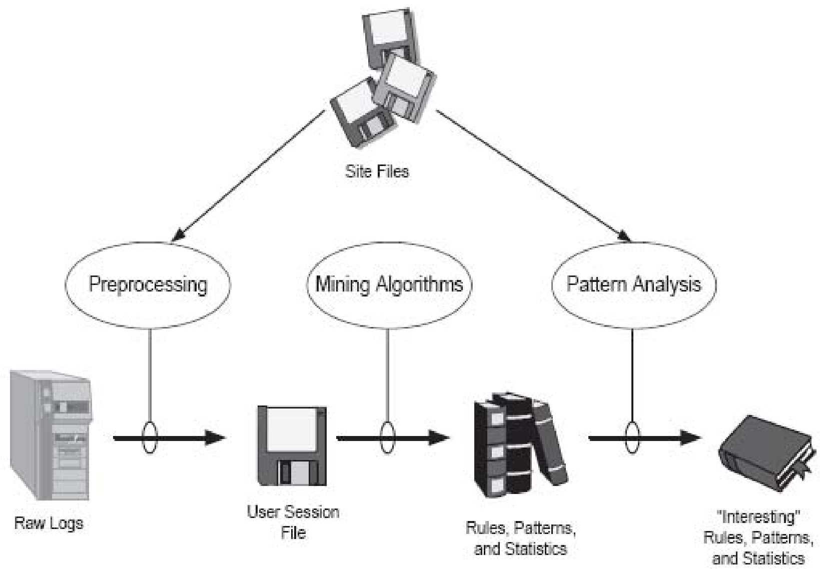


Figure 2.4 Web usage mining process (from Cooley *et al.* 1999)

Web data mining methods are also employed to discover relationships between pages (or items) based on navigational patterns such as association rule mining (Srikant and Agrawal 1997), sequential pattern discovery (Mobasher *et al.* 2002), clustering, and markov process model (Deshpande and Karypis 2004). Inspired by All K-th order Markov model, (Kim *et al.* 2004) proposed a hybrid prediction model that recursively applies four models: the Markov model, sequential association rule, association rule and a default model. These models perform predictions in tandem in their precision order from the highest to the lowest. As their experimental results showed, the hybrid model performs better than any individual model. They also used a learning approach to measure the performance of each individual model for a given context in order to set the sequence of them in the precision order. However, this hybrid personalization model is non trivial to be applied due to its computational complexity. (Albanese *et al.* 2004) presented a Web personalization strategy based on pattern recognition techniques, taking into account static registration information and dynamic user behaviors. The classification over the user domain is based on a clustered algorithm, and an iteratively repeated re-classification phase that ensures a suitable convergence of user clustering.

2.4.4 Machine learning for Web personalization

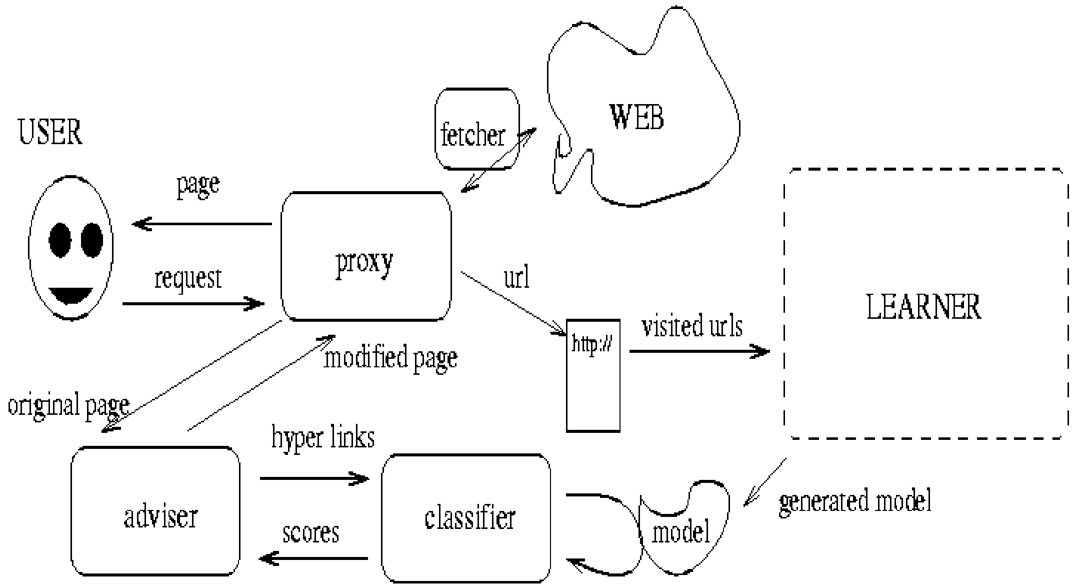


Figure 2.5 Machine learning in Personal WebWatcher (from Mladenic 1999)

Machine learning, in a broad sense, refers to the changes in systems or computer agents that empoy *artificial intelligence (AI)* to perform tasks such recognition, diagnosis, planning, robot control, prediction (Mitchell 1997). Machine learning for user model has awoke from the winter of last decade and resurged recent years. There are four critical issues that lie on the way to its widespread application: the need for large data sets, the need for labeled data, concept drift, and computational complexity (Webb *et al.* 2001).

- Large data sets: One important limitation of machining learning approaches to user modeling tasks is that they do not build a model with acceptable accuracy until they get a relatively large number of examples.
- Labeled data: Supervised machine learning approaches require explicitly labeled data (e.g. degree of interest), but correct labels are difficult to infer from the simple observation of user's behaviours.
- Concept drift: User modeling is a very dynamic task due to the fact that user's characteristics such as interests and preferences are likely to change over time. Therefore, learning algorithms should be capable of adjusting to these changes, this being a challenging problem known as concept drift.
- Computational complexity: Machine learning approaches applied on the Web have been fairly limited, this being caused by their computational complexity.

In Personal Webwatcher, machine learning is used to generate and model user's interests (Figure 2.5). Various machine learning approaches have been introduced for incrementally learning and revising user profiles from user feedbacks on interesting degree of Websites, e.g., naïve Bayesian classifier (Pazzani and Billsus 1997) and k-Nearest Neighbor and reinforcement learning (Mladenic 1999). The Webwatcher allows agents to learn user's interests and personalization strategies using a Q-learning function (Joachims *et al.* 1997). Q-learning (Watkins 1989) is a form of Reinforcement Learning algorithm that does not need a model of its environment and is very suited for repeated games against an unknown opponent. Q-learning algorithms estimates the values of state-action pairs $Q(s,a)$. The value $Q(s,a)$ is defined to be the expected discounted sum of future payoffs obtained by taking action a from state s and following an optimal policy thereafter. (Zhang and Seo 2001) proposed a method to learn user's interests in a personalized Web-document filtering system. In contrast to conventional approaches based on relevance feedback (Rocchio 1971), information filtering process was formulated as a *TD(0)* reinforcement learning problem, to learn user profiles and adapt the term weights in order to represent user's information needs and interests, and maximize the expected value of user relevance feedbacks. *TD(0)* is known as the simplest *temporal-difference* (TD) learning method. TD methods can directly learn from raw experience without a model of the environment's dynamics, and update estimates based in part on other learned estimations, without waiting for a final outcome.

2.5 Example systems

This section introduces and discusses some of the most popular Web personalization systems. Many of these systems apply a combination of several personalization approaches. There are few of system that solely developed with demographics-based approach. We thus discuss three groups of personalization systems: content-based, collaborative filtering-based, and hybrid. The hybrid personalization systems are based on personalization techniques optimised with some strategies, such as Web usage mining and machine learning. WebSifter is a personalized search system using ontology and knowledge representation. For complete issues, it is also discussed in this section. Major tools/systems for Web personalization include the followings (however by no means exhaustive):

2.5.1 Content-based filtering systems

Letizia (Lieberman 1995): Letizia is a "zero-input" software agent that assists Web browsing (Figure 2.6). It learns user profiles by recording and analyzing the user's browsing activity in real time, and provides a continuous stream of recommendation of Web documents. It works like a "friend" watching the user browsing on the Web, who soon learns what the user is interested in, then makes predictions and gives the user better recommendations. Browsing with Letizia is a cooperative activity between the user and an agent. It uses Netscape as its interface, and one or more additional windows to show recommendations continuously. While the user is looking at a page, Letizia conducts a search in the local Web neighborhood surrounding that page, and recommends further actions on the user's part. Letizia doesn't require the user to state her/his goals. Instead it infers user's interests or goals implicitly from user's browsing behavior. It is located on the user's machine and thus serves one particular user's current interests. **Let's Browse** (Lieberman *et al.* 1999) extends Letizia to support customized services for multiple users (Figure 2.7).

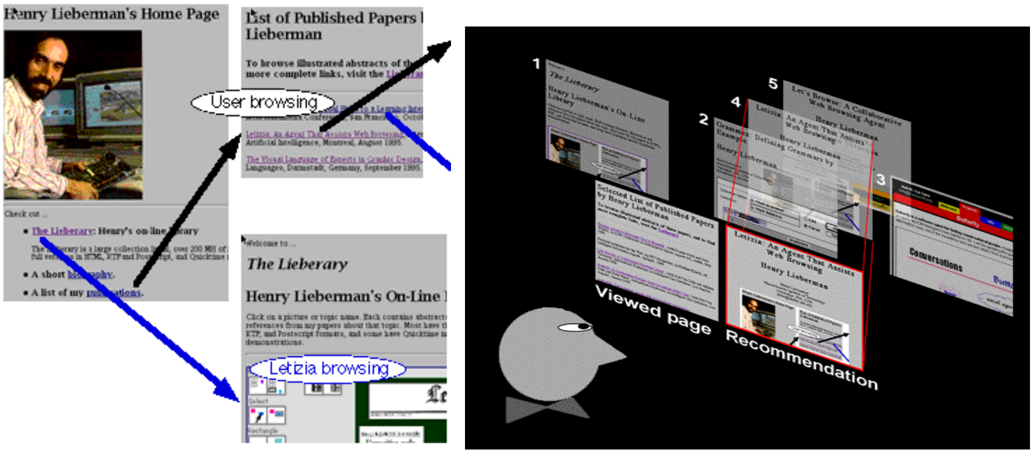


Figure 2.6 Letizia snapshots

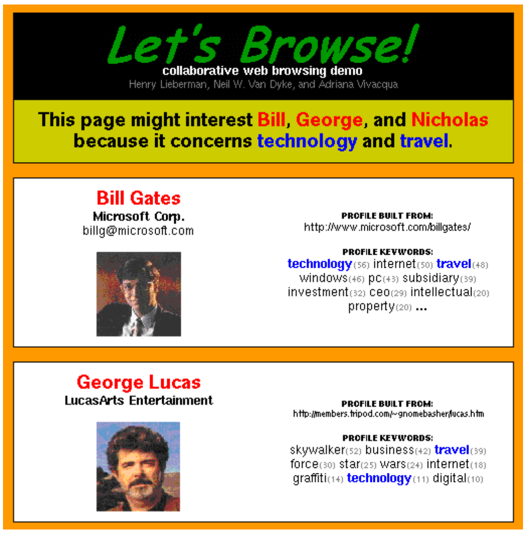


Figure 2.7 Let's Browse snapshot

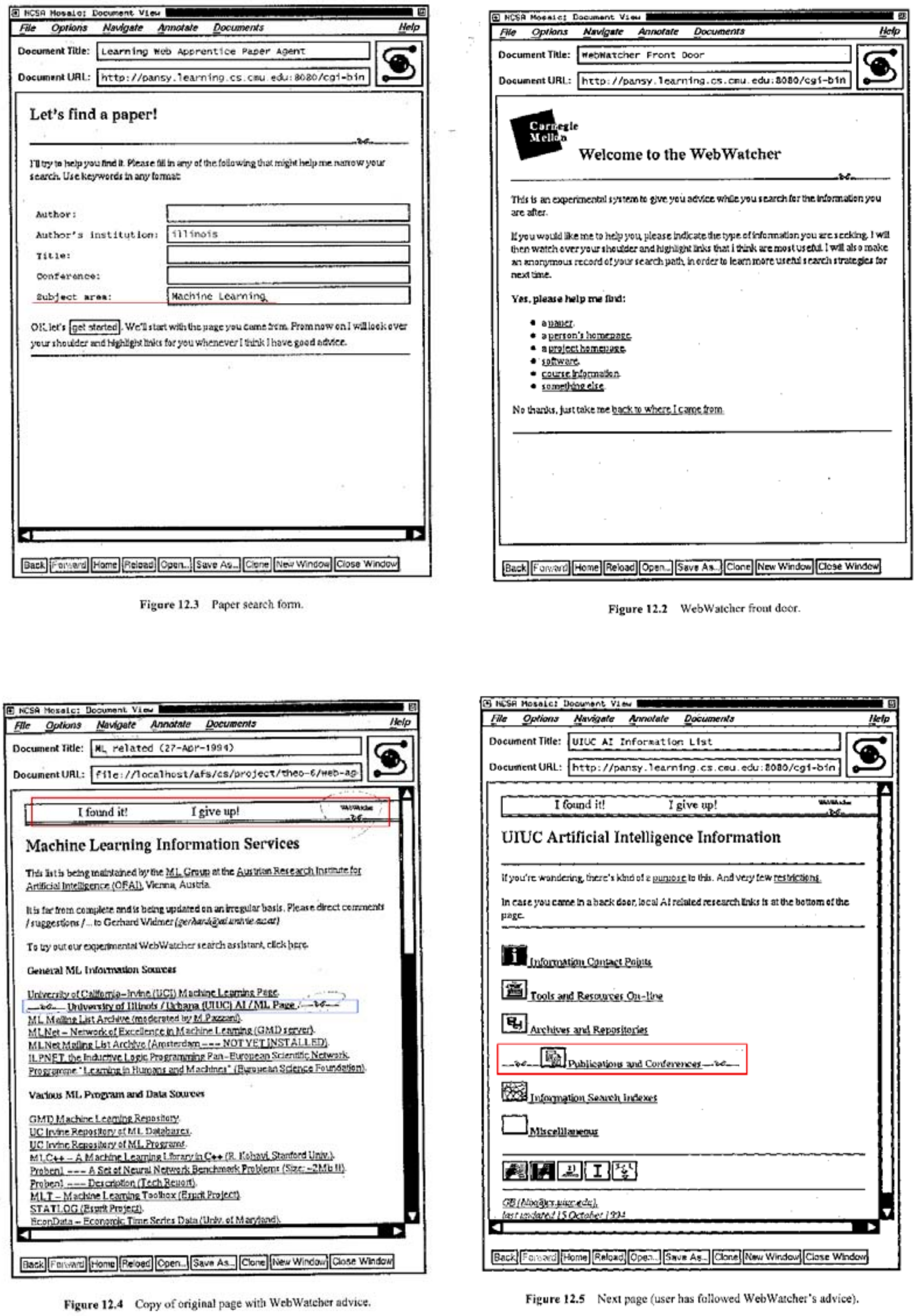


Figure 2.8 WebWatcher snapshots

WebWatcher (Armstrong *et al.* 1995, Joachims *et al.* 1997): WebWatcher is a tour guide software agent for assisting users to browse the Web (Figure 2.8). It learns user's interests while browsing and recommends Web documents. First, when the user invokes an instance of WebWatcher, where the user is asked for a short description of her/his current interests. Secondly, WebWatcher returns to the initial page, prepares to guide the user to navigation on the Web. Thirdly, the user can find some changes that

WebWatcher makes to the initial page. That is, this Web page is augmented with WebWatcher's additions, including a menu bar, a list of new hyperlinks, selected hyperlinks from the original page have now been highlighted with eyeball icons. Fourthly, at the end of the tour, the user shall give explicit feedbacks to show her/his satisfactions. WebWatcher is located on a separated server. It modifies every page the user asked, e.g., adding recommendation indexes.

Personal WebWatcher (Mladenic 1996): Personal WebWatcher is mainly inspired by WebWatcher: "a learning apprentice for the Web". It is an agent that assists users in finding information on the Web. Personal WebWatcher is oriented to a particular user, modeling its interests, without involving the user in the learning process, that is, it doesn't ask the user for any keywords or opinions about Web pages. It records the addresses of Web pages requested by the user and highlights corresponding hyperlinks. Personal WebWatcher is a personalized form of WebWatcher, since every user has its own copy of the system.

Lira (Balabanovic *et al.* 1995): Lira is an off-line Web-based search system based on some heuristics. First, it presents the best pages to the user using some selection heuristics. Secondly, it receives an evaluation from the user for each page presented, and updates the search and selection heuristics according to these feedbacks. It learned by asking the user to rate pages explicitly as the figure shows (Figure 2.9).

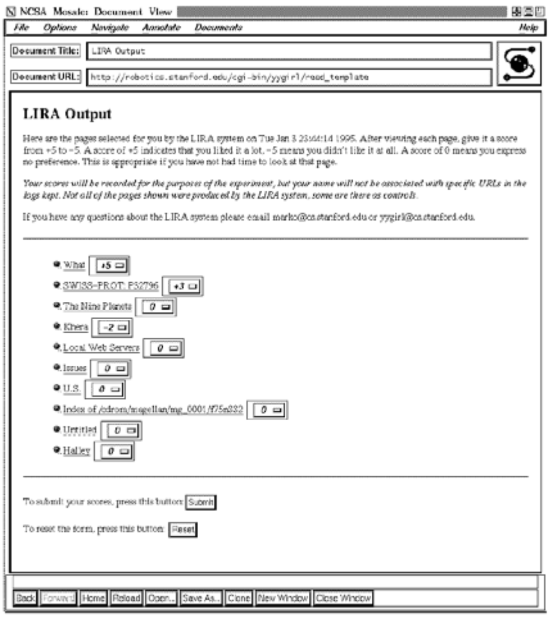


Figure 2.9 Lira snapshot

Syskill & Webert (Pazzani *et al.* 1996): Syskill & Webert is a software agent that learns to rate Web pages, deciding what page might be of interest to a given user. It organizes a separate profile for each topic to a given user (Figure 2.10). This allows the user to rate a page as hot (two thumbs up), lukewarm (one thumb up and one thumb down), cold (two thumbs down). Then, it learns the user's page ratings and uses this profile to suggest other pages accessible from the index page. It can also retrieve Web pages by turning the topic profile into a query, using search engines like LYCOS. Starting from a manually constructed index page for a particular topic, the user can rate hyperlinks of this page.

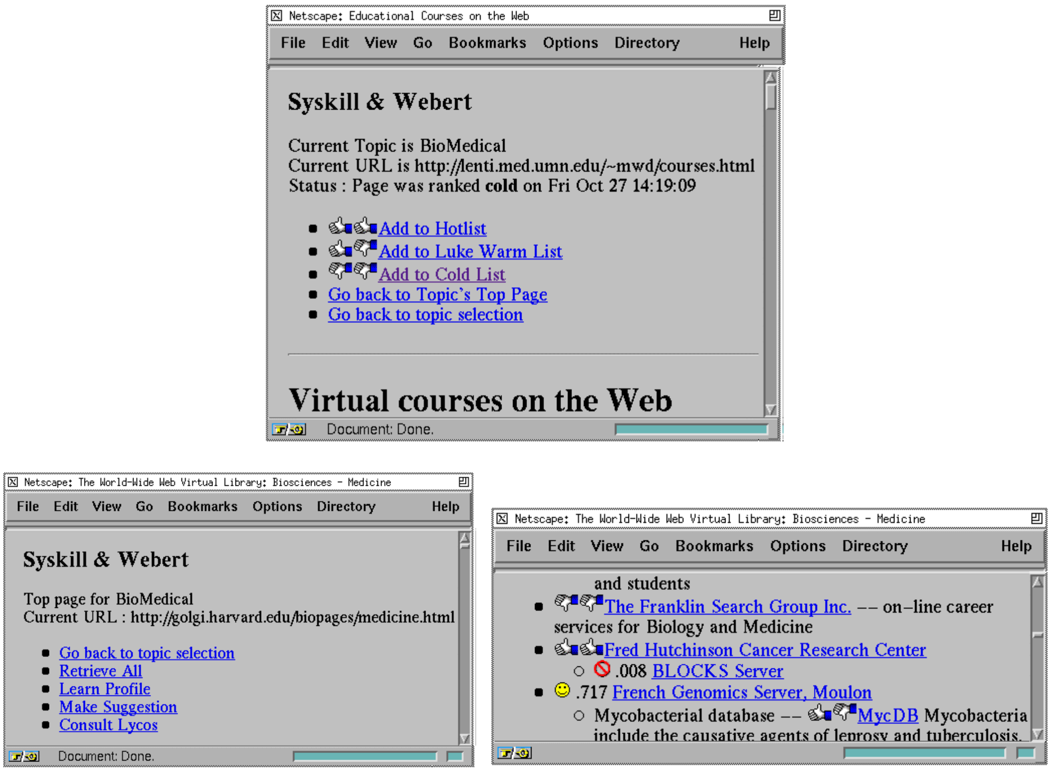


Figure 2.10 Syskill & Webert snapshots

NewsWeeder (Lang 1995): NewsWeeder is a netnews filtering system. It asks the users rate Web news articles on a five-point scale and learns user profiles based on these ratings. The system employs a learning algorithm based on Minimum Description Length principles to raise the article recommendations to the user.

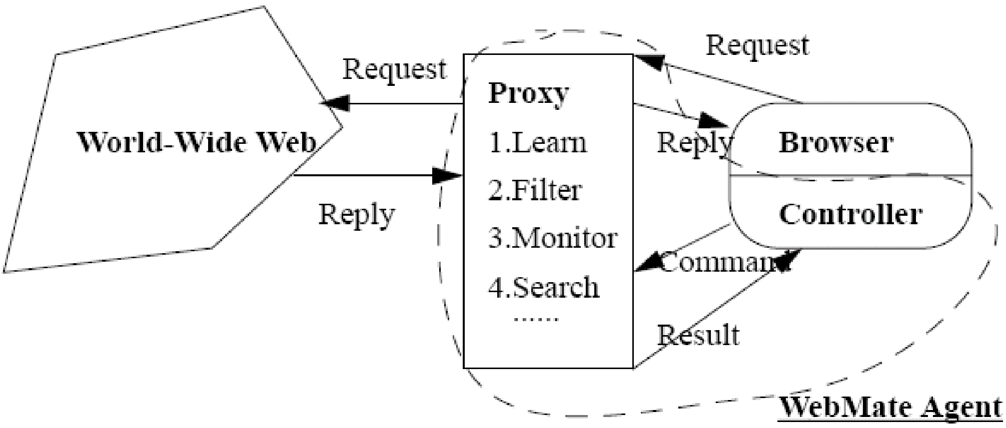


Figure 2.11 WebMate system architecture

WebMate (Chen and Sycara 1997): WebMate provides navigation assistance through observing user's behaviors. It works in a similar way (Figure 2.11) to WebWatcher but runs on the end user side. It uses multiple document vectors to track user's interests, learns user profiles, and helps the user to improve Web searches using keyword expansions and relevance feedbacks. The system extracts and combines relevant keywords from Web pages and uses them for keyword refinement.

2.5.2 Collaborative filtering systems

GroupLens (Konstan *et al.* 1997): GroupLens employs collaborative filtering algorithms to provide online recommendation services. The user's profiles are extracted from explicit and implicit ratings, which refers to navigational data such as the time that a user spent on a page.

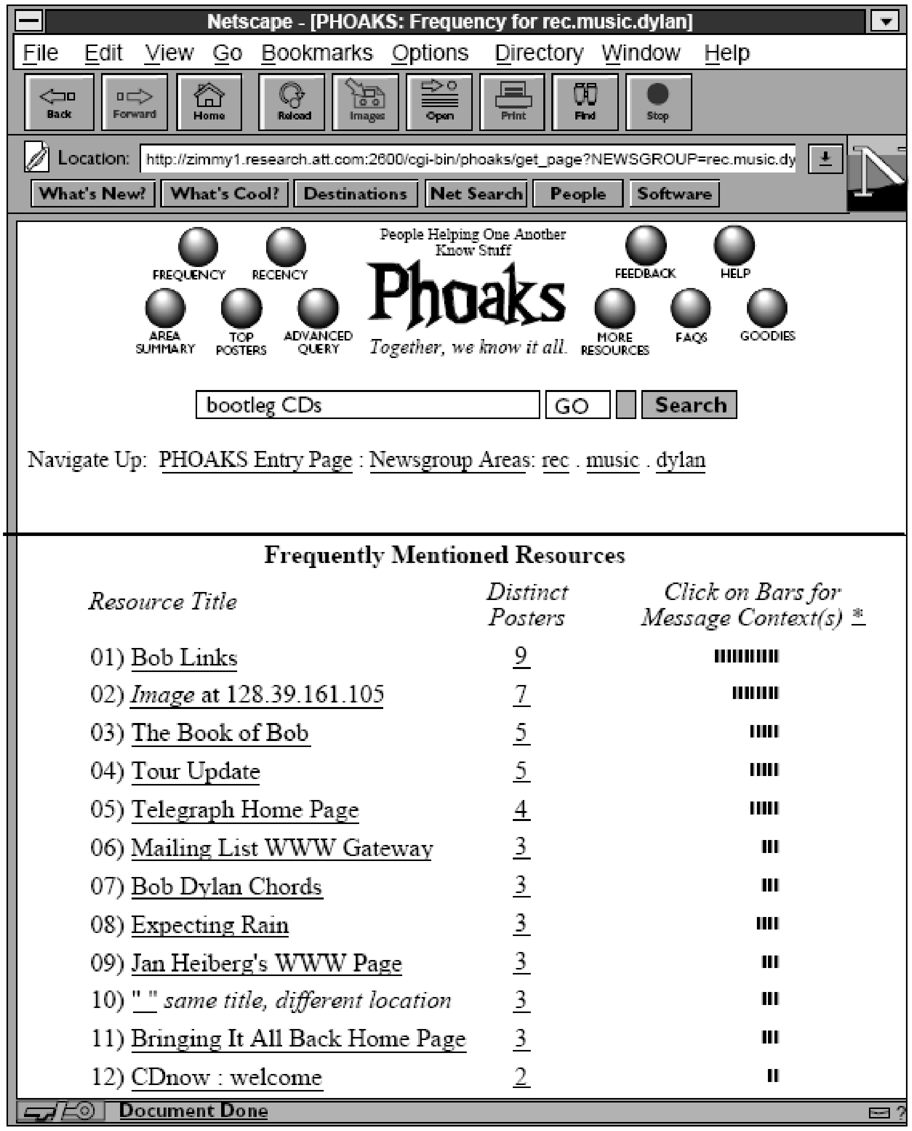


Figure 2.12 PHOAKS snapshot

PHOAKS (Terveen *et al.* 1997): PHOAKS (People Helping One Another Know Stuff) is another experimental application employing a collaborative filtering approach

(Figure 2.12). It is based on two major design principles: role specialization and reusability, compared to ratings-based systems. Role specialization refers to taking a different view to roles of recommendation provider and recommendation recipient, instead of role uniformity. Reusability denotes recognizing, tallying recommendation resources from online conversations, that is, collective assessment of newsgroups.

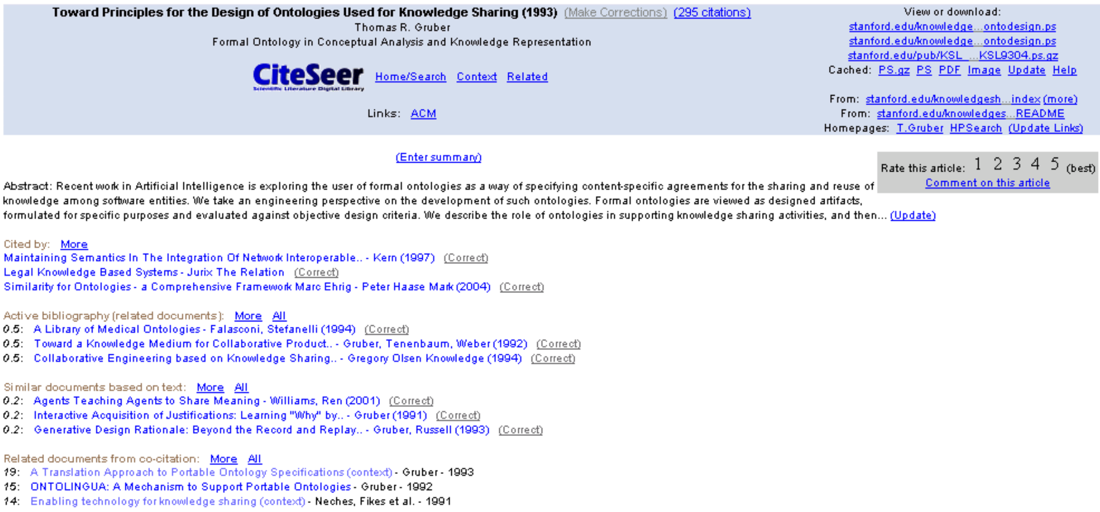


Figure 2.13 Citeseer snapshot

Siteseer (Rucker and Marcos 1997): Siteseer is a collaborative Web page recommendation system, based on a bookmark structure that describes the content of the selected documents (Figure 2.13). It uses individual's bookmark files and structure to predict and recommend relevant Web pages. User profiles are derived from bookmarks in a Web browser. These bookmark files are used to identify user groups with similar interests.

2.5.3 Hybrid Web personalization systems

Entrée (Burke et al. 1996, 1997): The restaurant recommender system Entrée (Figure 2.14) uses case-based reasoning methods to make recommendations in a city beginning with an original restaurant the user knows and likes. The refinement process allows the user to navigate or search by stating her/his preferences with respect to some features of a given restaurant.

Fab (Balabanovíc and Shoham 1997): Fab implements a hybrid content-based and collaborative filtering to recommend Web pages to the user. The user is recommended an item that is scored relatively high against her/his profiles, this recommendation being reinforced when it is also rated highly by other users with similar profiles. User profiles are represented as weighted keyword vectors. This requires the users to explicitly rate the recommendations given by the system.

Figure 2.14 Snapshot of Entrée

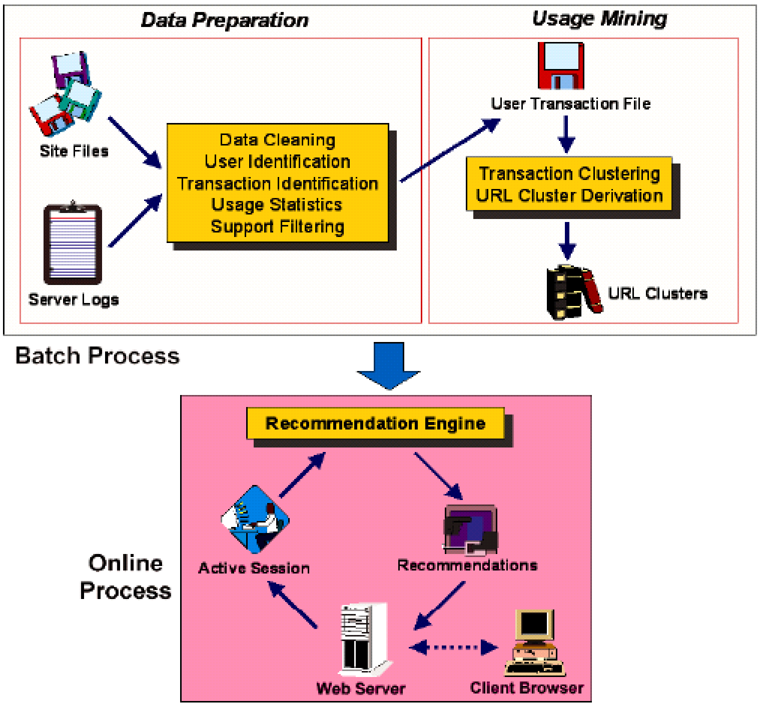


Figure 2.15 Architecture of usage-based Web personalization systems (from Mobasher 1999)

WebPersonalizer (Mobasher *et al.* 2000a, Mobasher 1999): WebPersonalizer is a usage-based Web personalization system developed by Mobasher and his colleagues.

Compared to previous personalization approaches based on explicit inputs, Web usage based personalization avoids static user profiles and biases, and makes the personalization process automatic and dynamic. WebPersonalizer can automatically offer the user effective navigational pointers based on her/his active session, and the aggregate usage rules and patterns of other similar users. It provides a general framework for Web usage based personalization that consists of an offline and an online process (Figure 2.15). The offline process prepares Web log files and uses data mining techniques such as association rules and sequential patterns to extract information and knowledge. The online process relies on a personalization engine to match user's active session and the discovered aggregate profiles in order to effectively recommend Web documents. The framework has been extended through integration of Web content and usage profiles to enhance Web personalization results (Mobasher *et al.* 2000b).

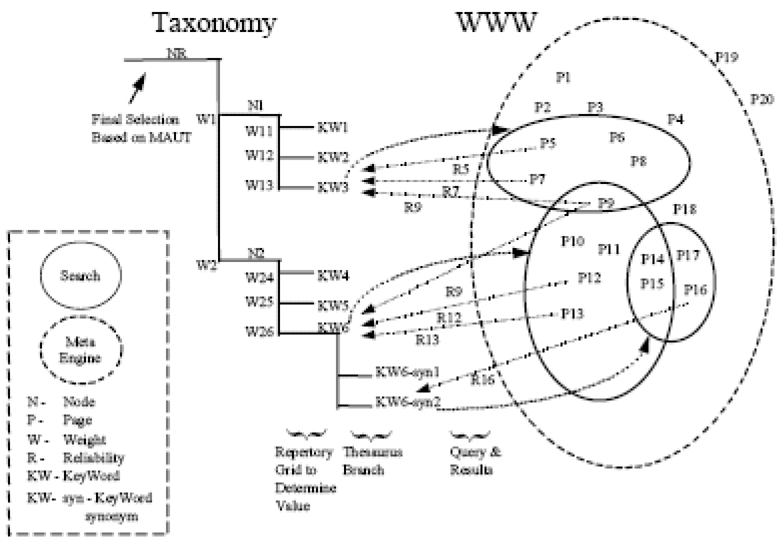


Figure 2.16 WebSifter retrieval and analysis architecture

WebSifter (Scime and Kerschberg 2001): WebSifer is a personalizable meta-search system based on a weighted semantic taxonomy (Figure 2.16). It typically employs a methodology and architecture for query construction and results analysis that provide the user a ranking of choices based on the user's determination of importance of query terms and Web pages. The user can define the information intent as an ontology of search keywords, which is complemented with a standard thesaurus to accommodate possible differences between the user's terminology and the search engine keywords. The design of search ontology can also refer to the taxonomy store for suggestions based on the work of previous Web searchers and the thesaurus to find semantically similar categories. The resulting search ontology is populated with appropriate Web pages selected by multiple search engines. A reliability store keeps track of Web pages visited and rated by previous Web searchers. WebSifter proposes a unique method for the syntactic and semantic rating of returned pages. The ranking of Web pages is a function of both availability of these pages and user's decision criteria, based on a combination of decision analysis and database management approaches. The ranking process gives more weight to Web pages matching multiple categories within the search ontology. It also takes into account of both the organization of the ontology and relative weights of categories in the ranking process.

WebSift II (Kerschberg *et al.* 2001) extends WebSifter principles with elaboration of user preference representation scheme based on various components, each of which represents a specific decision-criterion. The relevance of a page hit is rated by a decision-analytic rating mechanism combing a weighted semantic taxonomy tree model and the component-based preference representation. Websifter II is implemented through cooperating with Wordnet for concept retrieval and most well-known search engines.

MEMOIR (Roure *et al.* 2001): MEMOIR is a framework extensible and adaptable to a given application, with a standard Web browser as the user interface. It analyses user's trails, open hypermedia link services and a set of software agents to assist users in accessing and navigate vast amounts of information in the distributed or Internet environment. These software agents treat and mine user's trails as "first class" objects. Similarities between user's trails are based on the trailmarks contained in the trails, the keywords contained in these trailmarks and user's profiles are attached with weights.

This section illustrates and analyses major Web personalization systems and agents developed so far in the academic and industrial domains. Amongst them the ones described in sections 2.5.1 and 2.5.2 mainly use either collaborative filtering or content-based filtering approaches associated to appropriate similarity measures to discover and recommend Web pages and products in E-Commerce that might be of interest to the user. Research proposals illustrated in section 2.5.3 introduce various optimisation strategies to address the demerits of conventional approaches (e.g. collaborative filtering and content-based filtering). With the development and widespreading of Web services and increasing heterogeneity of user communities, this calls for personalization systems and tools that should be more intelligent, easy to operate, and capable of high-quality adaptive information services.

2.6 Research trends

Web personalization is widely acknowledged as an effective solution to improve quality of information services and Website design. In order to provide more intelligent, rich and effective personalization results, recently various research trends emerge in the Web personalization research community.

2.6.1 Semantic Web personalization

Conventional Website design, Web-based information retrieval and data mining view Web pages as the basic elements to present, organize and analyze information and items. With initiatives to deal with more complex relationships between entities embedded in Web documents instead of plain Web pages, there is an explosive growth of research on integration deeper semantic knowledge from domain ontology (Dai and Mobasher 2002), which is in envision of the semantic Web. The general solution is to automatically or semi-automatically extract objects and ontology from Web documents through some appropriate data mining approaches such as text classification algorithms.

The demand for capturing more complex relationships and patterns at a deeper semantic level is acknowledged (Berendt *et al.* 2002). With the emergence and proliferation of semantic Web (http://www.w3.org/2001/sw/, Berners-Lee *et al.* 2001, Fensel and Musen 2001), it facilitates the incorporation of semantic knowledge and domain ontology into personalization processes. Semantically enriched Web personalization systems manage and reason about user model, navigational pattern, and Web content at different stages to optimise the personalization processes and results. Semantic Web personalization is considered as a novel issue in the field of Web personalization, which leads to various research perspectives. (Dolog and Sintek 2004) described a service-based architecture for personalizing e-learning in distributed environments based on semantic Web technologies such as metadata and ontologies. The main components include personal learning assistant, user interaction, resources provision network and a series of personalization services.

In order to improve Web site design and personalization services based on Web usage mining, some efforts are oriented to the semantic enrichment of Web log files. (Dai and Mobasher 2002) proposed an ontology-based approach to aggregate more general concepts from Web usage profiles, thus to identify common user profiles/preferences at more generalized level of abstraction. Then user's navigation activities can be described and predicted with more rich semantics at different levels of abstraction. A step further, (Dai and Mobasher 2005) explored personalization strategies based on semantic knowledge about the underlying domain. They apply ontology-based methods to extract semantic features from the textual Web content and integration with Web usage mining, to provide insightful and smart personalization services. (Oberle *et al.* 2003) presented a framework to enhance Web usage profiles with formal semantics based on an ontology underlying the site relevant domain. The semantic Web usage mining consists of three steps: description of raw data of transactions, mapping the recorded URLs to Web entities based on a domain ontology, and mining and identification of user navigation patterns over the set of meaningful Web entities.

2.6.2 Cross-system personalization

One of the main hindrances to personalization is the limited representation of user's personal information and preferences (Callan and Smeaton 2003). This is aggravated when a given user's interactions and operations with a Website is viewed as a part of a larger task that covers several such kinds of interactions with different systems. Therefore there is still a need for more standardized, generic user models (Kobsa 2001) for cross-system personalization (Niederée *et al.* 2004, Mehta *et al.* 2005).

Another underlying challenge in cross system personalization is how to deliver personalized representation to users across the boundaries between the user and the services (Stewart *et al.* 2004). (Thomson 2005) proposed a standardized cross-system personalization framework, addressing some issues such as design goals, user identifications and privacy policies. (Cingil *et al.* 1999) presented a broader approach for Web personalization based on W3C standards. In this approach, Web resources, user log files, user profiles are described with XML (extensional Markup Language)/RDF (Resource Description Framework), are both human understandable and machine processable, then could better support efficient and automatic agents to work in close cooperation with the Web. The user privacy is managed with Privacy Preference (P3P).

2.6.3 Logic-based personalization

Description logics (DL) based approaches attempt to support the reasoning tasks over quantitative information, conflicting and incomplete information. (Berghell *et al.* 2005) presented a system, so called SmartDate, which can automatically arrange dates for candidates and seekers, through matching user's profiles for dating and reasoning about demand and supply profiles. Recommendation and matchmaking techniques are based on DL (Calì *et al.* 2004), and take into account the user's location and the time at which she/he issue a query. The service is fully accessible via mobile services through a mobile browser or a Java Midlet.

2.6.4 User markup languages

Ubiquitous computing (Weiser 1991) brings some challenging but interesting issues to the semantic Web personalization and user modeling. It leads to virtually invisible technology embedded in everytime, everywhere, everything in our daily life. (Heckmann and Krueger 2003) introduced a user modeling markup language (UserML) for ubiquitous computing to enable communication about partial user models via the Web. The UserML language is based on XML/RDF since the latter has the advantage to be used directly in the Web environment. The content of a UserML document consists of metadata, user Model, inference explanations, context model and environment model. In (Heckmann *et al.* 2005a) a general user model ontology

GUMO was developed for the universal interpretation of distributed user models in intelligent semantic Web enriched environment (Heckmann *et al.* 2005b). It uses the Web ontology language OWL as an ontology language for the representation of user model terms and their relationships. The user model dimensions are divided into three parts: auxiliary, predicate and range (*subject {auxiliary, predicate, range} object*). The user's interests and preferences information for example "interests in footable" can be expressed as: *auxiliary = hasInterest, predicate = football, range = low-medium-high*.

2.7 Conclusion

This chapter gives a survey of the state of the art on Web personalization, from personalization components, algorithms, system examples and recent research trends. Web personalization is widely recognized to address several problems in the Web research and engineering community: information overload, irrelevant information supply and mismatch between Web authors' and Web consumers' understanding of Web resources. A variety of Web applications employ personalization techniques to improve user's satisfaction and utility of online information services.

Web personalization techniques started with conventional approaches such as collaborative filtering and content-based filtering, and evolve to provide more intelligent and semantic-rich information services. Collaborative filtering and content-based filtering can be viewed as two basic orthogonal dimensions of Web personalization techniques, focusing on knowledge extraction from the user domain and the application domain, respectively. Going a step further, several hybrid Web personalization approaches are introduced to improve personalization performance. The first is the combination of the two basic approaches, that is, collaborative filtering and conctent-based filtering. Secondly, it integrates dimensionality reduction techniques such as similarity indexing, Latent Semantic Indexing, to address sparcity and scalability in personalization systems. Thirdly, it introduces data mining to extract information and knowledge from user's historical trails on the Web, and machine learning approaches to infer and model user's interests and preferences to provide more intelligent personalization services.

The advent of the semantic Web supplies various opportunities and challenges in the Web personalization community. Recently, many research proposals address various aspects of semantic-enriched personalization. Semantic Web related techniques such as ontology engineering are used to explore information and knowledge among entities embedded in Web resources. A reliable solution for cross-system personalization applies ontology and metadata description languages such as XML/RDF(S) to build generic user models and preference elicitation rules, and to markup user's characteristics such as interests and preferences.

Chapter 3 Spatial Web Personalization

Despite the novel research achievements described in section 2, personalisation of spatial information on the Web is still an issue that has not been fully considered. Semantic Web related techniques can be used to markup spatial information resources on the Web, to design a generic user model and inference rules for the approximation of user preferences applied to spatial entities embedded in Web documents. This chapter introduces current researches on the spatial Web, particularly spatial Web personalization, and a preliminary outline of our research objectives.

The remainder of this chapter is organized as follows. Section 3.1 discusses spatial information retrieval on the Web. Section 3.2 introduces spatial entities on the Web. Section 3.3 gives a brief survey on spatial information personalization and related applications. Section 3.4 discusses research issues on spatial Web personalization. Finally, section 3.5 concludes this chapter.

3.1 Spatial information retrieval on the Web

A reasonable proportion of Web resources can be mapped to some degree to geo-referenced entities associated to a location in the geographical space. (Winter and Tomko 2005) argued that the World Wide Web is closely coupled with geographical structures and spatial entities embedded in Web documents, and that this continues to deepen with the emergence of the ubiquitous computing age. The fact that 90% of business data is geographically related (Moloney *et al.* 1993), emphasizes the potential role of geo-referenced entities on the Web. Statistics collected by search engines and systems on the Web found that spatial information is pervasive on the Web, and that many queries contain spatial information (Silva *et al.* 2004). Many objects in digital libraries (in local network or on the Web) are related to places in the real world. This leads to many digital libraries to consider geographic querying techniques to facilitate interactions with information resources that contain geographic characteristics (Larson and Frontiera 2004b). Moreover, electronic information on the Web such as IP address, and personal information can be considered as “spatially relatedness”. For example, an IP address is directly or indirectly associated with telephone area codes, place names, and spatial coordinates (Buyukokkten *et al.* 1999). According to estimations from the Kelsey Group[2], about 40 percent of search engine queries fall into a sort of local search (Bishop 2005). This includes for example the search for a specific business or service in a local area. Therefore, a local search engine should be ideally initialised with two kinds of inputs:

[2] http://www.kelseygroup.com/

a search term representing information of a product, service or business the user is looking for, and geographical criteria.

There are several weaknesses associated to current Web search engines and systems when the objective is to manage and search for spatial information. First, most search engines are still in keyword-based style. Web users have to input terms as exactly as possible to explicitly describe their information needs in order to get the relevant information. This leads to many approximations, or even inappropriate results due to the multiple semantic significance of a given search term or expression. Yet no sufficient efforts are put to identify and infer user's interests and preferences. In particular, it is non trivial to formulate a spatial query because of the inherent semantic complexity of spatial information. The second problem relies in the fact that search results are identical, and independent of the user and the contexts in which the user makes the request. In spatial information retrieval, users at different locations may expect different retrieval results, even from a same query. Finally, spatial queries on the Web should take into account a specific property of space that result from the First Law of Geography (Tobler 1970). This law stresses the role of proximity in a spatial environment in that it influences the distribution of things, interrelationships and consequently spatial searches. For instance, things in the neighborhood of a spatial entity returned by a query are likely to have a higher interest than the ones which are more distant. Thus, spatial proximity plays a central role in spatial information retrieval and personalization. However, there is a lack of appropriate mechanisms for the evaluation and integration of spatial proximity and semantic similarity in spatial information searches on the Web.

Spatial documents describe vast and rich range of information, which can be extracted into multi-dimensional forms, with each dimension corresponding to a semantic class or to a point of view. Recently, spatially enhanced information retrievals and search engines on the Web attracted wide interests and research efforts (e.g. Jones *et al.* 2002, Chrisment *et al.* 2004). (Larson and Frontiera 2004a) developed a ranking algorithm based on a logistic regression model to measure the spatial proximity between two regions in spatial information retrieval. The logistic regression model of information retrieval introduced by (Cooper *et al.* 1992) estimates the relevance probability between a particular query and a candidate record in the database as the log odds relevance between the query and a candidate record.

There emerge many research proposals on inference and analysis of spatial information and criteria extracted manually or through data mining techniques such as text learning. Retrieval and extraction of spatial and temporal attributes of spatial entities from the Web were applied with matching mechanisms on HTML tag structures that index spatial and temporal information (Tezuka and Tanaka 2004). (Tezuka *et al.* 2001) argued that geographical boundaries and relationships are distorted in conceptual maps by several factor such as the popularity of geographical objects *per se*, paths between them, and borderlines. A Web crawling algorithm is applied to search through the Web to obtain popularity and co-occurrence rates to regulate query resolver (Tezuka *et al.* 2003). However, the quality of query results depends on the performance of the Web crawling algorithm and Web author's style. (Bera and Claramunt 2005) presented an approach for inference and analysis over a semantic network with spatial and semantic properties extracted from the Web. The approach was applied to two conference Websites of the spatial reasoning scientific community, with an objective to infer and analyze the structural relationships among different research groups.

In order to deal with more advanced forms of spatial information retrieval, semantic information and relationships among entities embedded in Web documents shall be discovered. Spatial ontologies describe different spatial concepts, relationships and terminologies, which is expected to support query interpretation, disambiguation and expansion to improve spatial information retrieval on the Web (Fu *et al.* 2005a, 2005b), relevance ranking and Web resource annotation, that is, geo-markup of geo-referenced information (Hiramatsu and Reitsma 2004). (Jones *et al.* 2001) made an attempt to build an ontology of places, which describes some basic information of place and qualitative spatial relationships between places to support information retrieval tasks. A hierarchical distance measure, combined with a semantic distance measure based on classification semantics, was introduced to rank retrieval results. The intuition behind the hierarchical distance measure is that two spatial entities with common ancestors in the spatial hierarchy have closer spatial relationships (e.g. proximity).

The user community in spatial Web domain is inherently heterogeneous. This implies to manipulate spatial information with consideration user's information needs and background knowledge. Yet, few research works are devoted to identify user's interests and preferences, and to provide personalized spatial information services for the user on the Web or in mobile environments.

3.2 Spatial Web entities

The spatial Web stores and manipulates information with direct or indirect spatial semantics with some appropriate mechanisms and techniques, and provides services taking into account of features, attributes and relationships in the spatial, temporal and semantic domains. What is hereafter referenced as the "spatial Web" in a broader sense contains not only Web applications that diffuse electronic maps and manipulation languages on the Web, but also systems where spatial information is embedded on the Web using either textual symbolic or interactive map components.

The information, data and knowledge on the spatial Web are geo-referenced, visual, and explicitly or implicitly mapped to real objects in the physical environment, either urban or natural. For instance, in a Web urban environment, the landscape perceived by humans is composed of landmarks, edges, districts, paths and nodes (Lynch, 1960).

We call this kind of entities “spatial entities”. A spatial entity has physical or virtual locations and boundaries, and semantics assigned with respect to its properties and functions. It can be represented by geometrical primitives and semantic information, and connected by spatial relationships amongst them some are fundamental for navigational knowledge such as distance and cardinal relationships (Thorndyke and Hayes-Roth 1982, Benelli *et al.* 2002).

In the context of this thesis Web resources are made distinction between **spatial Web resources** and **a-spatial Web resources**. The former refers to any form of information, data, and knowledge related to the spatial dimension on the Web, while the latter is not related to the spatial dimension. Information entities are embedded in Web resources, so-called **Web entities**. Similarly Web entities are divided into **spatial Web entities** and **a-spatial Web entities**, which correspond to spatial Web resources and a-spatial Web resources, respectively. A spatial Web entity can be considered as the mirror of a spatial entity in the underlying physical environment, described and embedded in Web documents. In Figure 4.4, for example, sightseeing places of the Shaanxi Province (China) are described as a set of spatial Web entities embedded in a Web user interface.

Different kinds of spatial Web entities of interest (e.g. sightseeing places, hotels, universities) are embedded in multi-media Web documents either in textual or map forms. An urban ontology describes a set of objects, relations, events and processes related to a given application domain at conceptual level (Fonseca *et al.* 2000). It can be described as a container of a set of heterogeneous categories at different levels of abstraction, e.g. Sightseeing places, Hotels, Residences. It allows for interoperations between different urban models and databases, and communications among between various actors in urban management and planning (Keita *et al.* 2004). From a representation perspective, maps represent and display spatially referenced data to the users, whereas other media forms such as text and graph prevalent in Web documents serve as supplementary means to describe semantic and spatial contents. Spatial Web entities represented on maps denote explicitly their spatial locations and their overall distribution, potentially linked to some semantic documents that can describe additional spatial and semantic properties. Maps are the most intuitive way to represent spatial referenced information, although these are not always provided in an interactive way. Maps on the Web are so far provided either by graphic files or by interactive maps software. Interactive maps provide effective framework to present spatial information on the Web using human-Web interaction modes. Interactive maps (Putz 1993, 1994) allow the user to have an access to various kinds of interactions with the Web, and provide customized user interfaces for browsing spatial entities on the Web (OGC 2004, 2005). It is for instance possible under these principles to link the image of a spatial entity back to an interactive map viewer interface, allowing thus the user to perform some map-oriented operations and hyperlink interactive modes (e.g. clicking on the map to view detailed information of a spatial entity). In contrast to map-oriented spatial representations, other media forms represent spatial entities implicitly with un-structured descriptions of geo-references or embed relevant information within semantic descriptions such as postal codes or textual addresses.

Spatial Web entities identified and extracted from Web documents have additional relationships, that is, hyperlinks that connect them, which describe the location of spatial entities in the Web space and relationships among them (Figure 3.1). Web personalization on spatial information should discover spatial proximity and semantic similarities among spatial entities, match semantic and spatial properties with user preferences, and personalize Web services and experiences to users. User's interests and preferences can be deduced implicitly or presented explicitly through unobtrusively observing user's behaviours such as visiting spatial Web entities.

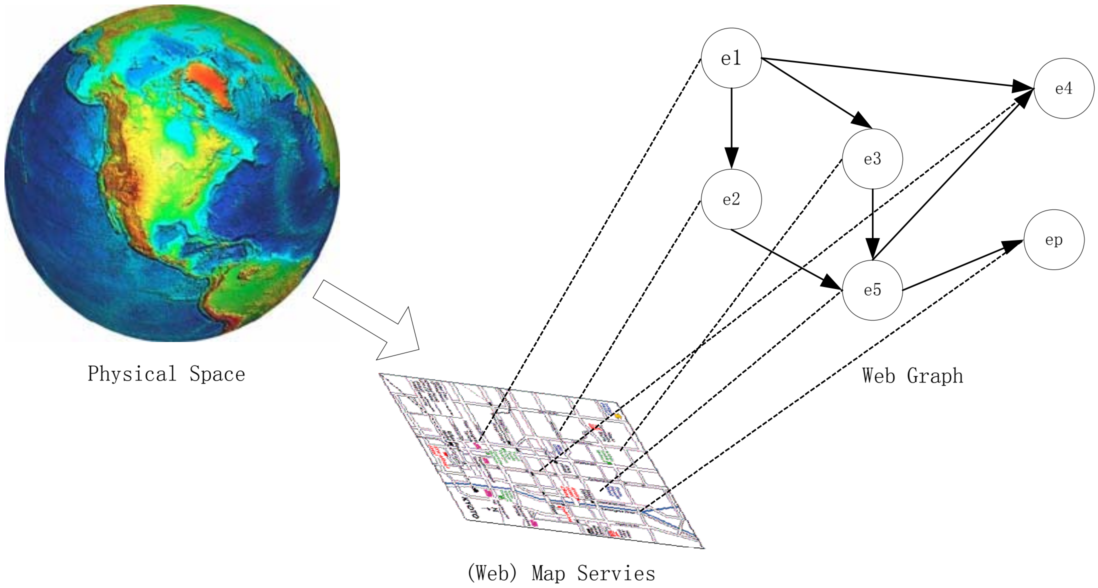


Figure 3.1 Physical space, Web map and Web space

3.3 Spatial personalization services

In (Spaccapietra *et al.* 2005) two kinds of information services were discussed within the ubiquitous computing community: broadcast-based and location-based services. They also analyzed two essential components: user profiles and contexts with the goal of providing personalized and context-aware information services to the mobile user. Going one step further, (Gu *et al.* 2004) argued to separate low-level context obtained directly by physical sensors from high-level context derived by a context interpreter. Specifically, low-level context contains and provide information on the user's locations and the spatial entities nearby. High-level context so-called situation, deduces information about user's behaviours. Information about the user, context and situation at different levels of abstraction serves as primary inputs for personalization components to provide customized services to the user. This section gives a brief survey on spatial personalization services, with special attentions to personalization and adaptive maps in the tourism domain.

3.3.1 Tourism personalization

In (Fink and Kobsa 2002), a generic user modeling system was proposed to personalize tourism and travel services, to watch and analyze user's behaviors and make generalizations and predictions about the user's actions in the future. Anderson and her colleagues (Anderson *et al.* 2001) built a Web site personalizer *PROTEUS* to retarget existing Web contents for mobile clients in order to reduce time and efforts to attain information-seeking goals. They emphasize on the limitations of mobile devices compared to desktop clients such as low bandwidth networks, small interface and slower processors, while less attentions to spatial semantics. In addition, user-centric factors (e.g. demographics, interests and preferences) and spatially related aspects (e.g. location, context) should be considered when designing guides for spatial Web users and mobile users.

Spatial information personalization should be capable of integrating spatial and semantic criteria within user's queries and about application resources. Accordingly, it goes beyond simple database querying in that it can help the user to evaluate and rank decision alternatives based on multiple criteria. The implication of personalized multi-criteria decision strategy for spatial information personalization is to employ user preference to adjust parameters of a multi-criteria evaluation method. (Rinner and Raubal 2004) implemented a case about hotel booking, which supports personal decision-making based on the Ordered Weighted Averaging (OWA) operator. The OWA model contains a set of order weights and corresponding importance weights, that are used to emphasize the better or the poorer properties of each decision alternative (Yager 1988). The OWA-based approach allows the user to standardize selected criteria using qualitative utility values, and to weight the relative importance of properties of each decision.

Location-based mobile services take user's location and environment to deliver relevant information services (Schiller and Voisard 2004). (Yu *et al.* 2003) used user profiles to refine queries in location-based services, but avoid the issues such as acquisition and construct user profiles, and learning and update with the evolution of user's interests and preference. (Weibenberg *et al.* 2004) discussed some issues on ontology design, user model, contexts and situations of situation-aware services to provide the user with information tailored to her/his interests and preferences. Their works are still to transfer from a demonstrator to a prototype system. The situations can be viewed as the high-level contexts, the inference engine abstracts from the low-level context dimensions by translating specific contexts into logic situations.

The project Deep Map (Malaka and Zipf 2000, Malaka *et al.* 2000) aims at the design of smart user interfaces and software agents for access to personal information and handling the user's context and the availability of resources. It develops an intelligent spatial information framework and related techniques for tourism services. Preliminary efforts generate tour proposals based on personal interests and preference of a tourist, e.g., a route with high user dislike of smoke and noise. In this research framework, they recognize the requirement of exploiting user models and context-aware knowledge, while still inferring user's interests and preference and integration of user model components pose problems to be solved. Another project, CRUMPET (Poslad *et al.* 2001, Schmidt-Belz *et al.* 2003), is intended to create "user-friendly mobile services personalized for tourism". User's interests and preferences are modeled based on a domain dependent taxonomy, e.g., a taxonomy of tourism-related services (Schmidt-Belz *et al.* 2002). The domain taxonomy is a structured set of concepts and attributes describing a tourism service. The user's interests are given as a set of probabilities corresponding to concepts and attributes in the domain taxonomy, with a mapping function. An adaptive map prototype implements a step-by-step solution to dynamically generate tourist maps according to a range of variables from user's interests and preference, the given task, cultural aspects, to context and location (Zipf 2002).

3.3.2 Adaptive map

Web-based interactive adaptive map generation and visualization are one of the major applications for spatial Web personalization techniques, which could provide adapting Web maps customized to the user's needs and contexts (Cecconi *et al.* 1999, Cecconi and Galanda 2002, MacEachren and Kraak 2001). Task that tourists want to accomplish is one of major situational factors, which represents an important basis for the design of interactive maps for mobile guides (Hunolstein and Zipf 2003). This is motivated by the observation that the user's expectation from the map is not exactly the same as the represented map (Figure 3.2) Digital maps are derived from conventional maps, which are organized with several layers of interest (e.g. roads, builds) based on layer oriented concepts. A layer of interest contains a set of spatial entities whose presentation depends the user's needs, interests and preference and the context of map usability (Nafaa 2005). In this thesis research, he proposes a multi-agent approach based on the use of multiple representation and cartographic generation for Web map generation, among them one so-called Coordinator agent performs some personalization actions before transferring a given layer of interest. The system developed offers the user several simple options for map browsing such as the modification of maps styles, and a registration form to specify the type of entities of interest.

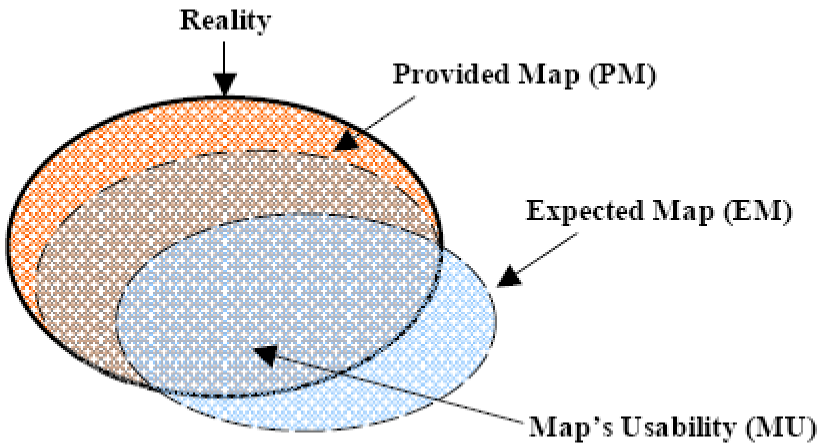


Figure 3.2 Reality, provided map, expected map (From Nafaa 2005)

In a project on map personalization, (Doyle *et al.* 2004) applied personalization techniques to map producing systems in order to provide the user personalized maps. They developed a system on desktop, and then migrated it to mobile environment called *CoMPASS* (Combining Mobile Personalized Application with Spatial Services) (Weakliam *et al.* 2005a). The main idea is to record user's interactive actions on map frameworks and features from which user's interests and preference in terms of spatial content are identified to facilitate the next visit for the individual user (Weakliam *et al.* 2005b). The system requires only each unique user's name/ID to discern her/his user profiles. This engenders a problem when two different users who use the same name/ID to retrieve maps, however no efforts towards user identification. System log files recording the user's actions aren't analysed to update user profiles until the end of a user session. Then it lacks of on-the-fly processes to personalize map services according to user's current behaviours, e.g. to predict what the user want to do next. It focused on user's map operating actions but without any attempts to record and analyse user's trails on the map content, which is believed to be quite important in location-based mobile services. In addition, the authors seem to under-estimate actions on map frame/interface, which may be beyond their scope but believed to be valuable to personalize user interface.

3.4 Spatial Web personalization

3.4.1 Motivation

On the one hand, and in spite of successful personalization techniques and applications, to the best of our knowledge, personalisation of spatial information on the Web is still an issue that has not been fully considered. Current algorithms applied in user preference elicitation and personalization do not consider the spatial dimension although this might be of interest for many application areas such as Web-based travel planning. User preference elicitation and Web personalization almost neglect to consider space as an essential component.

On the other hand, a few spatial Web applications supply personalized services in terms of spatial information. This is probably due to less attention to user's information needs and preferences to the end user especially novice users. Different users with various tasks and preferences might get the exact same experience and results from spatial Web applications. Therefore, customized spatial information services become necessary to improve end-user's satisfaction, taking into account user's tasks, knowledge, interests and preferences.

Spatial Web personalization can be viewed to fill the gap between conventional Web personalization and spatial information retrieval. This implies to personalize information services related to the spatial dimension, with the explosive growth of spatial information on the Web, communication services and portable services. This requires integration of spatial properties and relationships into personalization techniques, in order to provide the user relevant spatial information services according to her/his interests and preferences.

Spatial Web application research can be categorized into two groups with respect to an application environment, that is, Web-based and location-based spatial information service personalization. Web-based and location-based spatial information personalization services regard the Web as the main information and service source. They require the design of inference rules and personalization techniques related to the spatial dimension. Differences lie in the facts that 1) the Web-based spatial personalization emphasizes to organize effective and flexible Web structure, content and usage, while location-based approaches also take into account user's location and the context to a higher degree. 2) In the former case the user accesses spatial information online through mobile devices, while she/he is physically acting in a physical environment in the latter case. This thesis research focuses on spatial information personalization in the Web environment, in order to design and manipulate mechanisms for user preference elicitation and personalization services at the design level.

3.4.2 Research problem statement

A same spatial Web entity may hold different contents from different user's viewpoints and preferences. In an information retrieval case, that is, the retrieval results expected may depend on the user's personal information, e.g. knowledge, social ground, interests and the current context. Let's consider an historical museum:

- A tourist who has little knowledge about the city demands information about its location, open hours, traffic means to get there, and exhibitions of interest.
- A general citizen may pay attention to more specific details and events in it.
- A specialist in architecture may have interests in its structure and historical

heritage.

In spatially related environments, this becomes more complex, due to the essential characteristics inherent in space. Besides, another factor is the disparity between cognitive maps in human mind and the real world, which is non-straightforward to understand and evaluate. In addition, spatial information personalization on the Web should emphasize on the extra criteria necessary to consider spatial properties and relationships. Furthermore, spatial components bring several additional considerations to Web interfaces that enlarge the dimensions offered to users.

Our research domain is the one of the understanding and modelling of dynamic interactions between the user and the spatial Web that combines different sources of multi-dimensional and geographical-related information. Spatial Web personalization should fully consider the cognitive way of human beings activities while interacting with spatial entities in a given environment. Understanding the way human manipulate geographical data is a complex and non-deterministic task that implies to model how human perceive and interact with geo-referenced information space. The cognitive ways of human interaction with spatial entities and environment act as the fundamental basis for spatial Web personalization. This should provide some clues or at least an approximation of user's intentions and preferences during the procedure of manipulating geographical information on the Web. Representing geographical and semantic data of a given country, region or city on the Web is very expressive in terms of information content and interaction opportunities offered to users.

This implicates that spatial criteria and relationships are the predominant factors to identify and infer user's interests and preferences, to provide personalized services in spatially related applications in various environments. For example, planning actions in urban spaces involves a range of choices and user preferences that relate to spatially related domain knowledge particularly diverse and stochastic. In particular, this relates to the planning of tourism activities where one of the complex problems faced is the lack of understanding of the way tourists use to select and arrange their activities in the city (Brown and Chalmers 2003).

Furthermore, spatial Web personalization amounts to the design of inference rules for user's interests and preferences and personalization engines over spatial information entities on the Web to support personalized spatial information services in diverse Web applications. Possible information about the user include demographics, task, knowledge, interests and preferences, these information comprise a user model and profile. Besides, inference rules should offers two distinct ways for user's personal information and preference elicitation. The first is to infer implicit preferences from explicit user assumptions in the user model, another to identify user's information, e.g., preferences through unobtrusively observing human-computer interactions. In order to provide personalization services across different systems and Websites, a decent way to maintain user profiles with a general user model based on ontology, so-called user ontology. A user ontology should describe concepts and relationships concerning the user's demographics, knowledge, activities, interests and preferences relevant to a given application domain. Personalization engines over spatial entities could provide flexible means to personalize accesses to spatial information since relevant domain information and relationships are organized at entity level. Flexibility refers to the fact that personalization engines could provide various ways for service customization.

Presentation of personalized spatial information content still needs some appropriate relevance ranking functions to show the user the final results through appropriate way, for example, convenient and acceptable to the user. This implies to explore semantic similarity and spatial proximity measures, and relevance ranking functions on the behalf of the user. The semantic domain is essentially multi-dimensional, and thus combining of spatial and semantic criteria is one of the remaining challenging issues in the field of spatial Web personalization. This computational and reasoning issue is related to multi-criteria decision-making and analysis relevant spatial information (Malczewski 1999, Thill 1999). Contextual factors and the user's personal information can be used to adjust the weights of the relevant criteria in a specific application domain, which forms a context-aware and personalized multi-criteria decision-making methodology.

3.5 Conclusion

Spatial Web personalization is acknowledged as a promising research issue that deserves more efforts to integrate spatial semantics with personalization techniques. Consideration of the spatial dimension should lead to the combination of spatial and semantic criteria with respect to a series of relevant techniques, e.g., similarity measure, relevance ranking algorithms and personalization processes.

Spatial Web personalization research opens several challenging research perspectives that are investigated in this PhD thesis.

- First of all there is a necessity of an integrated research framework that should combine both the spatial and semantic dimensions. It should act as a basis for user preference elicitation and personalization services.
- Secondly, user modelling and preference elicitation mechanisms should be explored to identify user information, e.g., interests and preferences.
- Personalized search and recommendation strategies should be able to use relevant user information to search for and recommend spatial entities that might be of interest to the user.

Chapter 4 Towards an integrated framework for spatial Web personalization

4.1 Introduction

Personalization on the spatial Web should tailor spatial Web contents and presentation to user's tasks, interests and preferences. This implies to model and integrate information about the user (e.g. demographics, historical activities) and available spatial entities embedded in Web documents, to infer and prioritize personalized results. Interactions between the user and a spatial Web space are to some degree related to her/his perception of the underlying physical environment. Intuitively, the user constructs her/his own concept map of an underlying spatial environment when interacting with spatial Web services. Spatial Web personalization implies to design a Web-based interactive and inductive learning system that could reflect user's perception of the underlying physical and Web space, and approximate user's interests and preferences.

A Web urban space can be viewed as a mirror of a given urban space on the spatial Web. It can be informally defined as a set of image schemata, spatial and semantic information related to a given city, and presented to the user by means of a Web site. It consists of a set of spatial entities and a specific information environment materialized on the Web. Spatial entities are explicitly or implicitly embedded in a variety of Web documents that present relevant information to the user. These spatial entities and the associated environment are stored in centralized or distributed Web sites, and are associated with spatial and semantic properties and relationships.

There are still many open and challenging issues on spatial Web personalization. This is further reinforced by the increasing amount of online and distributed spatial information. The modeling and integration of spatial entities within a Website lead to several research issues that, if successfully addressed, could open several opportunities for Web designers whose objectives are to develop personalized services on the spatial Web. The most important amongst these issues are to our opinion as follows:

(1) Representation of spatial entities within a given spatial Web environment.

(2) Exploration and identification of semantic and spatial relationships that relate these spatial entities.

(3) Modelling the mappings between the semantic description of these spatial entities and information presentation within the spatial Web environment. This implies to relate Web content (e.g. semantics exhibited by spatial entities) to Web presentation (e.g., Web pages at the user interface level).

(4) Presentation of spatial information to the user, that is, at the user interface level.

(5) User preference elicitation, and search for and delivery of personalized spatial information services to the user.

This PhD research addresses the above issues as a whole at the Web design level. We believe that Web user preference elicitation and personalized search processes applied to spatial information should consider and explore the way spatial entities are semantically and spatially related. Personalized search should integrate these semantic and spatial properties within a process of user preference elicitation, these being deduced implicitly or explicitly, in order to provide customized results.

The objective of this research is to design appropriate inference rules to identify user's interests and preferences, and to personalize user's navigation and experience on the spatial Web. The way a specific user interacting with the Web is located and moves in a given physical spatial environment or not, leads to different human perception and cognitive processes that influence her/his expectations. The former refers to mobile environments where the user interacts with the environment through portable and embeddable devices. The latter denotes conventional interactions between the user and a Web information space using a "table" computer, but where the user plans to act in the city later. Our research work mainly focuses on the second application context in order to design a conceptual framework for user modelling and personalization in spatial Web applications. Our research framework and principles are applied as a demonstrative case study in the tourism domain, e.g., Web-based travel planning in a Web urban space.

We assume no prior – if any – little knowledge of the spatial Web environment presented by the Web interface, neither experiential nor survey knowledge[3]. Our framework is intended to manipulate and deliver customized information services during interactions between the user and a given Web urban space. Potential applications include any form of spatial Web interactions, e.g., search for information on a spatial entity, or travel in a Web urban space. Given a spatial Web environment of interest, such as in the tourism domain, the user is expected to plan a trip and where she/he would like to find out some spatial entities of interest, and a reference entity from which she/he will be able to act in the environment.

The remainder of this chapter is organized as follows. Section 4.2 introduces an

[3] Experiential knowledge is derived from direct navigation experience while survey knowledge reflects geographical properties of the environment (Thorndyke and Hayes-Roth 1982).

integrated research framework for spatial Web personalization, and focuses on one of its components: a conceptual framework for personalization services. Section 4.3 develops the modelling principles of our framework. Section 4.4 introduces an approach for the construction of a user-centric conceptual map, spatial proximity and similarity measures with consideration of the overall contextual knowledge. Finally, section 4.5 concludes the chapter.

4.2 Research framework

4.2.1 A conceptual framework for spatial Web personalization

Spatial Web personalization is intimately linked to spatial Web design. The design of a personalization framework for spatial Web applications requires a user model and associated flexible user preference elicitation mechanisms, a personalization engine that combines spatial and semantic criteria, and an intuitive user interface enriched with spatial components (Kuhn 1996). These three components should be used to personalize Web services and interactions between the user and Web-based spatial information systems. Spatial Web personalization implies the modelling and representation of user features, particularly the ones relevant to the spatial domain. Accordingly, and instead of the consideration of conventional user modelling and preference elicitation techniques, there is a need to explore user modelling and preference elicitation mechanisms appropriate to spatial Web applications.

This PhD research proposes an integrated conceptual framework for user modelling and preference elicitation, and personalization services on the spatial Web. The framework identifies spatial personalization services and a semantic user model. These two components communicate information and knowledge about the user through inter-process communications: "*tell*" and "*ask*" operations (Figure 4.1). This chapter motivates and introduces a conceptual framework for personalization services on the spatial Web, within which personalized search strategies and personalization engines are developed to customize spatial information services.

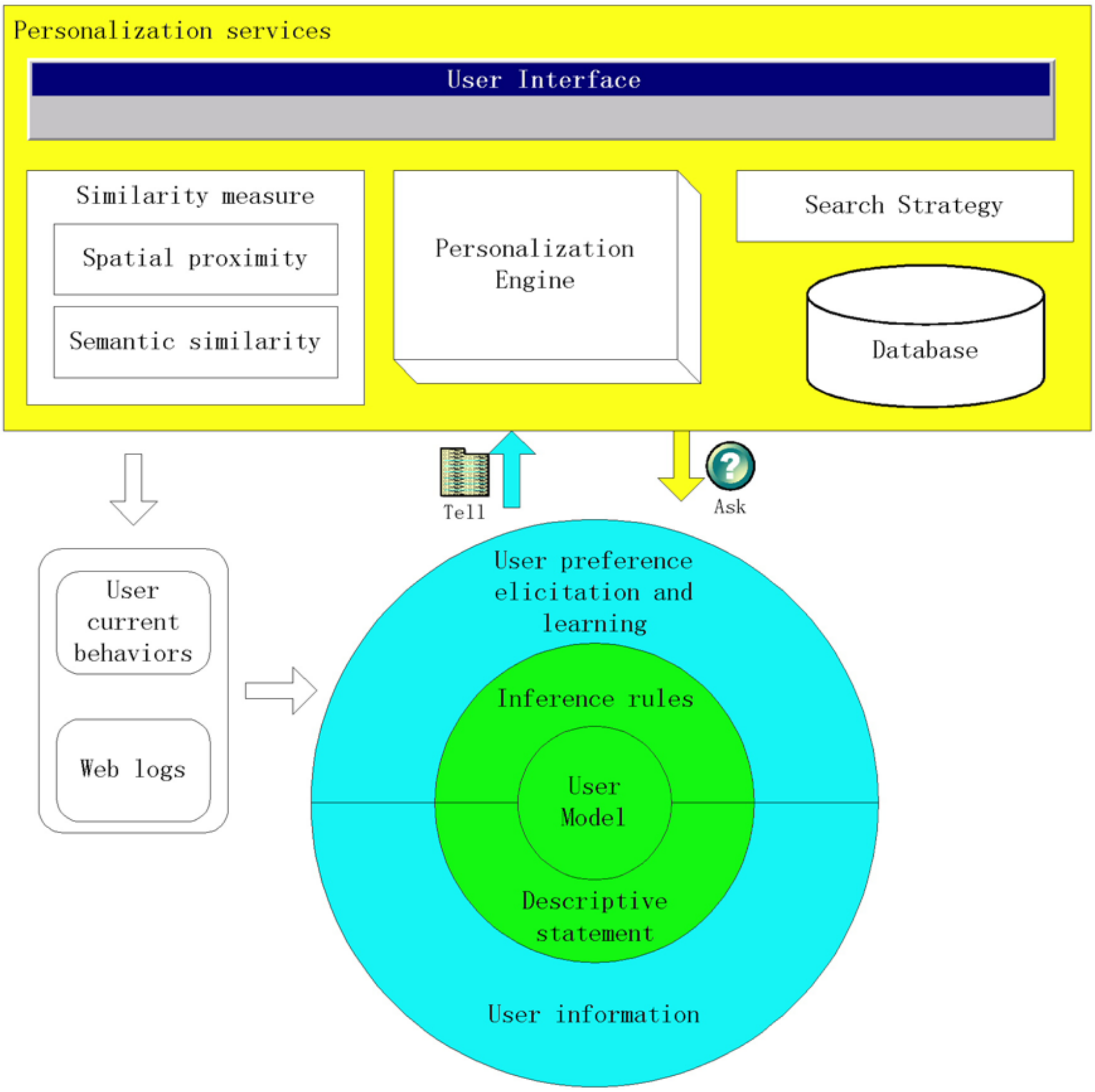


Figure 4.1 An integrated framework for spatial Web personalization

4.2.2 A framework for personalization services

Spatial Web design and information retrieval take spatial entities embedded in Web documents as the basic modelling concepts. Proximity between spatial entities is one of the most fundamental relationships when considering the case of a user physically acting and moving in a given physical environment. Proximity, and its inverse notion of distance, are also a primal component of an ontology of space as denoted by its role in the First Law of Geography (Tobler, 1970):

> *Everything is related to everything else, but near things are more related than distant things*

Proximity is to a high degree involved in the way human beings perceive and act in the real world, and thus influences queries and manipulation of spatial information

(Leuski and Allan 2004). Usually, human beings consider proximity with reference-oriented expressions, e.g., "*which is the nearest restaurant?*". According to this example, the user's current location may act as a reference for further interactions with the underlying physical space. We define a reference entity in this context as a salient location from where a given user can act and interact with the underlying physical environment.

Our objective is to design a spatial Web personalization system that manages and combines spatial and semantic criteria, infer user's interests and preferences, and retrieve a set of spatial entities that might be of interest to the user. Spatial entities in a given spatial environment are inter-related by spatial and semantic relationships, which are valuable criteria for the representation and manipulation of spatial information on the Web. We introduce a two-level framework for personalization services on the spatial Web based on a BNAM architecture that manipulates spatial entities (Figure 4.2). The Bi-directional Neural Associative Memory (BNAM) provides efficient means to store and recall pairs of associated information items. In a spatial Web context, the BNAM can be used to discover associated patterns exhibited either by the spatial distribution and semantic contents of spatial entities of interest, or by user's interests and preferences. Furthermore, spatial and semantic associations that present interests to a given user can be reinforced with explicit or implicit user's feedbacks. The framework consists of two levels that describe a set of spatial entities of interest and a set of reference entities, respectively.

In order to provide a semantic representation for a given spatial environment, fuzzy logics are used to relate the semantic features of a spatial entity to several classes of interest (Figure 4.5) identified according to a domain ontology (Figure 4.6). Fuzzy quantifiers standardize and formalize the semantic content of spatial entities with reference to an ontology, which organizes and describes the concepts and relationships relevant to the target domain. The choice is motivated by the fact that a given spatial entity can belong to different classes, with different orders of magnitude, while Crisp logics can not flexibly describe such properties. Last but not the least, the advantage of fuzzy logics is that it introduces variability in the associations that relate the spatial entities and reference locations in the BNAM framework. Spatial entities are classified using fuzzy quantifiers according to degrees of membership to several domain-dependent classes of interest. For example, in a given urban environment, a temple surrounded by a garden is likely to have high degrees of membership to predefined classes of interest: garden and temple, relatively high to a class museum and low to a class urban. The second important parameter considered in the spatial entities search and ranking process is given by an aggregate evaluation of the proximity of those spatial entities to some reference locations.

The work distinguishes Web information description and presentation. The descriptions about spatial entities and their properties are encapsulated within the considered Web site. The proposed framework employs image schemata and affordance concepts at the interface level. A Web-based interface provides an interacting level where user preferences are first encoded using an image schemata-based selection of the spatial entities that present an interest for that user. The concepts of image schemata and affordance enrich Web user interface to facilitate interactions between the user and a given Web urban space, and act as a bridge for spatial information presentation and content description in a Web database. The Web user interface provides dynamic interactions between the user and the Web environment. The principles of the user preference model are supported by a flexible interface that encodes user preferences in the selection of spatial entities of interest in a spatial Web environment, and ranks these entities that best fit user preferences. The main components of the user preference model are motivated and described in the following sections.

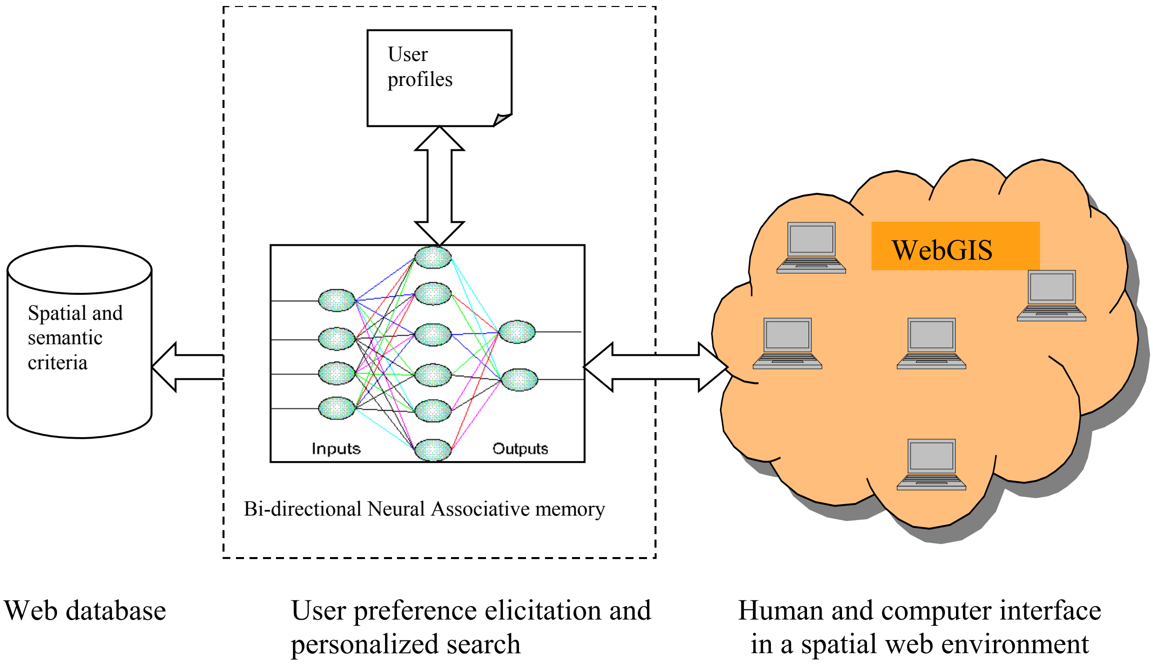


Figure 4.2 Framework for personalization services on the spatial Web

The research framework takes advantage of the high degree of semantics owned by the specific properties of spatial information, and develops a Web-based model and interface for the evaluation of user preferences in the manipulation of visually explicit spatial information. One of the constraints of the development is to be flexible on the one hand, and intuitive and limited in terms of explicit user inputs on the other hand. For this reason, we also explore to infer implicit user's interests and preferences through observing and analyzing user's behaviors during the interactions between the user and the spatial Web.

4.3 Framework principles

4.3.1 Bi-directional Neural Associative Memory (BNAM)

The personalization service framework supports flexible personalized search algorithms and a personalization engine where the input is given by a set of spatial entities and a set of reference locations. We use a Bi-directional Neural Associative Memory (BNAM) architecture as the basic mechanism to elicit user's interests and preferences, and to develop personalized search algorithms. The BNAM defines the minimal two-layer nonlinear feedback network in that it uses less information than other feedback networks (Kosko 1987, 1988). It can be considered as an extension of the Hopfield network, which allows the storage and recall of heteroassociated patterns (A_1, B_1), …, (A_m, B_m), where $A \in \{0, 1\}^p$ and $A \in \{0, 1\}^q$ (p and q are respectively the number of neurons that activate pattern A/B) by a recurrent network. The term "bi-directional" refers to forward and backward information flows to produce a two-way associative search for stored associations (Ai, Bi).

The BNAM uses iterative processes to recall the first pattern at the first layer (i.e., the input layer) and the second pattern at the second layer (i.e., the output layer). Information passes forward from the input layer to the output layer through the p-by-n connection matrix M, and backward through the transposed matrix M^T. The BNAM recall information through performing the following steps (Freeman and Skapura 1991, p133):

1) Apply an initial vector pair (x0, y0) to the processing elements of the BNAM.
2) Propagate the information (activated neural pattern) from the x input layer to the y output layer, and update the values on the y layer units, which is so-called forward.
3) Propagate the updated y layer information back to the x layer and update the x-layer units, which is so-called backward.
4) Repeat step 2 and 3 until there is no further change in the units on each layer.

For example, suppose that pattern A is associated with pattern B. The BNAM will recall (a part of) pattern B when a part of pattern A is activated at the first layer. Through iterative processes, the network will evoke a complete version of pattern A at the input layer and a complete pattern B at the output layer.

The BNAM encapsulates different forms of semantic and spatial associations between a set of spatial entities of interest and a set of reference locations. An algorithm output returns the reference entity that is the most centrally located with respect to a set of spatial entities that represent user's interests and preferences. The reference entities and the set of spatial entities of interest are linked according to those associations that combine spatial and semantic criteria, and user's interests and preferences. They are implemented as a BNAM that bears several advantages to spatial Web personalization: unsupervised search and learning, no input/output data samples and maximum flexibility with no training during the computation processing (cf. refer to see Kosko, 1992 for a survey on BNAM). This BNAM employs a form of "winner takes all" mechanism. The computation is unsupervised, and the complexity of the network construction is minimal.

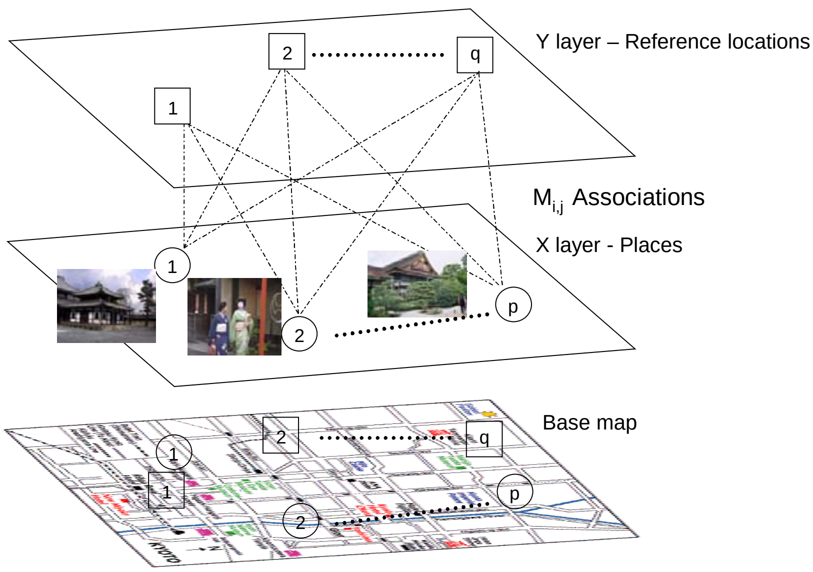


Figure 4.3 Bi-directional Neural Associative Memory (BNAM) principles

The BNAM provides an efficient means to store and recall paired associations between a set of spatial entities of interests and a set of reference locations. The BNAM is initialised by two layers X and Y where X={x_1, x_2, …, x_p} denotes the set of spatial entities of interest, Y={y_1, y_2, …, y_q} the set of reference locations (no semantic criteria are attached to these reference locations but they can be added to the associative memory with some minor adaptations) (Figure 4.3). The BNAM has *p* vectors in the X layer, *q* vectors in the Y layer. We define a weight matrix M where $M_{i,j}$ reflects the strength of the association between x_i and y_j for i=1,.., p and j=1, …, k. These matrix values are flexibly initialised and defined by various combinations of spatial and semantic criteria, user preference pattern, which corresponds to different personalized search algorithms. We give the user an opportunity to choose among several search algorithms in order to explore different output alternatives and evaluate the one that is the most appropriate to her/his intentions. One peculiarity of the BNAM applied in the spatial Web environment, relies in the fact that the user selects the nodes of the two layers dynamically. The BNAM is able to explore different output alternatives and to evaluate the reference location that is the most appropriate to the user intentions.

The BNAM-based search strategy consists of an iterative forward and backward process according to a combination of spatial and semantic criteria, the elicitation of user preferences. The forward part of the associative memory recalls the best reference entity from where the user would like to plan her/his actions in the spatial Web environment. The backward part of the associative memory recalls the spatial entities that might be of interest to the user. The iterative search processes keep recalling until arriving at the final search results that could best fit user's interests and preferences in a given spatial environment. The final results consist of a set of spatial entities of interest and the best reference location.

4.3.2 Representation of spatial entities on the Web

In the context of this research a set of spatial entities and a set of reference entities is represented preferably at the Web design level. The spatial and semantic contents of a Web urban space are encoded by a design process that identifies and models spatial entities of interest. This section introduces some basic notations to support our modelling approach. A given set of spatial entities E embedded in a Web urban space is represented as

$E = \{e_1, e_2, \dots e_p\}$
where p is an integer that denotes the number of spatial entities of interest

Spatial entities implicitly materialised on the Web and associated to hyperlinks generate a graph that relates them in the Web space. A spatial entity is likely to possess a semantic content and some spatial properties:

$e = (spatial(e), semantic(e))$
where *spatial(e)* denotes the spatial component and *semantic(e)* the semantic component of the spatial entity e

The spatial component describes the location of a spatial entity as an abstract data type and is given as follows

$spatial(e) = (x, y)$
where (x, y) denotes the coordinates of the spatial entity e in a two dimensional space

The semantic component can be considered as an h-dimensional vector that specifies the semantic parameters of a given spatial entity

$semantic(e) = \{w(c_1, e), w(c_2, e), \dots w(c_h, e)\}$

where h is an integer that denotes the number of semantic parameters; $w(c_i, e)$ gives the relevance of the spatial entity e associated to a semantic class c_i.

A semantic class corresponds to an abstract form of entities that share some semantic properties. Relevance values can be given by membership values given by the unit interval [0,1] and representing the degree of association or membership degree of an entity with respect to a given semantic class C_i.

4.3.3 Semantic information

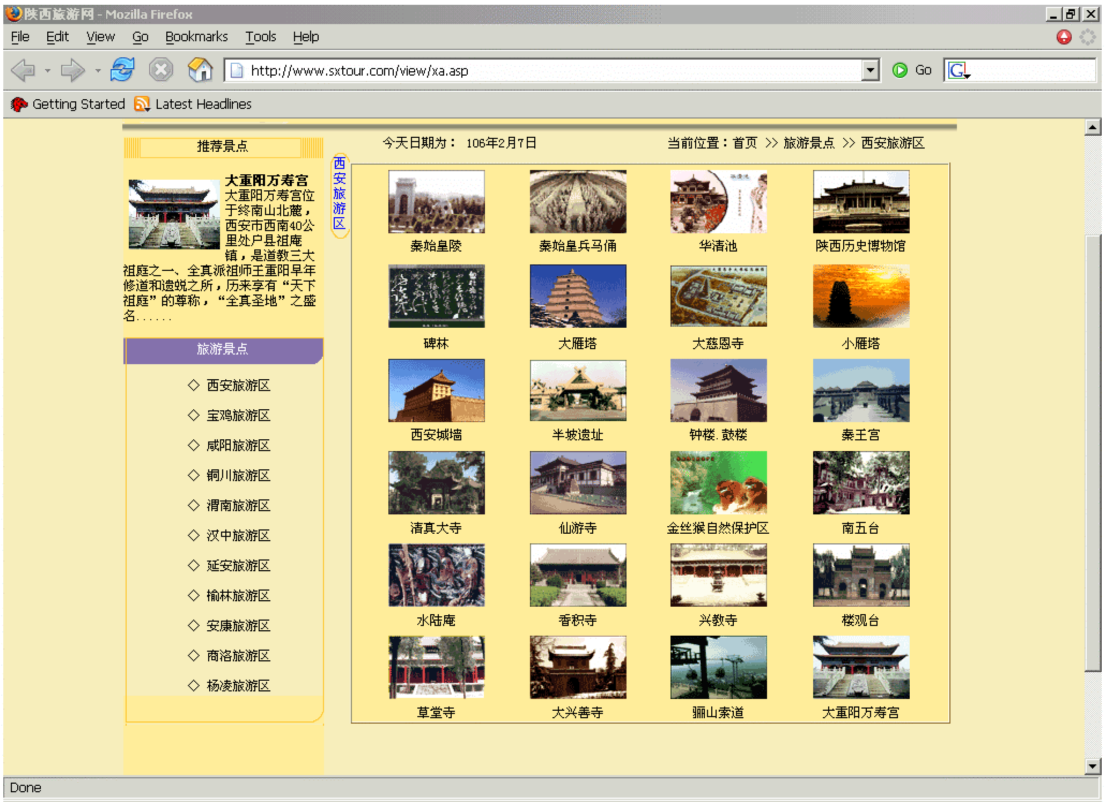


Figure 4.4 A tourism service interface[4]

In order to keep user inputs minimal, we use the concepts of *image schemata and affordance* (Gibson 1979) to approximate the user intentions. Affordance relies in the idea that the appearance of a tool or agent suggests what function can be expected from it (Lieberman and Selker 2002). This principle can be modelled as a relationship between an observer and the environment. A close concept related to the one of affordance is the notion of image schemata. An image schemata is associated to the graphical representation of an affordance. Image schemata are recurring imaginative patterns that help human to comprehend and structure their experience while moving and acting in their environment (Johnson 1987). They are closely related to the

[4] http://www.sxtour.com/view/

concept of affordance that qualifies the possible actions offered to do with them (Gibson 1979). Image schemata and affordance have been already applied to the design of spatial interfaces to favour interaction between people and real-world objects (Kuhn 1996). These concepts are applied to the selection of the spatial entities that are of interest to the user, thus assuming that these image schemata and affordance relate to the opportunities and actions she/he would like to take and expect in the environment materialized by the spatial Web interface.

Figure 4.4 illustrates a user interface enriched with these concepts in the tourism domain. This illustration is the main interface for e-tourism services in Shaanxi, a famous tourism place for Chinese culture and religion. Spatial entities are presented to the user using image schemata in order to approximate her/his range of interests. Image schemata reflects one or several application-dependent aspects relevant to spatial entities in a given environment in a graphic form.

Spatial entities of interest are represented as modelling objects classified semantically and located in space. An entity x_i is symbolised by an image schemata that acts as a visual label associated to it. The memberships of an entity x_i with respect to some thematic classes C_1, C_2, …, C_k are given by the values x_i^1, x_i^2, …, x_i^k that denote some fuzzy quantifiers with $1 \le i \le p$ (Figure 4.5). The semantic classes are conceptually represented in a domain ontology. Ontology is "specification of a conceptualization" (Gruber 1995). An ontology can be viewed as a special kind of semantic network representing the terminology, concepts, and the relationships among these concepts related to a particular application domain. In a Web urban space, a domain ontology (e.g. for tourism) can be designed as a set of concepts (e.g. garden, museum), which are linked with some relationships (e.g., semantic and syntactic similarity). x_i^h denotes the degree of membership of x_i to the class C_h and it is bounded by the unit interval [0,1], with $1 \le h \le k$. A value x_i^h that tends to 0 (resp. 1) denotes a low (resp. high) degree of membership to the class C_h. An entity x_i can belong to several classes C_1, C_2, …, C_k at different degrees, and the sum of the membership values x_i^1, x_i^2, …, x_i^k can be higher than 1. This latter property reflects the fact that some classes are semantically close, i.e. they are not semantically independent. Reference locations refer to some possible locations where the user could act from (e.g. hotels), to visit the spatial entities of interest. This is exemplified by the fact that a spatial entity x_i with a high degree of membership x_i^h to a class C_h is likely to also have high membership values with respect to the classes that are semantically close to C_h.

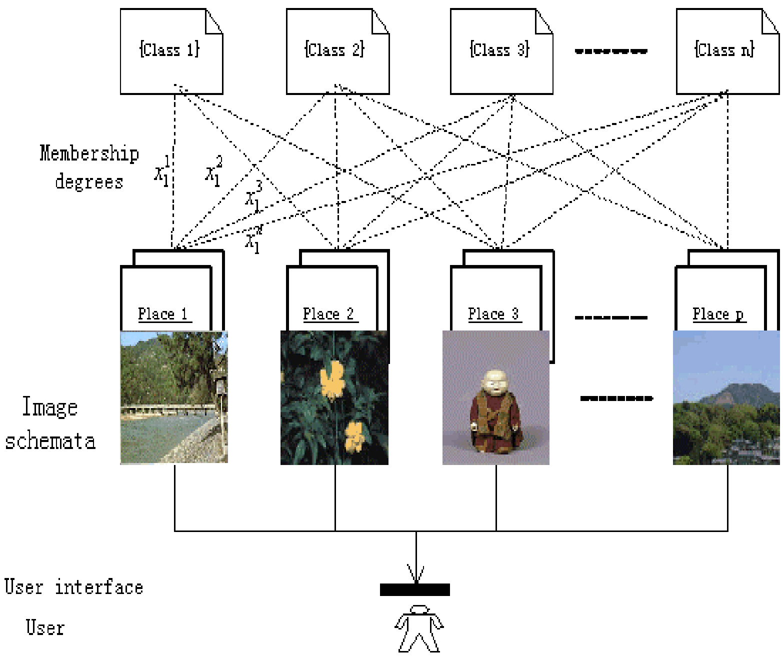


Figure 4.5 Semantic components

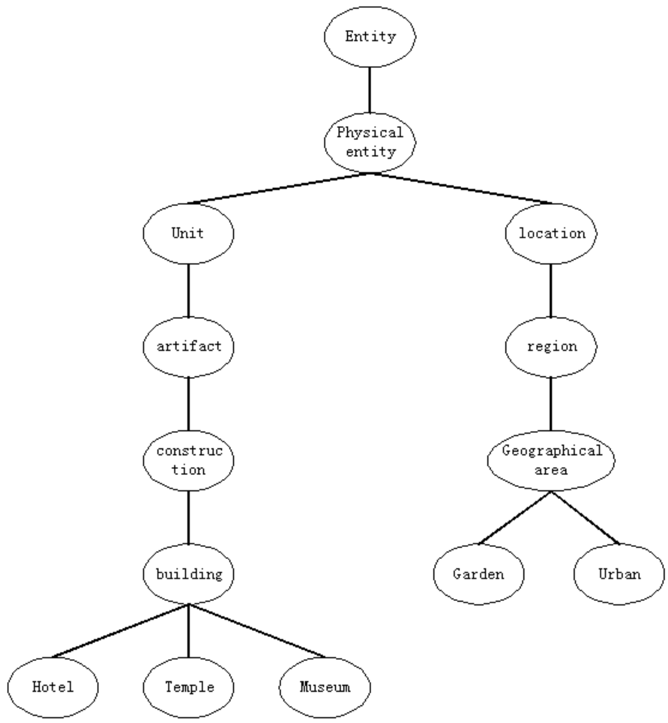


Figure 4.6 Terminological ontology extracted from WordNet

Membership degrees:

Museum: $x_1^1 = 0.6$,

Temple: $x_1^2 = 0.9$,

Garden: $x_1^3 = 0.8$,

Urban: $x_1^4 = 0.05$

Figure 4.7 Spatial entity example: Toji Temple

As an example, let us consider some spatial entities of interest in a given urban space. We build a terminological ontology relevant to travel and tourism applications from WordNet (Miller 1995) (Figure 4.6). Then spatial entities can be classified according to a set of classes $\{C_1, C_2, C_3, C_4\}$ with C_1='Museum', C_2='Temple', C_3='Garden', C_4='Urban'. The image schemata presented in Figure 4.7 illustrates the example of the Toji Temple in Kyoto labelled as x_1. This photograph exhibits a view of the temple surrounded by a park. This can be intuitively interpreted by a relatively high membership to the classes C1, C2 and C3 (one can remark a semantic dependence between the classes C1 and C2), and low to the class C4.

4.4 Spatial proximity and similarity measure

A central relationship of interest in information retrieval and personalization on the Web is the notion of similarity (Baeza-Yates and Ribeiro-Neto 1999, Mobasher *et al.* 2000, Pitkow *et al.* 2002, Jin and Mobasher 2003). In information retrieval, similarity measures are used to identify the degree of correspondence between two information entities, that is, how similar or dissimilar two entities are. With respect to our application domain, applying a similarity measure to spatial Web applications implies to explore to which degree a semantic relationship of similarity is influenced by space.

Spatial proximity is a non-straightforward notion distinct from similarity concepts developed in the semantic domain. It is directly related to complexity of spatial semantics (e.g., the spatial distribution) and influences user's information needs, interests and preferences, when a use is considering or acting in a spatial environment. Most spatial applications take spatial proximity as the Euclidean distance, rather than integrating a contextual component in the notion (Yao and Thill 2005). This leads to spatial proximity and similarity models which are limited by designer's perception, and that are not flexible enough for users with different backgrounds and interests, and from diverse communities. In particular, there is a lack of semantic-based techniques to personalize spatial information services with respect to diverse user communities that are undergoing an expansion and transformation with increasing user numbers and applications (Mountrakis *et al.* 2005).

To the best of our knowledge, a few similarity measures take into account user information such as interests and preferences. User's interests and preferences substantially influence human's perceptions of and interactions with an information space. Although personalization techniques applied to the Web and that recently emerged are expected to provide customized information services. Much useful information is still not considered. This section introduces an approach for the refinement of spatial proximity and similarity measures taking into account user's perceptions, interests and preferences. The approach is based at the conceptual level, on a user-centric conceptual map that reflects user's interests and preferences. The approach also allows for an integration of spatial proximity and semantic similarity measures within the context of spatial Web personalization.

One of our objectives is to explore the effects that user's interests and preferences have on spatial proximity and similarity measures for spatial Web personalization. This requires comprehensively understanding and modeling dynamic interactions between human and the spatial Web.

4.4.1 User's interests and preferences

User's interests and preferences can be represented with a preference pattern $prefPattern(pref_1, pref_2, ..., pref_n)$, where $pref_i$, denotes a user preference index, that is, to which degree a given user is interested in properties related to a semantic class C_i. A user preference index $pref_i$ is bounded by the unit interval [0, 1], close to 0 when a given user dislikes entities relevant to a semantic class, conversely close to 1 when the user likes them very much. User preference indexes can be extracted from user's behaviors in a Web information space.

User preference patterns can be used to distinguish whether a given user likes or dislikes with respect to a semantic class. Usually an average value (e.g. 0.5) is used, a preference index $pref_i \geq 0.5$ denotes that the user has positive evidence for corresponding interests, otherwise $pref_i < 0.5$ the user dislikes. Then user preference indexes can be divided into two sets: a positive set $like\ (pref_1, pref_2, ..., pref_{like})$ and a negative set $disl\ (pref_1, pref_2, ..., pref_{disl})$, where $like+disl=n$, *like* denotes the number of positive preference indexes, and *disl* the number of negative preference indexes in a given user preference pattern. Correspondingly membership degrees of entities can also be divided into user preferable set $like\ \{w(C_1, e), w(C_2, e), ...\ w(C_{like}, e)\}$, and user nonpreferable set $disl\ \{w(C_1, e), w(C_2, e), ...\ w(C_{disl}, e)\}$.

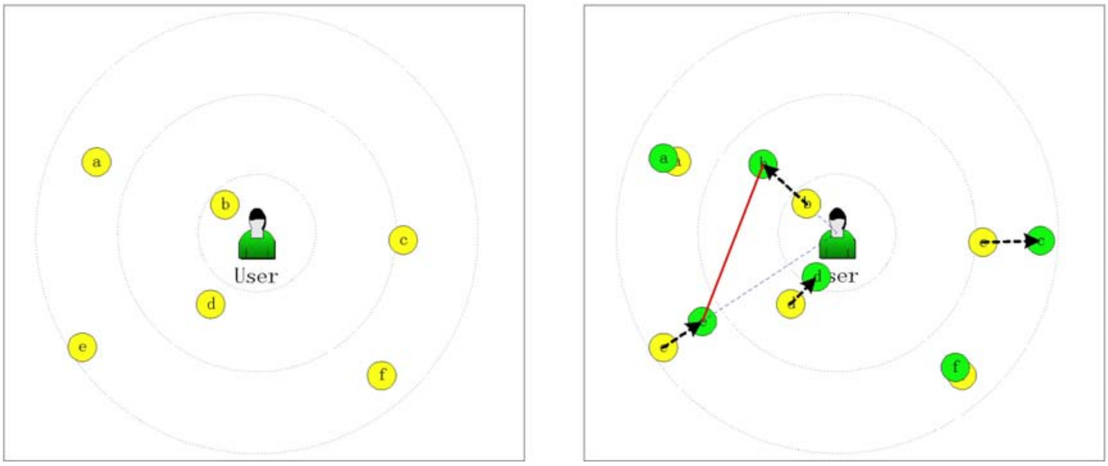
4.4.2 User-centric conceptual map

A conceptual map is a hierarchical diagram used to represent a set of concepts in the form of propositions (www.ettc.net, Novak 2003). It's an effective tool for representing and organizing knowledge, and for interpreting and conveying complex conceptual information in a clear and understandable way. A conceptual map of an urban space lies in a form relatively similar to a sort of geographical map (Cahuzac *et al.* 2004), this allowing people to remember and directly or indirectly interact with the embedded spatial entities. It describes a special image of the city space human beings have, characterised by informal landmarks, subjective distances and sizes, and intuitive way findings (Lynch, 1960, Gaye *et al.* 2003).

Generally, human beings perceive and interact with spatial entities in a given environment in an egocentric way (Franklin and Tversky 1990, Franklin *et al.* 1992, Hörnig *et al.* 2000). This gives at least four cognitive clues to understand the way that human perceive and interact with the spatial environment. First, a user's conceptual map of a spatial environment is surely a very personal representation. Secondly, a conceptual map is designed from an egocentric point of view where the user is generally the central of interest. When such a user is acting in a spatial environment, which makes her/his current location the central of the conceptual map. Thirdly, this leads to the fact that human evaluate spatial entities and relationships on the basis of such an egocentric conceptual map. Last but not the least, human make decisions influenced by views from their own conceptual map, and update their conceptual map continuously according to their experiences.

In (Tezuka *et al.* 2001) Web-based inference rules are introduced to infer conceptual prepositions that denote spatial relationships (e.g. near, between). Their study showed that geographical relationships are distorted in a conceptual map by the popularity of geographical objects *per se*, path between them, human-made and/or physical borderlines. (Hirtle and Jonides 1985) believed that the hierarchical arrangement of spatial entities in a given environment affects the representation of a conceptual map. We argue that user's interests and preferences are amongst the most important factors of distortion in a conceptual map. Conceptual maps are surely influenced by social, cultural and knowledge criteria, and by the experiences and interests the user have of an underlying space. For example, A historian may consider it's "near" to a museum but "far" to a garden given other factors are similar or equivalent, e.g. distance.

In order to illustrate the impact of user's interests and preferences on the derivation of conceptual maps, let us consider a scenario of a given user interacting with some spatial entities in a spatial environment. We also assume that a given user usually considers spatial entities of interest nearer and nonpreferable entities further to her/his current location. The way that these spatial relationships are interpreted and represented has a direct influence on the structure of a conceptual map. For instance, things of interest are likely to be decreased, and conversely. Figure 4.8 illustrates a scenario of a user-centric conceptual map representing spatial entities in a given environment. The left is a sketch of the original map of some spatial entities located in a given environment. The right presents a conceptual map as it might be conceptualised in the user's mind. The grayish circles denote the "real" locations of the spatial entities while the dark-gray circles the locations of these entities in the conceptual map. Compared to the original map, distance from the user to these spatial entities (d, e, f) are decreased when they present of interest to her/him, and conversely (a, b, c). This generates a contextual representation where spatial and semantics form the underlying dimension of a conceptual map.



(a) Distribution of spatial entities in real world (b) user's conceptual map

Figure 4.8 User-centric conceptual map

Conceptual maps take into account the distribution of spatial entities in a given environment, and user's interests and preferences. Contextual distances as they appear in a conceptual map depend on Euclidean distances between spatial entities and user's location, and a semantic distortion denoting to which degree the distance is distorted in the conceptual map. Such a distortion can be qualified as a distorted distance of a spatial entity with respect to a given user's conceptual map is given as,

$$distortD(e,u) = d(e,u) \times distortDeg(e,u) \tag{1}$$

where $d(e,u)$ denotes the Euclidean distance between aspatial entity e and

user's location u, and $distortDeg(e,u)$ the distorted degree.

The distorted degree describes to which degree the location of a specific spatial entities is distorted in a user's conceptual map. This is related to its semantic content and user's preference indexes that is the basis for the design of conceptual maps. It is given as

$$distortDeg(e,u) = \frac{1}{like}\sum_{i=1}^{like}[pref_i \times w(C_i,e)] + \frac{1}{disl}\sum_{j=1}^{disl}[(pref_j - 1) \times w(C_j,e)] \tag{2}$$

The distorted degree is bounded by the unit interval [-1, 1]. The distorted degree is represented as a function of user's interests and preferences (e.g. $pref_i$), and membership degrees of a spatial entity. The distortions in positive direction and in negative direction are determined by preference indexes in positive set (*like*) and those in negative set (*disl*), respectively. The distorted degree combines positive and negative evidences of user's preference patterns. A conceptual map provides the basis for spatial proximity and semantic similarity measure between spatial entities in a given environment.

4.4.3 Spatial proximity and semantic similarity

Geometric models and multidimensional scaling models are the most influential approaches to analyze the similarity between some entities (Goldstone and Son 2005). In multidimensional scaling routines, the distance between entities x_i and y_j is computed as follows,

$$d(x_i, y_j) = [\sum_{k=1}^{n} | x_i^k - y_j^k |^r]^{(1/r)} \tag{3}$$

Where n is the number of dimensions of entities, x_i^k, y_j^k denote the membership degrees relevant to the semantic class C_k of entities x_i and y_j, r a parameter for different spatial metrics.

The Euclidean distance is a specialized form of multidimensional scaling distance with r=2, a popular metric used in human similarity judgment in the underlying physical environment.

Proximity and similarity measures are sensitive to judgment contextual knowledge (Roberts and Wedell 1994, Goldstone *et al.* 1997). Semantic similarities are usually represented and judged by contextual factors such as, range-frequency principles of stimulus values of entities (Parducci 1965), relationships with other entities (Sjöberg 1972) especially with close neighbors (Krumhansl 1978), and variation of features of entities (Tversky 1977). Perceptions in the environment, so-called contextual knowledge, are influenced by several cognitive factors that range over the semantic, temporal and spatial dimensions. We already mentioned that one of the important factors that constrain human actions in the environment is the notion of spatial proximity. This should be modelled by a rule stating that the interest showed by a given user to a specific spatial entity increases when similar entities are located nearby. This is particularly important for a specific user acting or planning to act in a given urban space, as she/he will consider space as a recipient where her/his interests for a given spatial entity can be reinforced in the presence of similar entities nearby.

In the context of a Web interface, and at the design level, although spatial entities of interest and reference locations are geo-referenced, this information should not be presented to the user in order to not interfere with the approximation of her/his preferences. Ideally, a Web interface should encompass information sometimes explicitly (image schemata of the representative spatial entities) but also implicitly (location of the spatial entities and the reference location in the city, proximity between them).

In the spatial domain, the distance between spatial entities is influenced by the overall structure of a given spatial distribution. The proximity between two locations is usually approximated as an inverse of the distance factor. (Worboys 1996) defined a "relativised distance" concept to measure spatial proximity between spatial entities. In a related work (Worboys 2001), three approaches are discussed to experiential analysis on spatial proximity in environment space: nearness neighbourhoods as regions with broad boundaries, fuzzy nearness and distance measures, and four-valued logic. This reflects the fact, observed in qualitative studies (Sadalla *et al.* 1980, Tversky 1993), that the distance from a region α to a distant region β should be magnified when the number of regions near α increases, and *vice versa* (Worboys 1996). The relativised distance introduced by Worboys normalises the conventional Euclidean distance between a region A and a region B by a dividing factor that gives a form of contextual value to that measure. This dividing factor is the average of the Euclidean distance between the region A and all the regions considered as part of the environment.

However, these approaches are valid when considering a homogenous set of spatial entities, and where these entities are semantically different. We also believe that, semantic category, level of abstraction and scale have reasonable effects on spatial proximity measures. Let us consider the case illustrated in Figure 4.9. The overall distribution changes from spatial structure (a) to (b), but has less influence on the perception on spatial proximity between London and Paris. The reason behind is that the four cities added in spatial structure (b) are smaller than London and Paris, that is, at lower level (than Paris and London) in hierarchical taxonomy of cities. Whereas, what if several grand cosmopolises (similar to or even larger than London and Paris) emerge near London? In the latter case, the approaches described above work well: the distance between London and Paris is magnified.

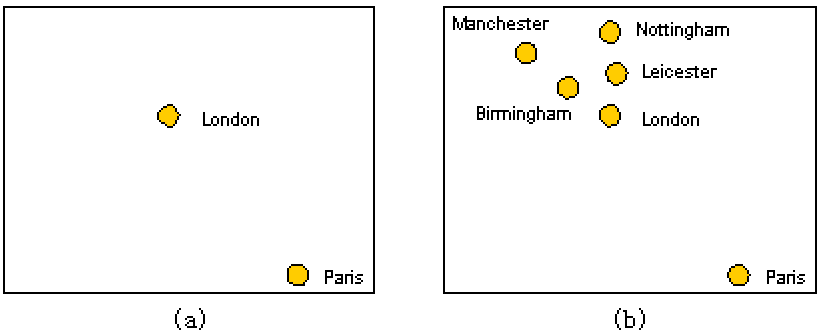


Figure 4.9 Spatial proximity scenario

4.4.4 A definition for proximity and similarity measure

We retain a contextual modelling of the distance and then of the proximity between two spatial entities. A generic concept of "contextual proximity" is introduced to model a distance relationship, but also taking into account the notion of similarity within it. Our definition can be considered as a generic approach for spatial proximity and semantic measures, taking into account of contextual knowledge such as relationships and spatial distribution of spatial entities under consideration. It reflects two widely acknowledged rule that are observed both in semantic similarity and spatial proximity research fields:

Rule 1: *The spatial/semantic relationship between two entities is asymmetrical.*
Rule 2: *The spatial/semantic relationship from one entity to another is weakened when the number of its nearby neighbours increases; magnified when the number of its distant neighbours increases; and vice versa.*

However, to the best of our knowledge, few spatial proximity and semantic similarity measures applied in information retrieval and personalization services take into account the two rules, especially rule 2. We make the distinction between intra-contextual proximity measures designed for a homogenous set of spatial entities, and inter- contextual proximity for spatial entities from different categories and/or levels of abstraction, respectively. The *contextual proximity* gives a form of inverse distance bounded by the unit interval [0,1].

Intra- Contextual proximity
The contextual proximity between two spatial entities x_i, x_j $\in X = \{x_1, x_2, ..., x_p\}$, where p denotes the number of elements in set X, is given as,

$$CP(x_i, x_j) = \frac{1}{1 + CD(x_i, x_j)^2} \tag{4}$$

Where $CD(x_i, x_j)$ denotes the contextual distance between x_i *and* x_j. The higher $CP(x_i, x_j)$, the closer x_i is to x_j, the lower $CP(x_i, x_j)$, the distant x_i is to x_j.

The contextual distance between x_i and x_j is given as,

$$CD(x_i, y_j) = \frac{d(x_i, y_j)}{d(x_i, X)} \tag{5}$$

Where $d(x_i, x_j)$ denotes the Euclidean distance between x_i and x_j, $d(x_i, X)$ the average distance between x_i and the other entities in X.

The average distance between x_i and the other entities in X is computed as,

$$d(x_i, X) = \frac{1}{p-1} \sum_{j=1, j \neq i}^{p} d(x_i, x_j) \tag{6}$$

Inter- Contextual proximity

The inter- contextual proximity is a generalized form of the relativised distance. The *contextual distance* normalizes the conventional Euclidean distance between a set of spatial entities A and a set of reference locations B by a dividing factor that gives a form of contextual value to that measure. The dividing factor is given by a function of two factors. The first is the average of all distances between the entities of one set A (in which α is located) with respect to the reference locations of a second set B (in which β is located). The second is the average of all distance between α and other entities in A. The *contextual distance* between a region α of set A and region β of set B magnifies when the number of regions of set B near the regions of set A increases, and *vice-versa*. The Contextual proximity between a spatial entity $x_i \in X$, and another $y_j \in Y = \{y_1, y_2, ..., y_q\}$, q is the number of elements in set Y, can be defined as follows,

$$CP(x_i, y_j) = \frac{1}{1 + CD(x_i, y_j)^2} \tag{7}$$

Where $CD(x_i, y_j)$ denotes the contextual distance between x_i and y_j. The higher $CP(x_i, y_j)$ the closer x_i to y_j, the lower $CP(x_i, y_j)$ the distant x_i to y_j.

The contextual distance between two entities from different sets is inversely proportional to two forms of distance: intra-distance and inter-distance. The former refers to the distance to entities in a same set, and the latter, the distance to entities in a different set with respect to entities under comparison. The inter-distance considers the first entity as the reference, reflecting an asymmetric characteristic of cognitive proximity measure. The contextual distance between x_i and y_j is given as,

$$CD(x_i, y_j) = \frac{d(x_i, y_j)}{\sqrt{d(x_i, Y)^2 + d(x_i, X)^2}} \tag{8}$$

Where $d(x_i, Y)$ denotes the distance between x_i and Y.

The definition above gives a form of generalisation of Worboys's definition of relativised distance as the dividing factor is here the average of all distances between the regions of one set with respect to the regions of a second set. We also slightly modify the definition of the relativised proximity introduced by Worboys in the same work (1996) by adding a square factor to the contextual distance in the denominator in order to maximize contextual proximities for small distances (vs. minimizing contextual proximities for large distances), and to extend the amplitude of values within the unit interval. The distance between xi and Y refers to the average distance between xi and spatial entities in Y, is given as,

$$d(x_i, Y) = \frac{1}{q} \sum_{j=1}^{q} d(x_i, y_j) \tag{9}$$

The contextual proximity concept still needs to be extended in the semantic domain. At least two minor adaptations need to be made. The first consideration is that the semantic domain is essentially multidimensional. Unlike in the spatial domain, the semantic dimensions are not perpendicular to each other, but have some interrelationships in between. The second is to employ user preference pattern for refinement of semantic similarity measures. User's interests and preferences are an essential component for similarity measures, and can be employed to refine a semantic similarity measure in that they are semantically related to entities of interest, and influence some processes such as decision-makings and way-findings.

In order to measure semantic similarities, it's necessary to take into account rich semantic dimensionality and intricate interrelationships. We thus distinguish between similarity contents of spatial entities related to user preferable dimensions and those related to user non-preferable dimensions. The semantic distance is given as follows,

$$d(x_i, y_j) = \lambda[\sum_{k=1}^{like} | x_i^k - y_j^k |^2]^{(1/2)} + (1-\lambda)[\sum_{k=1}^{disl} | x_i^k - y_j^k |^2]^{(1/2)} \quad (10)$$

Where λ is a constant valued as a real number between 0 and 1. It allows adjusting the respective influence of user preferable set and user nonpreferable set of semantic membership degrees of entities under consideration (a suggested value for λ is fixed to 0.8 in order to emphasize positive evidence on user's interests and preferences).

4.4.5 Refinement of semantic similarity measure

Qualitative measures of similarity between spatial entities require the consideration of contextual knowledge in semantic structures of an application domain. As we mentioned previously, the level of abstraction and category of spatial entities are two criteria for the refinement of spatial proximity and semantic similarity measurements. These factors can be extracted from semantic interrelationships between classes in an ontological taxonomy. We exploit a domain ontology to refine our approaches for proximity and similarity measure. The hierarchical/taxonomical domain ontology in our modeling context is based on is-a relationship (Figure 4.5). The is-a relation is transitive and asymmetric *per se*, describes paths and relationships from a specific to a more general concept in hierarchical structures.

A hierarchical domain ontology consists of a set of semantic classes N and links L. Classes are labelled with distinct labels. Links connect classes with different relationships e.g. is-a and part-whole. Let H be a hierarchical domain, Root (H) the root. The depth of a class is the number of links between Root (H) and the class. The least common ancestor of two classes is the deepest subsumer of them. The relationships between two semantic classes can be represented either by the number of links connecting them in the hierarchical structure, or by a function of the number of their common and distinctive super classes. The links and classes are also assigned weights denoting different importance, based on depth and density of semantic classes in class hierarchy.

Let $sup(C_1)$ be the set of super classes of C_1 in the hierarchical domain ontology,
$deep(C_1)$ the depth of C_1,
$sib(C_1)$ the number of siblings of C_1 with the most specific, common ancestor,
$sup(C_1/C_2)$ the set of super classes of C_1 but not of C_2,
$dis(C_1, C_2)$ the number of links between C_1 and C_2,
$LCA(C_1, C_2)$ the least common ancestor of C_1 and C_2.

Similarity between two semantic classes C_1, C_2 in a given hierarchical domain ontology is given as follows:

$$sim(C_1, C_2) = \frac{\eta \,|\sup(C_1) \cap \sup(C_2)|}{|\sup(C_1) \cup \sup(C_2)| + \alpha\,|\sup(C_1 / C_2)| - (1-\alpha)\,|\sup(C_2 / C_1)|} \quad (11)$$

Where α is the parameter bounded by the unit interval [0, 1], to adapt the weights of the distinct sets sup(C_1/C_2) and super(C_2/C_2) in comparison. η denotes the depth parameter.

The weight α is determined as a function of the distance between semantics C_1, C_2 and the least common ancestor of both classes, and the number of sibling C1, C2. It is given as

$$\alpha(C_1, C_2) = \frac{dis(C_1, LCA(C_1, C_2)) \times sib(C_1)}{dis(C_1, LCA(C_1, C_2)) \times sib(C_1) + dis(C_2, LCA(C_1, C_2)) \times sib(C_2)} \quad (12)$$

The depth parameter η is given as,

$$\eta = \frac{2deep(LCA(C_1, C_2))}{deep(C_1) + deep(C_2)} \quad (13)$$

The similarity function yields values bounded by the unit interval [0,1]. The maximum value 1 occurs *iff* the two semantic classes under comparsion are equivalent, that is, C_1=C_2. The similarity function reflects an asymmetric relationship between two semantic classes. However, the semantic interrelationships given in a hierarchical structure of semantic classes are not sufficient to distinguish one class from another (Rodriguez and Egenhofer 1999). For example, let us consider several classes *Hotel*, *Temp* and *Museum*, they have common super classes, deepth, and same similarity/distance to other semantic classes, thus it's not capable of distincting a hotel, a temp or a museum only with knowledge extracted from a hierarchical domain ontology illustrated in Figure 3. In our approaches these knowledge can be used to refine semantic similarity measures described earlier based on geometric model and membership degrees.

Usually an entity may pertain to more than one semantic class in terms of membership degrees. This requires studying the impact of interrelationships among several classes on similarities between two entities that are related to those classes. Since these semantic classes are somewhat interrelated to each other, we integrate these semantic relationships with the geometric model based similarity measure in semantic domain. In the case of an asymmetric similarity measure, there are an original entity *Oo* and a target entity *Ot*. Let *Oo(C)* denote the set of semantic classes to which the origin entity relates, *Ot(C)* the set of classes of the target entity. Given a class $C_i \in Oo(C)$, semantic relationships from it to the classes in *Ot(C)* is computed as,

$$sim_{C_i} = \frac{1}{|Ot(C)|-1}\sum_{j=1, j\neq i}^{|Ot(C)|} sim(C_i, C_j) \tag{14}$$

The semantic distance in Equation 10 is then refined as follows,

$$d(x_i, y_j) = \lambda[\sum_{k=1}^{like} |\frac{x_{ik} - y_{jk}}{sim_{C_i}}|^2]^{(1/2)} + (1-\lambda)[\sum_{k=1}^{disl} |\frac{x_{ik} - y_{jk}}{sim_{C_i}}|^2]^{(1/2)} \tag{15}$$

This section proposes an approach for the construction of a user-centric conceptual map based on user’s interests and preferences, and spatial proximity and similarity measures based on the geometric model. Moreover, it’s refined with semantic relationships extracted from hierarchical representation of an application domain. The spatial proximity and similarity model takes into account the overall structure of a given distribution of spatial entities in the environment.

4.5 Discussion

We introduce an integrated framework for user preference elicitation and personalized search strategy on the spatial Web, a novel issue in the context of “spatially enhanced” Web information retrieval. This chapter introduces a conceptual framework for spatial Web personalization services. The framework is based on the integration of spatial criteria and Web information retrieval to deliver Web information services taking into account user preferences. Within the context of personalization services, user preference elicitation and personalized search processes can be developed and supported by a BNAM that triggers an iterative forward and backward process that recalls the reference location and several spatial entities of interest that best fits user preferences. It also uses image schemata and affordance concepts for preference elicitation in a Web-based urban environment. Image schemata and affordance concepts are employed to design the spatial user interface and improve the interactions between Web information system and the user.

Chapter 5 Personalized search strategy on the spatial Web

5.1 Introduction

This chapter introduces a spectrum of personalized search strategies on the top of the personalization services framework developed in chapter 4. These strategies are used to search for spatial information on the Web taking into account user's interests and preferences. User's interests and preferences can be either explicitly provided by the user at the interface level, or inferred from user's descriptions and behaviors while interacting with spatial entities on the Web.

Differences inherent in applications should be reflected by the ways to personalize information services. Moreover, personalization services be flexible enough for satisfying user's demands, that is, opportunities to operate on options about personalization criteria. Google recently provides spatially personalized services, however, it is too limited to offer flexibility to the user. Google search engine ranks higher and presents Web pages whose language and content are related to where the user is linked to the Web, however, disregarding of user's backgrounds, interests and preferences. As the range of search mechanisms that can be implemented within the BNAM architecture is relatively large, we explore and study different personalized search algorithms in order to offer more flexibility for various kinds of application situations on the spatial Web. These algorithms are based on BNAM search and learning mechanisms, to infer user preferences and to recall a set of spatial entities tailored to user's interests and preferences. The personalized search component is ensured by a mechanism that derives user preferences, according to different criteria and ordering value functions. The basic rationale is, when the user shows interests to some spatial entities, the system induces her/his personal preference pattern, and then generates an inference process to search for and recommends the highest entities of interest. User preferences are inferred from the previous observation that, the higher the number of closer spatial entities of similar interest to a given spatial entity, the higher the value is given to this entity. The whole personalized search process is supported by the BNAM search architecture.

The entities of interest and reference entities act as basic nodes to construct a neural associative memory. The personalized search algorithms are expected to recall the best reference entity and a set of spatial entities according to user's interests and preferences elicited from which she/he would like to plan her/his actions in the city.

Without loss of generality, those reference entities can be illustrated by a set of salient locations distributed in the city. These spatial entities of interest in the city are ranked according to their associations with the best reference entity.

The remainder of this chapter is organized as follows. Section 5.2 introduces user preference elicitation processes based on BNAM-based learning mechanisms, and the refinement of user preferences and search strategies. Section 5.3 investigates semantic contents and user preference patterns, and proposes global and local personalized search strategies on the spatial Web. Section 5.4 concludes the chapter.

5.2 Personalized search strategies

5.2.1 Personalized search workflow

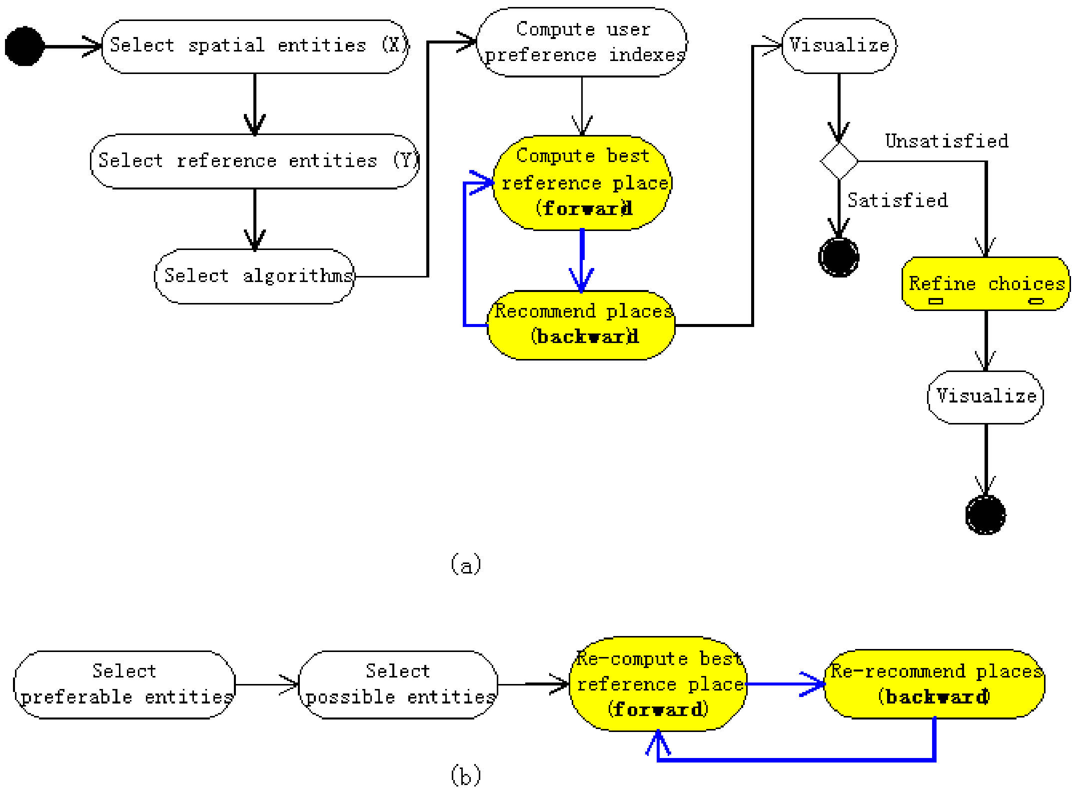


Figure 5.1 Personalized search strategy (UML activity diagram)

The personalized search model consists of two stages: search initialisation and refining process. User preference elicitation is involved in the personalized search model. Figure 5.1 describes the complete personalized search procedure. At the initial user interface, the user selects spatial entities and reference entities of interest. After the user submitted her/his initial selections, the first search stage computes user preference indexes. Therefore, the system performs an iterative process based on the the

BNAM architecture. This leads to recall the reference location which is the most centrally located with respect to the spatial entities selected, or activated (forward), and to recommend several spatial entities that reflect user preferences (backward). The search process is carried on according to the associations between spatial entities and reference locations. Finally, the user visualizes the results on the base map of the given city. The initial search stage stops here. In the case that the user is not completely satisfied with the results, the second search stage (Figure 5.1 (b)) provides the user additional opportunities to refine her/his selections. The spatial entities either selected by the user or returned by the system, will be discriminated as either spatial entities preferable to visit, or possible to visit, or of no interest. Then the system will re-compute the reference spatial entity (propagation), and re-recommends the spatial entities (back propagation) with the search algorithm selected in the first stage.

5.2.2 BNAM forward

The propagation rules of the BNAM are based on several semantic and spatial criteria that are described in the cases introduced below. First, propagation ensures selection of the reference entity that best fits the user's preferences according to some spatial and semantic criteria (BNAM forward). Secondly, back propagation to the layer of spatial entities ranks the spatial entities with respect to the selected reference location (BNAM backward). Let us first describe the encoding process. More formally, an input vector $x \in \{(0,1)\}^p$, is applied to the layer X and propagated from the layer X to the layer Y. This input vector represents the spatial entities of interest considered in the neural network (i.e. $x_i = 1$ if the spatial entity is considered in the computation, $x_i=0$ otherwise). Similarly an input vector $y \in \{(0,1)\}^q$ is applied to the layer Y where $y_j = 1$ if the reference location is considered in the computation, $y_j=0$ otherwise. The BNAM forward processes provide a diversity of mechanisms in the elicitation of user's preferences by allowing flexible combinations of user preference indexes, spatial and semantic criteria. These algorithms can be used to derive some clues for the further development of personalized search strategies. The possible combinations include:

- explicit choice of user's spatial entities of interest, and return of the most centrally located reference location, amongst the ones selected by the user, and according to some either spatial (cases A and B) or spatial and semantic metrics (case C);
- derivation of user's class preferences elicited from the spatial entities selected by the user, and return of the most centrally located reference location, amongst the ones that can best represent the user user's class preference pattern, according to some spatial and semantic metrics (case D);
- explicit definition of user's class preferences and return of the best reference location, amongst the ones that can best represent the user user's class preference pattern, according to some spatial and semantic metrics (case E).

The user's information input is kept minimal in all cases: selection of spatial entities of interest and of reference locations of interest. The criteria used for the propagation algorithm and the input values derived for each reference entity in the layer Y are described as the cases below (variables used several times in the formulae are introduced once).

Case A: Contextual distance

Based on the contextual distance, the algorithm returns the most centrally located reference location given a set of spatial entities. The input vector reflects the spatial entities that are selected by a user, those spatial entities corresponding to her/her spatial entity preferences. The algorithm below introduces the propagation part and the encoding of the algorithm:

$$f_{forwad}(x_i, y_j) = input(y_j) = y_j \sum_{i=1}^{p} x_i D(x_i, y_j) \text{ with} \tag{1}$$

$W_{i,j} = D(x_i, y_j)$

where x_i=1 if the spatial entity is selected in the input vector, x_i=0 otherwise; y_j =1 if the reference location is selected in the input vector, y_j =0 otherwise; $D(x_i, y_j)$ denotes the contextual distance between the spatial entity x_i and the reference location y_j; p denotes the number of spatial entities in the layer X, q of reference locations in the layer Y.

Case B: Contextual proximity

This algorithm models the strength of the association between the spatial entities and the reference locations using contextual proximities (one can remark that case B negatively correlates case A):

$$f_{forwad}(x_i, y_j) = input(y_j) = y_j \sum_{i=1}^{p} x_i P(x_i, y_j) \text{ with} \tag{2}$$

$W_{i,j} = P(x_i, y_j)$

where $P(x_i, y_j)$ is the contextual proximity between the spatial entity x_i and the reference location y_j.

Case C: Contextual proximity + degrees of membership

The algorithm below finds out the most centrally located reference location based on two criteria: contextual proximity, and overall interest of the spatial entities considered. This case takes into account both spatial (i.e., proximity) and semantic criteria (i.e., overall degree of membership to the classes given) to compute the strength of the association between a spatial entity and a reference location. Degrees of membership (x_i^1, x_i^2, …, x_i^k) weight the significance of a given spatial entity x_i with respect to the classes C_1, C_2, …, C_k. High values of class memberships increase the contribution of a spatial entity to the input values on the Y layer and its $input(x_i)$ in return. The propagation part of the algorithm is as follows:

$$f_{forwad}(x_i, y_j) = input(y_j) \ = \ y_j \sum_{i=1}^{p} x_i P(x_i, y_j) \sum_{h=1}^{k} x_i^h \quad \text{with} \tag{3}$$

$$\mathrm{W_{i,j}} = P(x_i, y_j) \sum_{h=1}^{k} x_i^h$$

where x_i^h stands for the degree of membership of x_i with respect to C_h; k denotes the number of semantic classes.

Case D: Contextual proximity + degrees of membership + class preferences

The propagation part of the algorithm adds another semantic criteria to the algorithm presented in case C: class preferences, a high-level semantic factor derived from the spatial entities selected. Formally, and for a given class C_h, its degree of preference p_h with respect to an input vector x $\in \{(0,1)\}^p$ is evaluated by

$$pref_i \ = \ \frac{\sum_{i=1}^{p} x_i \, x_i^h}{\sum_{i=1}^{p} x_i \sum_{j=1}^{j=k} x_i^j} \tag{4}$$

Those degrees of preferences form a class preference pattern *prefPattern(pref$_1$, pref$_2$, ..., pref$_n$)*, with respect to the classes C_1, C_2, …, C_h. At the difference of the previous cases, all spatial entities in X are considered as part of the input vector at the initialization of the neural network. The spatial entities selected by the user are taken into account only to derive her/his class preferences. Input values in the layer of reference locations are derived as follows:

$$f_{forwad}(x_i, y_j) = input(y_j) = \ y_j \sum_{i=1}^{p} P(x_i, y_j) \sum_{h=1}^{k} x_i^h \, pref_h \quad \text{with} \tag{5}$$

$$\mathrm{W_{i,j}} = P(x_i, y_j) \sum_{h=1}^{k} x_i^h \, pref_h$$

Case E: Contextual proximity + degrees of membership + user-defined class preferences

The approach is relatively close to the case D but with the difference that class preferences are user-defined. Input values in the layer of reference locations are calculated as follows:

$$f_{forwad}(x_i, y_j) = input(y_j) = \; y_j \sum_{i=1}^{p} P(x_i, y_j) \sum_{h=1}^{k} x_i^h \, pref_{uh} \quad \text{with} \tag{6}$$

$$W_{i,j} = P(x_i, y_j) \sum_{h=1}^{k} x_i^h \, pref_{uh}$$

Where $pref_{uh}$ denotes a user-defined class preference for class C_h, that is, an integer value given by the user at the interface level.

This case gives a high degree of flexibility to the user, it constitutes a form of unsupervised BNAM. The reference location y_j with the highest $input(y_j)$ value is the one that is the nearest to the spatial entities that are of interest, those being or not the ones given by the input vector.

5.2.3 BNAM backward

The backward algorithms are applied to all cases. The basic principles of the decoding part of the BNAM is to rank the spatial entities of the X layer with respect to the "winning" reference location selected in the layer Y. Output values are determined as follows.

Case A

$$y_j(t+1) = \begin{cases} +1 \text{ if } input(y_j) < input(y_{j'}) \text{ for all } j \neq j' \\ 0 \text{ otherwise} \end{cases} \tag{7}$$

for j =1, 2, …, q

Cases B, C, D and E

$$y_j(t+1) = \begin{cases} +1 \text{ if } input(y_j) > input(y_{j'}) \text{ for all } j \neq j' \\ 0 \text{ otherwise} \end{cases} \tag{8}$$

for j =1, 2, …, q

The patterns $y_j(t+1)$ produced on the Y layer are back propagated to the X layer thus giving the following input values on the X layer. The consistency of the algorithm is ensured as the decoding is made with the function used in the selective process in the X layer. The spatial entity with the best fit is the x_i where $x_i(t+2) \geq x_{i'}(t+2)$ at the exception of Case A where the spatial entity selected is the x_i where $x_i(t+2) \leq x_{i'}(t+2))$ for all $i \neq i'$. The other spatial entities are ranked according to their $input(x_i)$ values (ranked by increasing values for cases B, C, D and E, by decreasing values for case A). Input values for the X layer are given by

Case A: $$f_{backward}(x_i, y_j) = input(x_i) = x_i \sum_{j=1}^{q} y_j(t+1)\, D_{ij} \quad (9)$$

Case B: $$f_{backward}(x_i, y_j) = input(x_i) = x_i \sum_{j=1}^{q} y_j(t+1) P(x_i, y_j) \quad (10)$$

Case C: $$f_{backward}(x_i, y_j) = input(x_i) = x_i \sum_{j=1}^{a} y_j(t+1)\, P(x_i, y_j) \sum_{h=1}^{k} x_i^h \quad (11)$$

Case D: $$f_{backward}(x_i, y_j) = input(x_i) = \sum_{j=1}^{q} y_j(t+1)\, P(x_i, y_j) \sum_{h=1}^{k} x_i^h\, pref_h \quad (12)$$

Case E: $$f_{backward}(x_i, y_j) = input(x_i) = \sum_{j=1}^{q} y_j(t+1)\, P(x_i, y_j) \sum_{h=1}^{k} x_i^h\, pref_{uh} \quad (13)$$

with $x_i(t+2) = input(x_i)$
where $x_i=1$ if the reference location is selected in the input vector, $x_i = 0$ otherwise; $y_j(t+1)$ values are given by the $input(y_j)$ values above; q denotes the number of spatial entities in the layer Y.

The input functions above support a wide diversity of semantic and spatial criteria. This ensures flexibility in the elicitation process and maximization of opportunities despite the fact that user's data inputs are kept minimal.

5.2.4 Refine search processes

Amongst several search algorithms initially developed and implemented, the one retained (e.g., algorithm D) for the refining process first derives class preferences from the spatial entities of interest selected by the user. Secondly, this algorithm recalls the most centrally located reference location, amongst the ones selected by the user, according to some spatial and semantic metrics, user preference pattern. The information input by the user is kept minimal: selection of spatial entities of interest and reference locations. Without loss of generalization we consider the search algorithm D as the basis to develop the refining process. The motivation for algorithm D relies in the fact that it can 1) derive user's interests and preferences from user's behaviors while interacting with spatial entities of interest, e.g., selection, and 2) keep minimal user inputs: selections of spatial entities of interest and reference locations.

Due to the properties of the selected algorithm, some of the spatial entities of interest recalled may be not initially chosen by the user, while some of the spatial entities initially chosen can be even not selected by the search algorithm. When the user is not completely satisfied with the spatial entities selected by the system, the interface offers to her/him the option of refining her/his preliminary choices according to the system selections. A second pass of the personalized search algorithm is made by letting the user refining her/his choices, that is, re-selecting the spatial entities that interest her/him. The set of spatial entities manipulated at the interface level are as follows

A denotes the set of spatial entities initially selected by the user
B the set of spatial entities returned by the system
C = A $\cup$ B the set of spatial entities either selected by the user or returned by the system
D = A $\cap$ B the set of spatial entities selected by both the user and the system

The set of spatial entities D is likely to own the entities that present the higher degree of interest to the user, but some of the spatial entities of the set C might be also of interest. In order to refine user preferences and search strategies, after the first stage of the user preference elicitation process, the Web-based user interface monitors user's feedbacks on the results. The refinement of the user's choices is specified by evaluating to which degree she/he would like to see the spatial entities selected by the system. In order to compute these preferences, fuzzy scalars qualitatively model these different degrees of interest according to three levels: *Preferable*, *Possible*, and *No interest*. The spatial entities are grouped into three different sets whose intersection is null and the union equal to the set of entities presented initially to the user:

The set C_1 of spatial entities which are of preferable interest to the user
The set C_2 of spatial entities which are of possible interest to the user
The set C_3 of spatial entities that present no interest to the user

A second pass of the algorithm is made but with a quantitative modulation of $input(y_j)$ values. For a given spatial entity x_i, the refining coefficient xp_i is given by

$$xp_i = \begin{cases} 2.0 \text{ if the user would like to see the place } x_i, \\ 1.5 \text{ if the user would possibly like to see the place } x_i \\ 0 \text{ if the user would not like to see he place } x_i \end{cases}$$

The refining process is then given as follows

BNAM Forward

$$f_{forwad}(x_i, y_j) = input(y_j) = \ y_j \sum_{i=1}^{p} x_i xp_i P(x_i, y_j) \sum_{h=1}^{k} x_i^h \, pref_h \tag{14}$$

$$\text{with } \mathrm{W_{i,j}} = xp_i P(x_i, y_j) \sum_{h=1}^{k} x_i^h \, pref_h \, .$$

In the second step of the algorithm, the forward process from the X layer (spatial entities of interest) to the Y layer (reference locations) re-computes the best reference location according to the refinement of user preferences. The backward process from the layer Y to the layer X is given as follows based on the best reference location recalled,

BNAM Backward

$$f_{backward}(x_i, y_j) = input(x_i) \ = x_i \sum_{j=1}^{q} xp_i y_j (t+1)\, P(x_i, y_j) \sum_{h=1}^{k} x_i^h \, pref_h \tag{15}$$

The refining process is different from the first stage in different respects. First, the weights given to the spatial entities selected are changed according to the refinement process. Secondly, the search space is limited to the spatial entities of the set C (not all the spatial entities as in the first search stage), namely either selected by the user or the system in the first stage.

5.3 Global and local personalized search strategies

This section introduces two personalized search algorithms, that is, a “global” algorithm and a “local” algorithm, which optimises personalized searches through making further investigations into the semantic domain. These two algorithms make distinctions between positive and negative evidences in a user preference pattern, that is composed of preference indexes reflecting the fact that whether the user shows interest or not. The “global” strategy performs retrieval functions in the whole search space, while the “local”, in the user’s local space.

5.3.1 Evaluation of spatial and semantic interests

We introduce a function to evaluate the degree of spatial and semantic interest, exhibited by a spatial entity x_i and a reference location y_j, during a BNAM learning and search process. It is given as:

$$f(x_i, y_j) = f_{spa}(spatial) \times f_{sem}(semantic) \tag{16}$$

Both spatial and semantic contents are also expressed as functions. The spatial function $f_{spa}(spatial)$ is specified by the contextual proximity $P(x_i, y_j)$, previously defined, and bounded by the unit interval [0,1]. The higher $P(x_i, y_j)$ the closer x_i to y_j, the lower $P(x_i, y_j)$ the distant x_i to y_j, and *vice versa*. The spatial function is given as

$$f_{spa}(spatial) = p(x_i, y_j) \tag{17}$$

The semantic function $f_{sem}(semantic)$ is given as

$$f_{sem}(semantic) = \lambda \frac{1}{k} \sum_{h=1}^{k} g(x_i^h) like_h + (1-\lambda) \frac{1}{h} \sum_{h=1}^{l} g(x_i^h) disl_h \tag{18}$$

with $g(x) = \frac{1}{2}[sin(x - \frac{1}{2})\pi + 1]$

The domain of value of the semantic function is given by the unit interval [0, 1]. The semantic parameters x_i^h with $h \in \{1,..., h\}$ reflect the degrees of membership of a spatial entity x_i with respect to several semantic classes $C_1, C_2, ..., C_h$. The value k denotes the number of the top-k semantic parameters in the user preference pattern, which the user shows interests in. Similarly, l denotes the number of the low-l semantic parameters in the user preference pattern, which the user shows less interest (in the prototype developed so far the values k and l are tentatively fixed to 2). The constant λ is valued as a real number between 0 and 1, and allows adjusting the respective influence of the top-k and low-l semantic parameters (a suggested value for λ is fixed to 0.8 in order to emphasize positive evidences of user's interests and preferences in the search strategy). The function $g(x)$ describes a curve that increases slowly at the beginning and the ending parts, whereas sharply in the midway, to denote the assumption that people make a better distinction between "bad" and "good" than between "worse" and "worst" or between "better" and "best".

5.3.2 "Global" personalized search algorithm

A recurrent two-step search process is triggered by the BNAM-based framework to personalize access to spatial information on the Web (1) BNAM-forward: to determine the best reference location, (2) BNAM-backward: based on the best reference location to recall a set of top-n best spatial entities of interest. The recurrent bi-directional processes are respectively defined as follows

BNAM-Forward

$$f_{Gpropagation}(x_i, y_j) =$$

$$\sum_{j}^{q} y_j \sum_{i=1}^{p} x_i\, p(x_i, y_j) \times [\lambda \frac{1}{k} \sum_{h=1}^{k} g(x_i^h) like_h + (1-\lambda) \frac{1}{l} \sum_{h=1}^{l} g(x_i^h) disl_h] \tag{19}$$

where $x_i=1$ if the place is selected in the input vector, $x_i=0$ otherwise,
$y_j=1$ if the reference location is selected in the input vector, $y_j=0$ otherwise

BNAM Backward

$$f_{Gback-p}(x_i, y_j) =$$

$$\sum_{j=1}^{q} y_j(t+1) \sum_{i}^{p} x_i\, p(x_i, y_j) \times [\lambda \frac{1}{k} \sum_{h=1}^{k} g(x_i^h) like_h + (1-\lambda) \frac{1}{l} \sum_{h=1}^{l} g(x_i^h) disl_h] \tag{20}$$

Where $y_j(t+1)$ refers to the best reference location

The output of the BNAM forward process recalls the best reference location with respect to user preference pattern, while that of BNAM backward process results in a set of top-n spatial entities of interest. The best reference location and the set of top-n spatial entities of interests are linked with combination of spatial and semantic criteria, and user's interests and preferences.

The "global" personalized search algorithm derives user preference pattern from the entities the user selected, and then chooses the best reference location with traversal over all the entities of interest. This favors search and recommendation not limited to spatial entities initially selected by the user, but to the ones presenting an interest with respect to her/his class preference pattern while interacting with the whole spatial environment.

5.3.3 "Local" personalized search algorithm

The user commonly considers and interacts with a given spatial environment from his own and intuitive point of views. In other words, she/he focuses on spatial entities and reference locations of interest nearby or in her/his local space, not in the overall domain of the spatial entities available for selection. Accordingly, we also introduce a "local" personalized search algorithm, which makes an attempt to recall and determine the best reference location and a set of top-n best spatial entities that might interest the user in the "local" context of a specific user. However, the "local" personalized search algorithm does not consider the whole range of opportunities in the process of personalized search, as the search space is a local subset of spatial entities.

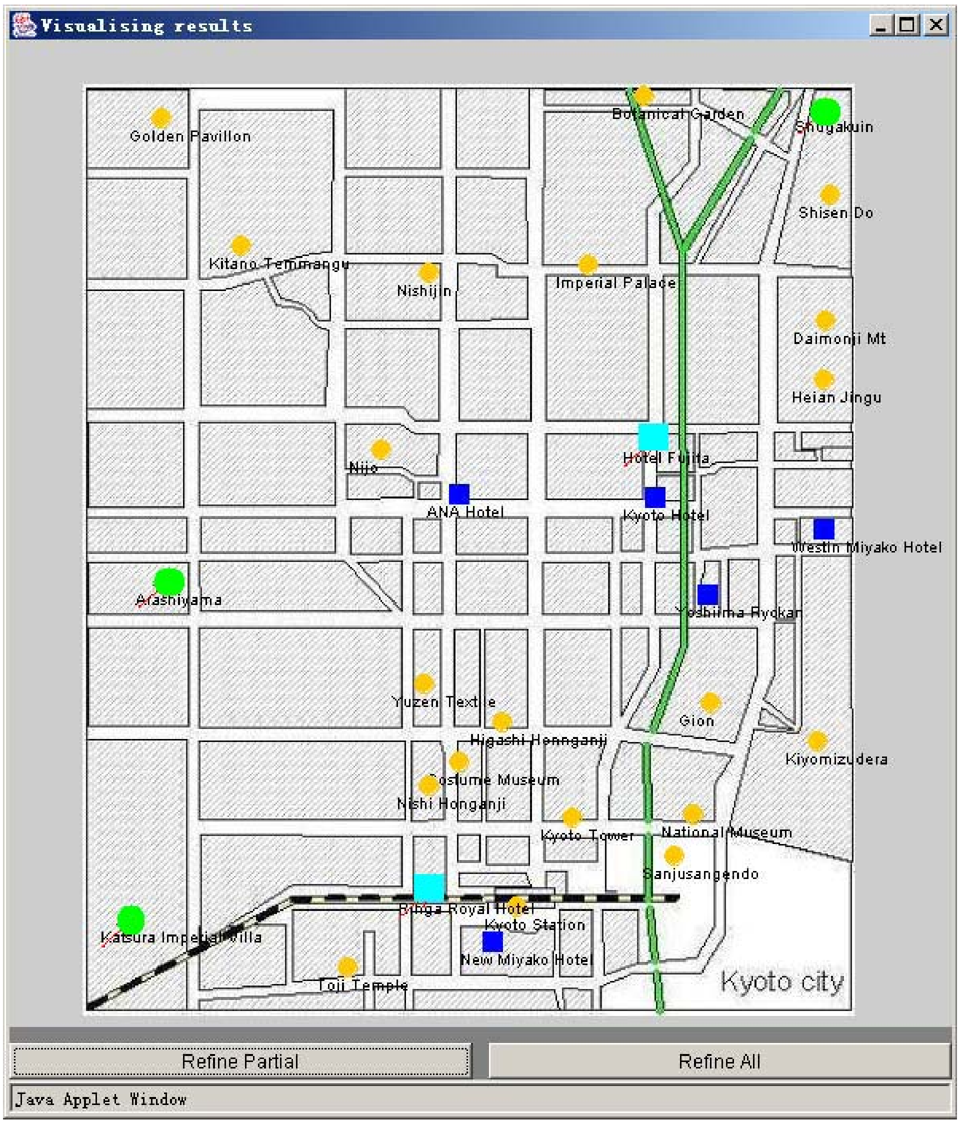


Figure 5.2 Example scenario

In the context of local search, we argue that the semantic value given by a spatial entity of interest surrounded by several similar entities should be reinforced. The interest of a reinforcement process is exemplified by the following case (Figure 5.2), where a user shows interests in a set of spatial entities e.g. *"Arashiyama", "Katsura Imperial Villa"* and *"Shugakuin"*, and chooses *"Rihga Royal Hotel"* and *"Hotel Fujita"* as reference locations. We consider that these three spatial entities have similar semantic parameters (e.g. relevant to semantic class Garden) and almost equivalent semantic value $f_{sem}(x)$ as introduced in section 5.3.1. In fact, *"Hotel Fujita"* shall be the better as the reference location rather than *"Rihga Royal Hotel"* in the case of a personalized search process emphasized on user preferences over gardens. The spatial entity *"Shugakuin"* should hold more weight (than the entities *"Arashiyama", "Katsura Imperial Villa"*) in the process of determining the best reference location, in that it is surrounded by several entities with similar semantic content. However, *"Rihga Royal Hotel"* is considered as the best reference location to take part in the process of recommending top-n spatial entities, when only entities user selected are considered {*"Arashiyama", "Katsura Imperial Villa"* and *"Shugakuin"*}.

To take into account this factor, we introduce a spatial auto-correlation component to infer a set of pseudo semantic parameters that are used to maximize the interest of a selected spatial entity when this spatial entity is surrounded by spatial entities of similar interest. For a given spatial entity of interest x_m, the pseudo semantic parameter is computed as

$$vx_m^h = \sum_{i=1}^{p} (1 - vx_m^h) x_i^h RP(x_m, x_i) \qquad (21)$$

where vx_m^h denotes a pseudo semantic parameter of a given spatial entity x_m with respect to a class C_h,; x_i^h the degree of membership of x_i with respect to C_h, p denotes the number of entities in the layer X, $RP(x_m,x_i)$ the inter-version of relative proximity between x_m and x_i.

The pseudo-semantic values are computed using an incremental process and bounded by the unit interval. Initial pseudo semantic values vx_m^h are set to zero. The reinforcement process is executed off-line on the Web server in order to avoid online heavy computation. In the procedures of extracting the pseudo-semantic values for each spatial entity, other entities are initialized as elements in an array rather. The order in the array is given by the proximity to the spatial entity under consideration. The algorithm of the reinforcement process is given as (Algorithm 1)

Algorithm 1.

X --- set of spatial entities of interest $X=\{x_1, x_2, \ldots, x_p\}$

vx_m^H --- set of pseudo semantic parameters of a given spatial entity x_m,

$1 \leqq m \leqq p$,

$$vx_m^H = \{ vx_m^1, vx_m^2, \ldots vx_m^h \}$$

1 Begin of pre-process

2 For each entity x_m in X
3 Sort routine over set of other entities in X, $X_{\sim m} = \{x_i \in X,\ x \neq x_m\}$ according to spatial proximity with x_m
4 For each pseudo semantic parameter vx_m^h in vx_m^H
5 Initialize as 0
6 Apply equation (21) to compute vx_m^h
7 End for
8 End for
9 End of pre-process

After the reinforcement process, the spatial entities have two sets of semantic parameters, that is, an original set ($x_m^1, x_m^2, \ldots x_m^h$), and a pseudo set ($vx_m^1, vx_m^2, \ldots vx_m^h$). The original semantic content of spatial entities serves the "global" personalized search algorithm, while the "local" personalized search algorithm needs to consider both the original set and the pseudo set.

Considering the additional component for the pseudo semantic content, the "local" personalized search algorithm is given as follows

BNAM Forward

$$f_{Lpropagation}(x_i, y_j) =$$

$$\sum_{j}^{q} y_j \sum_{i=1}^{p} x_i p(x_i, y_j) \times [\lambda \frac{1}{k} \sum_{h=1}^{k} \frac{1}{2}(g(x_i^h) + g(vx_i^h))k_h + (1-\lambda)\frac{1}{l}\sum_{h=1}^{l}\frac{1}{2}(g(x_i^h) + g(vx_i^h))l_h] \tag{22}$$

$$= \sum_{j}^{q} y_j \sum_{i=1}^{p} x_i p(x_i, y_j) \times [\lambda \frac{1}{2k} \sum_{h=1}^{k} (g(x_i^h) + g(vx_i^h))k_h + (1-\lambda)\frac{1}{2l}\sum_{h=1}^{l}(g(x_i^h) + g(vx_i^h))d_h$$

with k_h=$like_h$, l_h=$disl_h$

where x_i=1 if the place is selected in the input vector, x_i=0 otherwise,
y_j =1 if the reference location is selected in the input vector, y_j =0 otherwise

BNAM Backforward

$$f_{Lback-p}(x_i, y_j) =$$

$$\sum_{j=1}^{q} y_j(t+1) \sum_{i}^{p} x_i p(x_i, y_j) \times [\lambda \frac{1}{2k} \sum_{h=1}^{k} (g(x_i^h) + g(vx_i^h))k_h + (1-\lambda)\frac{1}{2l}\sum_{h=1}^{l}(g(x_i^h) + g(vx_i^h))l_h] \tag{23}$$

Where $y_j(t+1)$ refers to the best reference location

The "local" personalized search algorithm takes into account a sub-set of spatial entities the user shows interests in, instead of all spatial entities available. This facilitates its effectiveness in large applications with a large amount of heterogeneous spatial entities under consideration.

The personalized search algorithm recalls and recommends the best reference location and a set of associated spatial entities according to user preferences deduced from user's behaviors. In summary, the "global" algorithm searches the best reference location with consideration of all the spatial entities available in a "global search space", while the "local" only considers a subset of spatial entities that the user shows interests in "user-selected search space". The "local" search process is based on an offline pre-process to discover semantic information based on spatial relationships between spatial entities. The former is suitable to applications with small number of spatial entities, while the latter, to those manipulating large volume of entities, and where each entity possesses high semantic dimensionality.

5.4 Conclusion

A spectrum of personalized search strategies provides user preference elicitation and searching principles for spatial Web applications. The research also introduces two spatial Web personalized search strategies to customize spatial information services on the Web through the further investigation of semantic indexes and user preference pattern.

These search algorithms, on the one hand, calls for an evaluation process to measure their performances. On the other hand, current approaches for user preference elicitation and personalized search fall into the category of content-based personalization, in that less attention is paid to user identification, long-term user history profiles, and navigation patterns. Thereby, it is necessary to explore user navigation pattern over spatial entities using sequential Web data mining methods to personalize navigation in the Web space.

Chapter 6 A hybrid approach for spatial Web personalization

6.1 Introduction

Within our research framework, this chapter is intended to provide a personalization engine for spatial Web personalization. By contrast to personalized search strategies presented in chapter 5, the personalization engine is distinct with respect to several aspects, on the one hand. 1) It is more passive since it doesn't need any explicit user inputs. 2) It's a dynamical personalization process in that it takes into account user's current navigations to elicit user's interests and preferences, and to provide personalized information services to the user. User queries are generated through short-term user information (e.g. current user sessions). Personalization services are delivered through the matchmaking with long-term user information (e.g. historical transactions).

In the context of Web personalization, many research proposals have been oriented towards the modeling and prediction of user's behaviors on the Web, to reduce Web latency and improve Web prefetching (Padmanabhan and Mogul 1996), enhance search engine (Brin and Page 1998) and personalization engine (Mobasher *et al.* 2000), and ameliorate Website structure. In particular, Markov chains have proven to be suitable for predictive modeling of contiguous sequence of visits on the Web, that is, a potentially stochastic process (Pirolli and Pitkow 1999).

The research presented in this chapter integrates user's navigational trails, considering that they can improve user preference elicitation and personalization processes. We introduce a hybrid personalization approach and reinforcement process in order to facilitate navigations and interactions with spatial entities embedded on the Web. Such predictive mechanisms will facilitate Web recommendations and interface interactive opportunities offered to the users. Our approach combines semantic similarity, spatial proximity, and k-order Markov chains to predict the next spatial entity which is likely to be visited by a given user. Semantic similarity describes to which degree a spatial entity is close to another in the semantic domain, while spatial proximity is used to evaluate a contextual form of inverse distance with other spatial entities. A reinforcement process complements the approach by adapting both semantic content of spatial entities and transactions between them. Particularly through observing interactions between the user and the Web, a sequence of iterative negative/positive rewards evaluated on the basis of user's behaviors (e.g., selections) and feedbacks to personalized presentations.

The remainder of this chapter is organized as follows. Section 6.2 introduces basic notions of an entity-oriented user session and transaction. Section 6.3 proposes a hybrid Web personalization approach combining semantic similarity, spatial proximity and k-order Markov chains. Section 6.4 presents a reinforcement process to adapt the semantic content of spatial entities and navigational transactions. Section 6.5 gives some conclusive remarks.

6.2 Spatial entity-oriented user session and transaction

Users' navigational behaviors can be recorded using historical Web logs. The basic element of such a log information is a page-view, that is, a "Visual rendering of a Web page in a specific client environment at a specific point in time" as stated by the W3C Web Characterization Activity (http://www.w3.org/WCA/).

At the user level, navigation is triggered by a session. A user session is a sequence of user page-views made during a single visit by a given user. At a finer level of granularity, an episode or transaction denote a meaningful subset of semantically related user page-views within a user session. According to (Cooley *et al* 1999, Mobasher 2004) these notions lead to the following definitions:

Definition 1: Spatial entity-oriented user session
A *spatial entity-oriented user session s* is an n-dimensional vector $s = <e_1, e_2, \ldots e_n>$ that materializes a sequence of spatial entities "visited" by the user on the Web.

Definition 2: Spatial entity-oriented transaction
A *spatial entity-oriented transaction t* is a n-dimensional vector $t = <w(e_1, s_1), w(e_2, s_2), \ldots w(e_m, s_m)>$ that materializes a semantically related subset of a user session, and where each spatial entity e_i is associated with a weight s_m, that can be binary or values that denotes its semantic and spatial importance of a spatial entity on the Web.

We do not consider the case, which is far beyond the objective of our research, of spatial entities textually and implicitly embedded on the Web as this leads to explore and develop Web data mining and classification algorithms that automatically or semi-automatically extract and identify them. Instead we consider situations where spatial entities are explicitly embedded within a Web map interface or Web documents, and then more easily identified.

6.3 Personalization using Markov chains

6.3.1 Markov chains

Markov chains are used extensively to predict the next state of a system given a sequence of previous states (Sarukkai 2000). A Markov chain can be represented by a tuple with three parameters $<S, T, \lambda>$. $S = \{s_1, s_2, \ldots, s_n\}$ corresponds to the state space, namely the set of all possible states for the Markov chain; T is a transition probability matrix, where each entry t_{ij} represents the transition probability from a_j to state a_i; λ corresponds to the initial distribution of the states in S.

The state space of a Markov chain depends on the number of sequences of previous states available to predict the next state. A 0-order Markov chain is an unconditional base-rate probability of x_n denoted as $p(x_n) = Pr(X_n)$. In a 0-order Markov chain, the states are independent of each other. A first-order Markov chain considers one-step transition probabilities $p(x_2|x_1) = Pr(X_2 = x_2|X_1 = x_1)$ only, that is the probability of the next state given the immediately previous state. In a first-order Markov chain, each transition corresponds to a state. A k-order Markov chain considers the conditional probability by looking at the last k states to compute the predictions, $p(x_n|x_{n-1}, \ldots, x_{n-k}) = Pr(X_n = x_n|X_{n-1}, \ldots, X_{n-k})$. The state-space of a k-order Markov chain contains all possible sequences of k states.

The dimensionality of a Markov chain has a direct influence on the exhibited properties and performance of the prediction processes. Lower-order Markov chains cannot successfully predict next state because they don't look far enough into the past to correctly discriminate user's behavioral modes. High-order Markov chains result in high state space and low coverage, and sometimes even worse prediction accuracy due to the high number of sequential states (Deshpande and Karypis 2004). It has been observed, in an empirical analysis of data collected from xerox.com, (Pitkow and Pirolli 1999) found that using a 4^{th} order Markov chain is an optimal option upon an assumption that the benefit of making a correct hit equals the cost of marking an incorrect prediction.

6.3.2 Web personalization using Markov chains

Sarukkai (2000) introduced Markov chains for link prediction and path analysis to dynamically model URL access patterns, and to predict the next Web page accessed by the user. Padmanabhan and Mogul (1996) used n-top Markov models to improve prefetching strategies for Web caches. Pitkow and Pirolli (1999) explored the predictive capabilities of user paths and identified user access patterns on the Web. They introduced Longest Repeating Subsequence (LRS) models to predict world wide Web surfing. LRS models reduced predictive model size and complexity by nearly a third while retaining predictive accuracy. In order to improve prediction accuracy, and at the same time keep low state complexity, Deshpande and Karypis (2004) proposed a class of Markov models based on some prediction algorithms called *selective Markov models*, that are obtained by selectively eliminating a large fraction of the states of the All-K^{th}-Order Markov model. Empirical results show that the performance of selective Markov models is superior to that obtained by higher-order Markov models to predict Web accesses.

6.3.3 A hybrid Web personalization approach over spatial entities

Web personalization based on Markov chains predicts the next Web page a given user is most likely to visit by matching the user's current access sequence with historical Web access patterns. The entities extracted and identified from various Web documents constitute the state space. A state is defined as a Web entity, while a transition denotes a hyperlink from one entity to another. Markov chains use a sequence of Web page-views/entities the user accesses as inputs, with the goal of building Markov chains to predict the page-view/entity the user is most likely visit next. The predictive process is composed of a series of matching operations of user's current navigation trails with historical Web access sequences, to determine the next visit with transitional probability.

We propose a hybrid Web personalization approach integrating k-order Markov chains with a combination of semantic similarity and spatial proximity (denoted as *SemSpa similarity*). The intuition behind is as follows

> **Given *k* previously visited spatial entities $<x_{n-k},\dots x_{n-1}>$ on the Web, with consideration of conditional transition probability from $<x_{n-k},\dots x_{n-1}>$ to x_n, and *SemSpa* similarity between $<x_{n-k},\dots x_{n-1}>$ and x_n, the Web personalization engine predicts the n_{th} spatial entity which is likely to be visited by the user** (Figure 3)**.**

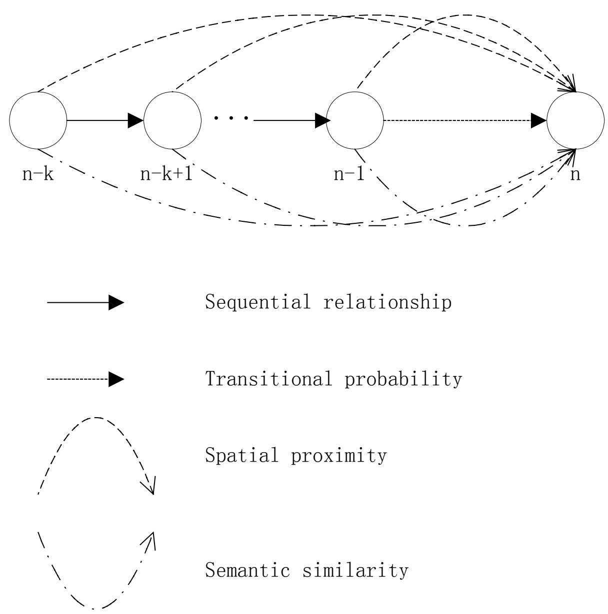


Figure 6.1 Relationships between spatial entities in k-order Markov chains

Figure 6.1 illustrates the principles of a k-order Markov chain related to a sequence of spatial entities of interest on the Web, namely sequential relationship in the k-sequence, semantic similarity, spatial proximity and transitional probability, that denote the relationships that influence the predictive k-sequence of Web entities.

As Web surfers may have two or more kinds of interests in mind, their interests may change from time to time when they are browsing on the Web. The latter is so-called "concept drift" (Webb *et al* 2001), an issue beyond the scope of traditional Web personalization applications. In order to address the "concept drift" issue we introduce a discount rate γ to adapt semantic similarity and spatial proximity values, on the basis of an assumption that following the user's navigational trails the "nearer" entities are more related than distant ones. The construction of the predictive Markov chain is decomposed into three steps as follows:

1) Determine *SemSpa* similarity between each entity in a k-sequence and a candidate entity connected to the current visit by a hyperlink.
2) Compute *kSemSpa* similarity between entities in a k-sequence and a candidate entity.
3) Compute *kSemSpaM*, a combination form of *kSemSpa* similarity and transitional probability in k-order Markov chains

1) SemSpa similarity between two spatial entities

The first step of our approach is to determine the *SemSpa* similarity. This corresponds to the combination value of Semantic similarity and spatial proximity between a spatial entity in a k-sequence x_i and a candidate spatial entity x_j, given as follows

$$SemSpa(x_i, x_j) = \sqrt{Sem(x_i, x_j) \times Spa(x_i, x_j)} \tag{1}$$

where $Sem(x_i, x_j)$ denotes the semantic similarity, and $Spa(x_i, x_j)$ the spatial proximity between x_i and x_j.

2) SemSpa similarity between spatial entities embedded a in k-sequence and a candidate spatial entity

The second step of the hybrid approach is to determine the *SemSpa* similarity between a sequence of k entities <x_{n-k},... x_{n-1}> and a candidate spatial entity x_n. In order to compute the *kSemSpa* similarity, we introduce a discount rate γ to refine the *SemSpa* similarity value between each spatial entity embedded in a sequence <x_{n-k},... x_{n-1}> and x_n. The *kSemSpa* similarity is given as follows

$$kSemSpa(< x_{n-k},...,x_{n-1} >, x_n) = \sum_{k=1} \gamma^{k-1} SemSpa(x_{n-k}, x_n) \tag{2}$$

where γ is a discount rate parameter, $0 \leqq \gamma \leqq 1$, that recursively decreases the SemSpa similarity between each spatial entities in a historical sequence $<x_{n-k}, \ldots x_{n-1}>$ and x_n with the historical length. The historical length denotes the number of steps from a spatial entity xi in a sequence $<x_{n-k}, \ldots x_{n-1}>$ to x_n.

3) Combination of k_SemSpa similarity and transitional probability

The third step of the hybrid approach is to combine the *kSemSpa* similarity with the transitional probability exhibited by the k-order Markov chain $<x_{n-k}, \ldots x_{n-1}$ in order to predict the n_{th} candidate spatial entity. The *kSemSpaM* value is given as

$$kSemSpaM(< x_{n-k},...,x_{n-1} >, x_n) = \sqrt{p(x_n | x_{n-1},...,x_{n-k}) \times kSemSpa(< x_{n-k},...,x_{n-1} >, x_n)} \tag{3}$$

where $p(x_n | x_{n-1},...,x_{n-k})$ is the transitional probability of the k-order Markov chain, statistically collected from the Web logs recording user's previous behaviors.

6.4 Reinforcement processes

6.4.1 Semantic content

The BNAM-based framework combines spatial and semantic criteria to identify user's intentions. The spatial component considered in the proposed framework reflects a *contextual proximity* factor, while the semantic dimension represented by the membership degrees of the spatial entities to some predefined classes of interest are *a priori* given at the system design level. In order to refine these membership degrees, a learning process adapts these membership degrees according to the user interactions and preference elicitation triggered by the interface. The learning process is applied to the first searching stage of user preference model, since the initial selection level is the most appropriate level to reflect intuitive user preferences.

The semantic parameters are reinforced either positively or negatively according to user preference patterns derived from the spatial entities selected. The semantic parameters of the user-selected spatial entities at the layer X are reinforced as follows:

$$x_i^h = \begin{cases} x_i^h + \dfrac{\alpha \times pref_h\left(1\text{-}x_i^h\right)}{\tau} & if\, pref_h = Max(f) \vee p^h = Max\left(f\text{-} \{Max(f)\}\right) \\ x_i^h - \dfrac{\beta \times pref_h\left(1\text{-}x_i^h\right)}{\tau} & if\, pref_h = Min(f) \vee p^h = Min\left(f\text{-} \{Min(f)\}\right) \end{cases} \tag{4}$$

Where f denotes a class preference pattern derived from equation 5.4, with *prefPattern(pref₁, pref₂, ..., prefₙ)*, $Max(S)$ is a set operator that returns the element of the set S with the maximum value, $Min(S)$ is a set operator that returns the element of the set S with the minimum value. α, β are the reinforcement factors with 0< α, β <1, both initialised once and relatively small, τ the number of times the elicitation process is triggered.

Positive reinforcements are given by the upper part of the equation (4), while negative reinforcements by its lower part given user class preferences derived by the reinforcing schema. In order to maintain an equilibrated strategy, positive rewards are given only to the spatial entities with the two higher input(x) values in the layer X, while the two smaller ones are reinforced with negative rewards. Let Ã denote the overall semantic values to a given spatial entity of interest (x_i), Ã = { x_i^1, x_i^2 , …, x_i^k }, Â provides an estimation of Ã. As the number τ of selection increases, then Â comes close to Ã, Â→ Ã. The learning process introduced is dynamic and ensures that the semantic parameters vary with the evolution of the overall users' attitudes to the spatial entities of interest, but converge after a reasonable number of times the elicitation process is triggered. This is due to the increase of the value of the coefficient τ over time.

6.4.2 Transitional probability

Interactions between the user and the personalization engine are iterative processes when she/he is surfing on the Web. These interactions consist of two kinds of process. First, the Web system provides personalized information services according to her/his interests; secondly the user gives relevance feedbacks through various behaviors. The user's relevance feedbacks reflect to which degree the user is satisfied with the personalized results, and thus are quite useful to improve the personalization engine. We take into account this component to adjust the transitional probability of the k-order Markov chains $p(x_n|x_{n-1},...,x_{n-k})$, through a reinforcement process. The reinforcement process is actually a learning process based on the observation of user's feedbacks to the predictive result provided (i.e., the n_{th} spatial entity). Possible forms of user's feedbacks to the n_{th} state is valued by two alternative Boolean values *satisfied* and *unsatisfied*. In the former situation, the user is likely to visit the recommended n_{th} spatial entity; while in the latter, the user will follow some other hyperlinks. The reinforcement process gives either positive or negative rewards. The reinforcement process is given as

$$p(x_n | x_{n-1},...,x_{n-k}) \leftarrow p(x_n | x_{n-1},...,x_{n-k}) \pm \eta \times r \tag{5}$$

where η is the learning rate, r is the reinforcement reward.

The value of η may be slightly smaller than 1 when learning begins, and then slowly decreased to 0 as learning progresses. A simple approach is to make use of τ, the number of visited times, to adjust these values dynamically. Then $\eta = \frac{\alpha}{\tau}$, α is the reinforcement factors with 0< α <1, is initialised once.

The reinforcement reward r is given as

(for positive reinforcement)

$$r = (1 - p(x_n | x_{n-1},...,x_{n-k})) \times (1 - kSemSpaM(< x_{n-k},...,x_{n-1} >, x_n)) \tag{6}$$

(for negative reinforcement)

$$r = (1 - p(x_n | x_{n-1},...,x_{n-k})) \times kSemSpaM(< x_{n-k},...,x_{n-1} >, x_n)$$

Then the reinforcement process is given as

(for positive reinforcement)

$$\Pr \leftarrow \Pr + \frac{\alpha}{\tau}(1 - \Pr) \times (1 - kSemSpaM) \tag{7}$$

(for negative reinforcement)

$$\Pr \leftarrow \Pr - \frac{\alpha}{\tau}(1 - \Pr) \times kSemSpaM)$$

with $\Pr = p(x_n | x_{n-1},...,x_{n-k})$, $kSemSpaM = kSemSpaM(< x_{n-k},...,x_{n-1} >, x_n)$

The transactional probability from a Web entity/page-view to another in sequential Web data mining is statistically calculated from Web logs recording user's historical navigation trails. The reinforcement process introduced provides a dynamic mechanism for conditional transition probability according to user's feedbacks to the personalized results.

6.5 Conclusion

This chapter introduces a hybrid Web personalization approach that combines Markov chains with spatial and semantic similarities, and with a reinforcement process in order to model and predict user navigational trails when interacting with spatial information materialized on the Web. The personalization process is based on a integration of two orthogonal dimensions that facilitate the approximation of user preferences, that is, semantic similarities and spatial proximities between spatial entities embedded in Web pages. Markov chains integrated with these relationships allow the system to recommend spatial entities that are of interest for the user.

Our approach introduces a discount rate parameter to compute the *SemSpa* similarity between a sequence of prior visits and a future visit (page view or Web entity), which effectively addresses the “concept drift” issue through attaching more weight to the “nearer” past visits in temporal dimension.

Chapter 7 Semantic user model

7.1 Introduction

This chapter introduces a semantic user model that describes and infers relevant user features to facilitate personalization services. As previously illustrated in figures 1.2 and 4.1, the user modelling component communicates user information with the personalization service framework. The derived user information is required by the personalization components and favours the customisation of spatial Web services to the user.

Personalization is a primary function to adapt various services to the heterogeneous user communities for semantic enriched applications, where user characteristics act as the main, determinant inputs. In particular, Fischer (2001) points out that

> "the challenge in an information-rich world is not only to make information available to people at any time, at any place, and in any form, but specifically to say the right thing at the right time in the right way."

This implies to explore efficient mechanisms for user modelling and user preference elicitation, and for personalization services in the context of information-rich applications. User modelling and preference elicitation are key prerequisite issues for the successful development of personalization systems. A user modelling component can be developed either as a customisable module that is independent from the applications, or as a part of a personalization system. The former refers to a user modelling shell system (Kobsa and Pohl 1995). System developers should select some components of a user modelling shell system, and fill it with relevant domain-dependent knowledge on the user community to a specific domain application. In the latter case, a user modelling module is specifically designed and integrated into a given application system. A user modelling module interacts with the application systems (or other system parts) through inter-process communications, e.g., "*tell*" and "*ask*" operations.

In most user modeling and preference elicitation applications, there are many cases where no sufficient information and assumptions about the user are available to support user preference elicitation and personalization strategies. This often results from the fact that the user may either be skeptical about a personalization system or be reluctant to be tracked by a user modeling component due to some understandable privacy issues. Where the user is asked to fill a registration form, she/he may do it so quickly that incomplete user profiles and inconsistency might occur. These make it a non-straightforward task to collect sufficient and correct user information.

This chapter introduces a user modelling approach based on description logics. The approach integrates static and dynamic user information to predict user features relevant to a given application domain. Description logics are effective to describe user's information and knowledge at the semantic level. Besides, they can efficiently handle inconsistency and incompletion issues, particularly to infer missing user information. Static user information refers to basic characteristics (e.g. demographics) explicitly presented by the user during a registration procedure. While dynamic user information is collected through observing user's behaviors.

We introduce a semantic user model illustrated and applied to the tourism domain, whose objective is to provide customized service and improve user's satisfaction. A tourism system usually should perform the following actions: (1) require each user to provide some information before delivering tourism services; (2) track what the user does, and (3) actively interact with her/him to dynamically adapt information services to her/his needs, interests and preferences. The user model uses different kinds of user information to identify and classify a given user to the right group in a domain-dependent stereotype hierarchy. It should be capable of predicting a set of user features relevant to personalization services. These operations, e.g. identification, classification and prediction, are implemented as inference rules using description logics. The logic-based user modelling framework acts as a support for a semantic Web personalization system to tailor information services to the user.

The remainder of this Chapter is organized as follows. Section 7.2 gives a brief discussion on logic-based semantic user models. Section 7.3 introduces description logics to design a user model knowledge base, and the description logic $\mathcal{SHIQ}(\mathcal{D}\mathcal{R})$. Section 7.4 proposes a user modelling approach and inference rules based on description logics. Section 7.5 illustrates the potential of the approach with an example in the tourism domain. Section 7.6 draws out the main research issues and explores further work.

7.2 Logic based semantic user models

Logic-based user models support user preference elicitation, semantic markup of user profiles, and added-value personalization strategies and services. Semantic user profiles can be universally distributed and employed in various applications with a multitude of different services and devices. They can also act as shared inputs for personalization services across multiple Websites or systems.

In a logic-based user model coupled to a specific application, inference rules can be employed to derive assumptions about the user, which are either executed prior to the runtime, or triggered by messages from the application. According to (Kobsa and Pohl

1995), four types of inference methods are characterized for user model acquisition as follows,

1) Inferences from user interviews: A user model could draw user's information needs, knowledge, interests and preferences from her/his answers through an interview component.
2) Inferences based on observed user information: User information on demographics, knowledge, interests and preferences observed and reported by the application become fairly explicit inputs to a user model. Then general and domain-specific inferences may be implemented to derive implicit assumptions about the user, these being required and used by the application.
3) Activation and retraction of user stereotypes: One of the pre-defined user stereotypes can be applied to a given user through certain activation and retraction conditions. Then, pre-defined assumptions of the stereotype are acquired to provide personalization services to the user.
4) Inference based on user's behaviors: In order to acquire assumptions about the user, frequently a number of either domain-independent or domain-specific heuristics are frequently used to describe inference rules. These heuristics-based inference rules identify assumptions when the user carries out specific actions while interacting with the application system at the user interface level.

The core of a user modeling system is a set of effective, flexible mechanisms for representing and reasoning about the assumptions about the user (Kobsa 2001). Assumptions about the user refer to what is taken for granted by a user modeling system. Different kinds of assumptions can be modeled and encoded, e.g., user's demographics, beliefs, goals, needs, interests and preferences. A user modeling system should offer a variety of representation and reasoning techniques that can meet the needs of application systems. There is a long tradition of employing logic-based mechanisms in user modeling system. The two main logic-based approaches to powerful user modeling representation and reasoning are the partition and the modal logics based approach (Pohl 1999). The partition approach allows for partial knowledge bases in a user model. These can offer several representation formalisms for different types of assumption contents in a user model. These assumption types are either domain-specific or domain-independent, which include goals, plans, capabilities, preferences, knowledge and beliefs. The latter approach covers user modeling requirements through the representation and reasoning capabilities of modal logics. The special operators for a variety of user assumptions are merely syntactic variants of the modal logic operators. Various modalities relevant to user modeling can be expressed using such a notation.

A logic based user model is a conceptual knowledge representation oriented to the description and inference about a user's assumptions, e.g. a User Model Knowledge Base (UMKB). A knowledge base consists of three components: a TBox (terminological axioms of an application domain), a ABox (e.g., assertions about named individuals), and a set of inference services. It provides interacting interfaces and a variety of information services to users and application systems. The interactions deal with the basic mechanisms of telling knowledge to a UMKB and retrieving knowledge from it (Figure 4.1). Ontology serves as a knowledge-base schema, that conceptually specifies a set of definitions and constraints with respect to concepts, roles, and attributes related to a given application domain.

An ontology can be considered as the kernel component of a semantic knowledge system, e.g., the semantic Web, which provides a common understanding of the basic semantic concepts used to annotate Web documents. Tim Berner-Lee visions the semantic Web as the extension of the current Web that can provide both human- and machine-understandable information (Berner-Lee *et al.* 2001). This can to a large degree facilitate the development of intelligent user agents that search and filter information, navigate in the Web space on behalf of the user. Ontology can also be used to reduce conceptual and terminological confusion, enhance information integration and high-quality information services. In the context of a UMKB, a user-oriented ontology is designed to conceptually describe concepts and relationships about user’s characteristics and classifications. This allows to identify and categorize a given user, and extract user’s personal information, e.g. interests and preferences. This information is contained in a user profile, and can be used as inputs for personalization systems.

Logic-based approaches, especially description logics, are dominant for the design and management of ontologies and knowledge bases, due to their strength of specifying primitive and defined concepts, and strong reasoning abilities (Calvanese *et al.* 1998a). Recently, ontology modeling issues are discussed from a conceptual modeling perspective (Borgida and Brachman 2002, Cullot *et al.* 2003). Conceptual modeling approaches such as Entity-Relationship models developed in the conceptual database domain have advantages such as better readability/understandability of an ontology content, and efficient management for large ontologies and knowledge bases. Logic-based approaches are capable of reasoning and inference to derive new knowledge from explicitly defined knowledge. Through analyzing some arguments about using database technologies for ontology design and analysis, Spaccapietra and his colleagues (2004) suggested a hybrid approach, namely combining database conceptual modeling and logic-based approaches. However, this remains an unexplored and promising issue.

7.3 Description logics for user modeling

Description Logics (DLs) are a family of knowledge representation formalisms that represent the knowledge of an application domain by defining the relevant concepts and specifying properties of objects and individuals in the domain (Baader *et al.* 2003a). DLs provide rigorous formalisms to describe user information, and derivation capabilities to infer missing characteristics and manage inconsistency in user models and profiles. Besides, DLs-based knowledge representation systems can be interpreted and processed by description logic reasoners such as Fact (Horrocks 1998, 1999), Racer (Haarslev and Möller 2001a, 2001b) and Pellet (Sirin and Parsia 2004).

Semantic user modelling and Web personalization on the semantic Web require techniques under the open-world assumption. One of the distinguishing features of description logics with conventional modelling languages is the non-finiteness of the domain and the open-world semantics. This implies that description logics based knowledge representation system can be applied to cases that one can not assume the information and knowledge contained in the system is complete. Description logics turn into relevant candidates for ontology languages due to their availability of a well-defined semantics and powerful reasoning tools (Baader *et al.* 2003b). Research achievements and insights from the description logics research community have a strong influence on the design of Web ontology languages such as RDF (RDF Core Working Group 2004) and OWL (Web Ontology Working Group 2004), particularly on the formation of semantics, the choices of language constructors, and the integration of datatypes and data values (Horrocks *et al.* 2003). Therefore a user model developed with description logics can be transformed with Web ontology languages to support personalized semantic Web services.

From a logical perspective, description logics become a member of the family of knowledge representation formalisms once they are equipped with a proper syntax and semantics, model and proof theory. Consequently, connections between description logics and other areas of logics particularly modal logics have received considerable research attention (Schild 1991, 1994, De Giacomo and Lenzerini 1994, Der Hoek and De Rijke 1995, De Rijke 1998). As first observed by Schild (1991), the description logics $\mathcal{ALC}$ can be viewed as a syntactic variant of multi-modal **K**. However, concrete domain constructor and n-ary relations that are the two main factors in our UMKB case, have no counterpart in modal logic, and there does not exist a translation from description logics with concrete domain extension. $\mathcal{ALC}(\mathcal{D})$ related concepts into formulas of the two-variable fragment of first order and modal logics or of the guarded fragment (Lutz 2003).

Recently, several research proposals apply description logics to represent and reason about user assumptions in user modeling systems. Description logic is proposed as a representation language for profile information to address matchmaking of demands and supplies of personal profiles in the business of recruitment cases (Cali *et al.* 2004). (Sinner *et al.* 2004) used description logics as a semantic language to describe services and user profiles, and determine whether a given profile is semantically compatible to particular services. Cali and his colleagues presented a richer set of description logic formalisms that encode various kinds of user information. Particularly, the former takes advantage of description logics with a concrete domain extension in order to manage “concrete” data in user profiles, e.g. the level of interest in a certain field (e.g. football). Semantic user profile representations are tailored for matchmaking operations and dating services, instead of user model construction and personalization services.

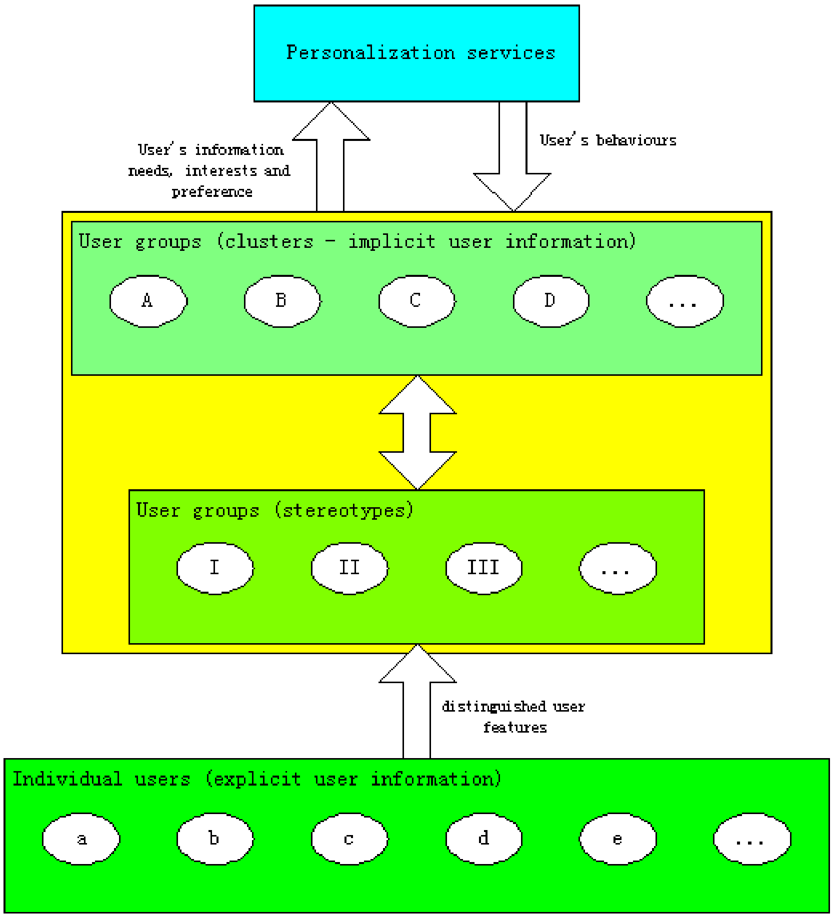


Figure 7.1 User model framework

This chapter proposes a user modeling framework based on description logics (Figure 7.1). The framework applies hierarchical user stereotypes identified according to several user features relevant to a specific application, e.g., the tourism domain. We combine two distinct types of user information to design a user model, that is, explicit, static user information from registration or interview processes, and implicit, dynamic user information derived from user's behaviors while interacting with the application system. Both explicit and implicit user features are used to identify a stereotype to which a given user belongs. While implicit user information is also used to dynamically update user stereotypes. The framework employs a description logics language $\mathcal{SHIQ(D_R)}$ to construct and reason an ontology-based UMKB.

The concept language $\mathcal{SHIQ}$ (Horrocks *et al.* 2000a, 2000b) is supported by modern description logic system such as FACT and RACER, which extends $\mathcal{ALC}$[5] with several expressive constructors: transitive roles, role hierarchies, inverse roles, and qualifying number restrictions. $\mathcal{SHIQ(D_R)}$ is an extension of $\mathcal{SHIQ}$ to encode concrete

[5] The DL $\mathcal{ALC}$ contains a variety of basic constructors: negation, conjunction, disjunction, existential restriction, universal restriction.

domains and n-ary relations, these being motivated by the requirement for modeling and reasoning about an ontology-based UMKB. In particular,

concrete datatypes such as number and string are used to represent user information contained in a stereotype (e.g. age, gender);
n-ary relations can describe roles that link an individual to more than just one individual or value (e.g. a user has interests in gardening);

The reasoning capabilities of description logics express inference rules for a user model. The user model can derive profiles for a given user who accesses to and interacts with the application system starting from her/his static (e.g. demographic data) and dynamic information (e.g. behaviors). A user profile contains the corresponding user's personal information, interests and preferences to assist searching, navigation in an information space.

The semantics of the DL $\mathcal{SHIQ}(\mathcal{DR})$ is specified as follows. An interpretation I is constituted by a pair (Δ_I, $.^I$), where Δ_I is a nonempty set, called the interpretation domain, and $.^I$ is the interpretation function. The interpretation function assigns to each concept name C a set of C^I of Δ_I, to each relation (role) name R of arity n a subset R^I of $(\Delta_I)^n$, and to each concrete feature g a partial function g^I to a partial function f from Δ_I to $\Delta_{\mathcal{D}}$.

Concrete domain
The necessities for the expressive power of description logics arise in almost all relevant application areas such as reasoning in conceptual database models and the construction of ontologies for the semantic Web. Amongst various extensions, concrete domain and n-ary relations are two aspects that play a important role to effectively address relatively complex information, e.g., particularly for database (knowledge base) modeling, management and queries. Following (Baader and Hanschke 1991), a concrete domain D consists of a set dom(D) (the domain of D) and a set pred(D) (the predicate names of D). Each predicate name P is associated with an arity n and an n-ary predicate over this set. For example, using the set of non-negative integer as concrete domain, we can describe a woman who is at least 20 years old as the concept:

$$\text{Human} \cap \text{hasGender.Female} \cap \text{hasAge.} \geq 20$$

Here ≥ 20 stands for the set of the nonnegative integer greater or equal to 20, $\geq$ the binary predicate.

N-ary relations
Calvanese and his colleagues introduced the description logic $\mathcal{DLR}$, which is capable

of handling a variety of data models with different forms of constraints, in terms of concepts and n-ary relations (Calvanese *et al.* 1997, 1998b, 1999). Let P and A denote a finite set of atomic relations having arity $n \geq 2$ and atomic concepts respectively, then arbitrary relations R and arbitrary concepts C are built according to the following syntax:

$$R ::= T_n \mid P \mid (\$i / n : C) \mid \neg R \mid R_1 \cap R_2$$
$$C ::= T_1 \mid A \mid \neg C \mid C_1 \cap C_2 \mid \exists[\$i]R \mid (\leq k[i]R)$$

where n denotes the arity of the relations P, R, R1, and R2, i a component of a relation ($1 \leq i \leq n$), k a nonnegative integer. As an example we consider to describe user's interests and preferences. User's interests and preferences are multi-dimensional, dynamic per nature. This implies to represent properties on user's interests and preferences over a set of activities or entities, e.g., *Christian* has high interests in gardening since 2005. In this example, the relation "hasInterest" is a 4-ary in that it links an individual (*Christian*) to other three individuals or values (gardening, high, and since 2005). We can conceptually describe "hasInterest" as:

//Terminology
hasInterest $\subseteq$ (\$1:Person)
$\cap$ (\$2: Entity)
$\cap$ (\$3: Interest_Degree)
$\cap$ (\$4: Period)

//Assertions
hasInterest(\$1:Christian \$2:Gardening \$3:High \$4: $\geq$ 2005)

Note that Interest_Degree is an enumerated class that consists of a list of individual names, namely Low, Middle, and High.

7.4 A user model knowledge base description logics

7.4.1 User stereotype construction and description

The stereotype approach (Rich 1979, 1989) to user modeling has been proven to be effective to address the cold-start problem in personalization applications that require a quick assessment of user's information. The "cold-start" problem results from the fact that there is often a lack of sufficient information either for a new user who accesses the application for the first time, or for a new item added recently. In (Kobsa 1993), Kobsa identifies three tasks for a developer of a user modeling component using stereotype approach,

(1) User group identification: The first step for building a stereotypical user model is to identify and differentiate user (sub)groups in a given user population for a given application.

(2) Identification of key characteristics: These user (sub)groups should possess certain homogeneous application-relevant characteristics, which allow one to identify the members of each user (sub)group.

(3) Representation in (hierarchically ordered) stereotypes: The collection of all represented characteristics of a user (sub)group is called a stereotype for this (sub)group. The application-relevant characteristics of these user (sub)groups should be formalized in an appropriate representation system, e.g., a hierarchical representation of stereotypes.

The stereotype user modeling approach is principally domain-dependent, in that the classification of the user community is determined by the requirements and characteristics of a given application domain. Let us consider the user community in the tourism domain. Tourism services aim at the delivery of adaptive information to help the user organize travel plans (Goy and Magro 2004), and physically interact with a given urban space. The classification of the tourism user community should take into account some user features, e.g., user's familiarity, interests and preferences, which are important criteria for evaluations of the utility of tourism services to a specific user. User's familiarity is an important factor to determine the level of information details, and interests and preferences for what kinds of information services providing to the user.

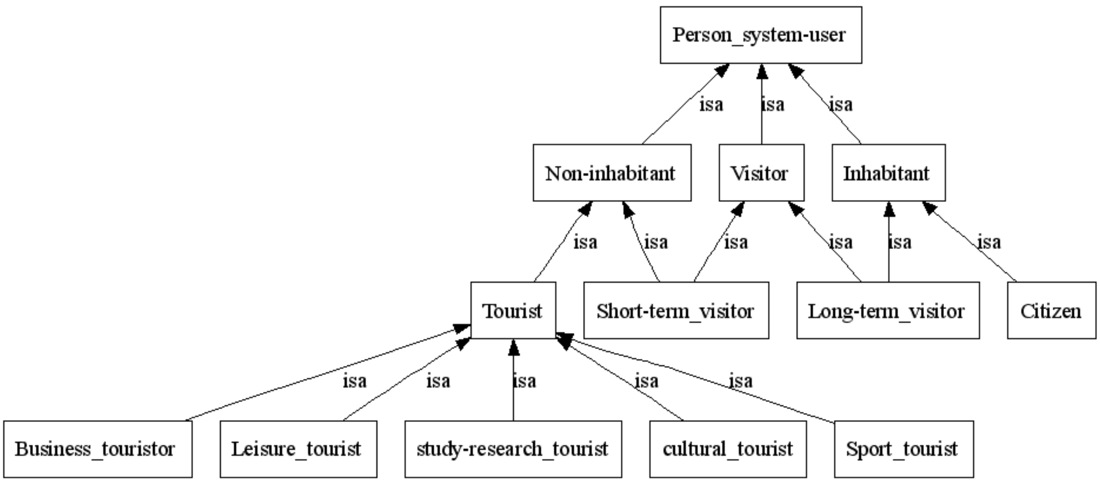


Figure 7.2 Hierarchical representation of stereotype

The tourism user community can be ordered in a hierarchy according to these user features, in which the contents of super stereotypes are inherited by the subordinate stereotypes. Figure 7.2 illustrates a stereotype hierarchy developed for user groups in the tourism domain. The root stereotype (*person_system-user*) contains the basic, universal user characteristics, e.g., hasCode, hasAddress, which are inherited by its subordinate stereotypes (Figure 7.3). Therefore, the deeper the user stereotypes are located in the hierarchy, the more detailed information they possess. Starting from these user characteristics either explicitly given by the user or observed by the system, some inference rules are applied to extract implicit user features that act as direct inputs for personalization applications. For example, a user who has permanent address in an urban space, may be of high familiarity to it. These user characteristics relate the *person* concept to other concepts in the tourism domain (Figure 7.4), e.g., a user has an address in a place or a city. In other words, our user model relates to a given domain ontology with some semantic relationships.

//Concept description
Person $\subseteq$ (Thing
//Basic information
$\cap$ ($\leq$ 1hasCode String)
$\cap$ (hasAddress Place $\cup$ City)
$\cap$ (hasAge Integer)
$\cap$ (hasBirthPlace City)
$\cap$ (hasNationality Country)
$\cap$ ($\leq$ 1hasGender Gender)
$\cap$ (hasHobbies Activity)
$\cap$ (hasProfession String)
$\cap$ ($\leq$ 1hasDisable B_State)
$\cap$ ([$1]hasAccessTimes $2: Spatial_Entity $3:Integer)
$\cap$ (hasActivity Place)
$\cap$ (isFamVisit B_State))

//Derived information
$\cap$ ([$1]hasInterest $2: Place $3:Interest_Degree)
$\cap$ (hasCogCapability Capability_Degree)
$\cap$ (hasPhyState Capability_Degree)
$\cap$ ([$1]hasFamilarity $2: City $3:Familarity_Degree)))

Figure 7.3 The *person_system-user* concept

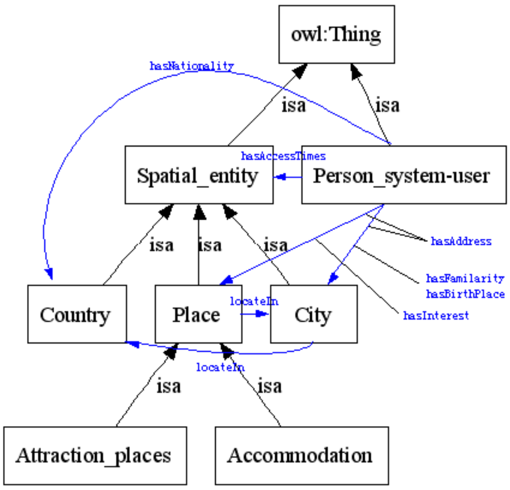


Figure 7.4 The tourism domain ontology (partial view)

Following the descending level of familiarity to a given urban space, people (system-users) can be divided to citizen, long-term_visitor, short-term_visitor and tourist. These relevant user stereotypes are represented with a set of identification constraints, through which a tourism system can classify an individual to a user stereotype. For example, a tourist is described as a non-inhabitant without temporal address in the residence center of the destination city (Figure 7.5)

//relationships

$\text{hasPAddress} \subseteq \text{hasAddress}$

$\text{hasTaddress} \subseteq \text{hasAddress}$

$\text{contain} = \neg\, \text{locateIn}$

//concepts

$\text{Destination} \subseteq \text{City}$

$\text{DCountry} \equiv \text{Country} \cap \exists\, \text{contain.Destination}$

$\text{DResidence\_center} \equiv \text{Residence} \cap \text{locateIn.Destination}$

$\text{Non\_inhabitant} \equiv \text{Person} \cap \exists\, \text{hasTAddress.Destination}$

$\qquad \cap \forall\, \text{hasPAddress.} \neg\, \text{Destination}$

$\text{Short\_visitor} \equiv \text{Non\_inhabitant} \cap \exists\, \text{hasTAddress. DResidence\_center}$

$\text{Tourist} \equiv \text{Non\_inhabitant} \cap \forall\, \text{hasTAddress.} \neg\, \text{DResidence\_center}$

Figure 7.5 The tourist concept

The identification constraints of the stereotype *Tourist* include user features hasCode inherited from its superordinate stereotypes (as described in the previous section). In the tourism domain, an application system should identify a foreign tourist from system users, since they may have little knowledge over the destination city. The concept of foreign tourist is given as:

$$\text{FTourist} \subseteq \text{Tourist} \cap \forall \text{ hasNationality}.\neg \text{DCountry}$$

In tourism applications, the stereotype approach still needs additional user's information to refine and update user classification since usually user's registration information might be either incomplete or inconsistent. Tourism users may not give explicit information e.g., about what kinds of attraction place she/he prefers or what kinds of tourist she/he is. This may result from either some privacy issues or the fact that the user has no clear clues on her/his own interests and preferences. These types of information may be derived from user's behaviors when interacting with a given urban space. The application system can observe and send a user modeling component some information about user's behaviors when interacting with a given urban space to a user modeling component. According to these actions, the user modeling system can produce the degree of user's interests and preferences, which is in turn used to classify the user to appropriate user stereotypes. Our user model for the tourism domain distinguishes five types of tourists: *business_tourist*, *leisure_tourist*, *study-research_tourist*, *culture_tourists*, *sport_tourist* (Figure 7.2).

- *Leisure_tourist*: This type of tourists usually has no explicit and detailed travel plan and destination, but only enjoys their time with some activities in some places, e.g., go to cinema in order to have a good time. They often go to cinema, theater, garden and beach.
- *Business_tourist*: Business tourists have explicit travel purposes in mind, e.g., trade, meeting, convention and exhibition. The common destinations for this kind of tourists include business_center and administrative_center.
- *Culture_tourist*: Culture tourists may prefer different cultures and cultural characteristics in the destination. They may often visit church, temple, museum, historic-site and culture area.
- *StudyResearch_tourist*: This type of tourists usually travels to visit several universities and research centers in the destination. Their activities may be arranged with reference to either a scientific event (e.g. conference) or a research invitation.
- ◆ *Sport_tourist*: This kind of tourists travels in order to either participate in or watch various sportive events. They may go to Gymnasiums, beach and other places where sportive events occur during a specific period.

These types of tourist can be distinguished according user features such as interests and preferences extracted from their behaviours while interacting with a given urban space. For instance, we identify and describe a business_tourist using the concept:

$$\exists \text{ hasActivity}.(\text{Business\_center} \cup \text{Administrative\_center})$$

$\Rightarrow \exists$ hasInterest$|_{\$2}$. Business
hasVisit $\Rightarrow$ hasInterest$|_{\$3}$.High
…
Business_tourist $\equiv$ Tourist $\cap$ [\$1] $\exists$ hasInterest.(\$2:Business, \$3:High)

On the other hand, tourist classification can also use information about user's profession and whether the user is in a professional travel or not. This can be inferred from some facts, e.g, the user is in a professional travel if she/her is a member of family visit. In a visit (during a period other than holiday seasons) to a given urban space, user's profession may be directly related to the activities and interacting styles that she/he will take, in case of being not a family tourist. For example, a visiting professor can be classified as a studyresearch_tourist since her/his activities might be arranged around a scientific event.

Person $\cap$ $\exists$ HasProfession(Professor) $\cap$ isFamVisit.No $\Rightarrow$ StudyResearch_tourist

In the tourism user hierarchy, each stereotype make a prediction on user features such as physical and cognitive function, familiarity to an urban space, and preference pattern, according to user's personal information and behaviors. A user preference pattern is constructed over several domain-dependent semantic classes, each of which is referred to a type of spatial entities in the tourism domain.

7.4.2 Implicit user information extraction

In our user model for tourism services, we mainly consider inference rules to derive user features highly relevant to the tourism domain: Familiarity to an urban space, physical and cognitive capability, and interests and preferences. In the application side these are used to build utility function of the tourism services to a given user.

Familiarity to an urban space
Familiarity is one of determinant features that can significant affect the interactions between the user and a given physical space. For instance, a given urban space to a user with little knowledge on it and another, who is familiar with it, is quite different. High familiarity to an urban space may facilitate user's decision-making, way-finding, travel routine and time, *vice versa*. In this user model, we identify and infer user's familiarity with two factors: role (e.g. tourist) and the access times. User's role may imply what type of activities a user takes in an urban space and user's familiarity. Each user group in a stereotype hierarchy refers to a specific role for users classified in it. We distinguish four roles: citizen, long-term_visitor, short-term_visitor and tourist, following the descending level of familiarity. Meanwhile, the access times can be used to decide user's familiarity degrees as: high, middle, and low. This tourism user model describes a person with low familiarity to a given urban space as

$$\begin{aligned}\text{PersonLF} \equiv \text{Person} \cap & \text{[\$1]hasFamilarity.(\$2: City \$3:Low)} \\ & \cup \text{[\$1]hasAccessTimes.(\$2: City \$3:} \leq 1) \\ & \cup \text{Tourist}\end{aligned}$$

Physical and cognitive capability

In comparison to many other kinds of electronic commerce, a tourism application is immaterial and difficult to model. Thereby the range of user's cognitive capabilities is of most importance: high demands for cognitive capability. Exceeding user's capabilities may give rise to service exclusion and difficulties for her/him to get the right information and make the right decisions. Tourism services are spatially related, and should act as an assistant for the user to physically interact with a given spatial environment. Accordingly, spatial proximity measures are an indispensable factor to tailor information to a specific user's personal and contextual situation. User's physical capability is in turn one of important factors that should be taken into account in spatial proximity and similarity measures.

According to Piaget's theoretical framework genetic epistemology, there are four cognitive structures (i.e., development stages): sensorimotor, preoperations, concrete operations, and formal operations (Piaget and Inhelder 1968). In previous stages children cannot reason abstractly or test hypotheses systematically. They begin to reason abstractly in the fourth stage (12-15 years). Statistical surveys on physical and cognitive capability shows that physical and cognitive function decrease with age among people aged over 50, and the trends of cognitive and physical decline are in prevalence over the age of 75 (Steel *et al.* 2004). They also find that cognitive capabilities are likely to be related to measures of physical functions, particularly the ability to perform instrumental activities of daily living. For instance, using a map to figure out how to act in an unknown environment, requires both physical and cognitive capabilities. Therefore, user's physical and cognitive capability with age can be considered as a determinant factor. Physical and cognitive capabilities are classified as three degrees: low (less than 12 years, over 75 years), middle (from 12 to 18 years, from 50 to 75 years), high (from 18 to 50 years). Moreover, users with a variety of physical impairment conditions such as athetoid, ataxic, require more efforts and attentions for tourism service personalization. There are very important performance and acceptance differences between them and able-bodied user, when they interact either with computers or with physical environments. These differences in the interaction cycles should be qualified at the level of user model design (Keates *et al.* 2002), since disabled users may demand additional services especially while delivering personalized services to a user group, e.g. a family (Ardissono *et al.* 2001). A disabled user's physical state is often at the low level. Then we describe a person who has low physical capabilities using the concept:

$$\text{PersonLP} \equiv \text{Person} \cap \text{hasAge.} \leq 12 \cup \text{hasAge.} > 75 \cup \text{hasPhyDisable.true}$$

User’s interests and preferences

User’s interests and preferences can be implicitly identified either from user’s personal information, or from user’s behaviors, as the way that an adept guide does in the tourist domain. First of all, we can infer user’s interests and preferences from her/his personal information, e.g., hobbies. Profession and hobby, to some degree, can be employed to infer user’s interests and preferences. A person who likes football may have strong interests in sportive issues. A professor in history may prefer to visit historic_sites and museums.

$$
\begin{aligned}
&\text{Activity} \subseteq \text{Thing} \cap \text{occurIn.Spatial\_entity} \\
&\text{relateTo} = \neg\, \text{occurIn} \\
&\exists\, \text{HasHobby(Activity)} \Rightarrow \\
&\qquad \exists\, \text{hasInterest(\$2:Place} \cap \text{relateTo.Activity \$3:High)}
\end{aligned}
$$

Secondly, user’s interests and preferences evolve over time. This implies to explore user’s dynamic interactions with an information space. It also reflects the observation that a given user shows some interests to a type of spatial entities if she/he performs some actions on an entity that falls into this type, either physically or on the Web. The possible actions include visit, pass by, browse, search for further information, include in travel routine, etc. These actions lead to different degree of user’s interests and preferences according to a type of spatial entities. For example, visit, search for further information, select and include in travel routine reflect high interests and preferences. The family of tourist stereotype classifies the user according to her/his interests and preferences in the topics of the tourism domain. For instance, a person with high interests and preferences over the domain *Culture* can be described using the concept:

//Interests and preferences

$$\text{hasActivity. Place} \Rightarrow \text{hasInterest}|_{\$2}\text{. Place}$$

$$
\begin{aligned}
&\text{hasVisit} \subseteq \text{has Activity} \\
&\text{hasSearchFurther} \subseteq \text{has Activity} \\
&\text{hasBookmark} \subseteq \text{has Activity} \\
&\text{hasBrowse} \subseteq \text{has Activity} \\
&\text{hasSelect} \subseteq \text{has Activity} \\
&\text{hasPassby} \subseteq \text{hasActivity}
\end{aligned}
$$

$$
\begin{aligned}
&\exists\, \text{hasActivity. (Business\_center} \cup \text{Administrative\_center)} \\
&\Rightarrow \exists\, \text{hasInterest}|_{\$2}\text{.Business} \\
&\exists\, \text{hasActivity. (Gymnasium} \cup \text{Beach)} \Rightarrow \exists\, \text{hasInterest}|_{\$2}\text{.Sport} \\
&\exists\, \text{hasActivity. (Historic\_site} \cup \text{Culture\_area} \cup \text{Museum} \cup \text{Church)} \\
&\Rightarrow \exists\, \text{hasInterest}|_{\$2}\text{.Culture} \\
&\exists\, \text{hasActivity. (Research\_center)} \Rightarrow \exists\, \text{hasInterest}|_{\$2}\text{.StudyResearch} \\
&\exists\, \text{hasActivity. (Garden} \cup \text{Theater} \cup \text{Cinema)} \Rightarrow \exists\, \text{hasInterest}|_{\$2}\text{.Leisure}
\end{aligned}
$$

$\exists$ hasVisit $\Rightarrow$ $\exists$ hasInterest$|_{\$3}$.High
$\exists$ hasSearchFurther $\Rightarrow$ $\exists$ hasInterest$|_{\$3}$.High
$\exists$ hasBookmark $\Rightarrow$ $\exists$ hasInterest$|_{\$3}$.High

$\exists$ hasBrowse $\Rightarrow$ $\exists$ hasInterest$|_{\$3}$.Middle
$\exists$ hasSelect $\Rightarrow$ $\exists$ hasInterest$|_{\$3}$.Middle

$\exists$ hasPassby $\Rightarrow$ $\exists$ hasInterest$|_{\$3}$.Low

Thirdly, it's reasonable and natural to recommend particular places in a given urban space, e.g., the Effeil tower in Paris, especially if the user has visited them. Sometimes these places might act as one of the main impetus to the interactions between the user and a city. These inference rules can be given as:

Spe_place $\subseteq$ Place
$\exists$ visit.City $\Rightarrow$ $\exists$ hasInterest.(\$2:Spe_place $\cap$ locateIn.City \$3:High)

7.5 DL user model inferences: an illustrative case

A DLs based knowledge base can perform a variety of inference rules, beyond storing terminology and assertion axioms (Baader and Nutt 2003). Inference problems in description logics are used to extract knowledge that is contained only implicitly in a given ontology or knowledge base. These rules can assist knowledge engineers to construct complex knowledge bases over a given domain. The inference capability in a logic-based user model can be used to infer new knowledge either prior to the runtime or when required by application systems. The basic inferences operating on concept description and a terminological axiom include subsumptions and satisfiability.

- Subsumption checks whether a concept description is more general than another, which is employed to organize the concepts in a taxonomy according to their generality.
- Satisfiability concerns if a concept description is satisfiable with respect to a terminological axiom.

Concerning a knowledge base with a TBox and an ABox, the inferences are consistency and instance checking.

- Consistency is to check if an ABox is consistent with respect to a TBox, e.g. there is an interpretation that is a common model of the two.
- Instance checking verifies that if an individual in an ABox is an instance a concept description with respect to a TBox.

In order to achieve an efficient implementation, it's necessary to consider more complex inferences that can be reduced to multiple invocations of the more basic inference problems mentioned above: retrieval and realization.

- Retrieval is, given an Abox A, a TBox and a concept C, to find all individuals a such as that $A|=_T C(a)$.
- Realization is, given an Abox A, a TBox, and an individual a, to find the most specific concept C from the set such that $A|=_T C(a)$.

All the relevant inference problems can be reduced to the consistency problem for ABox, provided that the DLs at hand allow for conjunction and negation. The Tableau Algorithms have turned out to effectively handle this issue, and to obtain sound and complete satisfiability algorithms for a great variety of DLs (Baader and Sattler 2001).

Subsumption and concept satisfiability inferences are used to check if there exist some inconsistencies in the terminological axioms of the user modeling knowledge base, and to add a new user concept to a stereotype hierarchy. As an example we consider a person who visit a given urban space, e.g. Christian visits Beijing for the first time (Figure 7.6). Besides her registration information, we get the fact that she also visited the Great Wall after arrival.

Name: Christian
hasCode: BJ0001
Gender: Female
Age: 30~40
Address: Roma
Nationality: French
Temporary address: Beijing Hotel
Family visit: No
Profession: Professor
haHobby: Gardening…
hasAccessTimes: the first time
hasVisit: GreetWall
…

Figure 7.6 Christian's personal information

Consistency inference will be used to check if Christian's description is consistent with the terminological axioms, and instance checking to identify her to the right stereotypes. The Great Wall is figured out to be a historic site (realization). The user with similar features can be sorted out through retrieval inference. The logic-based user modeling system encodes and analyzes Christian's personal and behavioral information, then infers some implicit information about her, e.g. familiarity, physical and cognitive capability, and interests and preferences.

//Asserations
//Basic user information
hasCode(Christian, BJ0001)
hasGender(Christian, Female)
hasAge(Christian, $\geq 30 \cap \leq 40$)
hasPAddress(Christian, Roma)
hasTAddress(Christian, BeijingHotel)
hasNationality(Christian, French)
hasProfession(Christian, Professor)
hasHobby(Christian, Gardening)
hasAccessTimes(Christian, Beijing, ≤ 1)
hasVisit(Christian, GreatWall)
…

//Destination information
Destination(Beijing)
Spe_place(GreatWall)
Spe_place(ForbiddenCity)
…

//Derived user information
hasFamilarity($1:Christian, $2:Beijing, $3:Low),
hasCogCapability(Christian, High)
hasPhyCapability(Christian, High)
FTourist(Christian)
Culture_tourist(Christian)
StudyResearch_tourist(Christian)

Figure 7.7 User profile for Christian

These inference results show that various user features about Christian, and who is identified as a culture tourist, studyresearch tourist (Figure 7.7). The contents of the Culture_tourist and StudyResearch_tourist stereotypes with respect to user's interests and preferences are used to predict some information about her through a conjunction operator. The multi-dimensional user information is merged through a disjunction operator to form a complete user profile (Carmagnola *et al.* 2005). The application system employs this information to tailor information services delivered to her in order to facilitate her travel in Beijing.

$$\text{Culture\_tourist} \cup \text{StudyResearch\_tourist} \subseteq \text{Tourist} \cap$$
$$\cap\ [\$1]\ \exists\ \text{hasInterest.}(\$2\text{:Business\_center } \$3\text{:Low})$$
$$\cap\ [\$1]\ \exists\ \text{hasInterest.}(\$2\text{:Culture } \$3\text{:High})$$

$\cap$ [$1] $\exists$ hasInterest.($2:Leisure $3:Middle)
$\cap$ [$1] $\exists$ hasInterest.($2:Sport $3:Middle)
$\cap$ [$1] $\exists$ hasInterest.($2:Research_center $3:High)
$\cap$ [$1] $\exists$ hasInterest.($2:Transportation $3:Middle)
$\cap$ [$1] $\exists$ hasInterest.($2:Administrative_center $3:Low)

7.6 Discussion

This chapter proposes a user modelling approach based on description logics, for spatial Web personalization applications, as applied to the tourism domain. We believe that user's multi-dimensional characteristics are intimately and semantically inter-related. A logic-based user model should principally infer as much as user information from limited inputs from a given application system. We employ an extension of $\mathcal{SH}$- description logics, $\mathcal{SHIQ}(\mathcal{DR})$ to represent and reason about stereotypical user modelling. Adding numerical knowledge representation to the $\mathcal{SHIQ}$ enhances the expressive power without increasing the worst-case complexity of reasoning (Lutz 2002). This work is based on a reasonable assumption of reasoning with the $\mathcal{SHIQ}(\mathcal{DR})$ to be decidable in EXPTIME. There still remains open issues for further work, e.g., complexity of reasoning with the $\mathcal{SHIQ}(\mathcal{DR})$.

Our semantic use model has shown to be capable of describing various types of user information and inferring user features to facilitate semantically enriched services. This chapter makes some preliminary efforts on the elicitation of implicit assumptions from user characteristics explicitly provided by the user in registration procedure. There are still some inference rules among user's characteristics that remain unexplored. For instance, some useful information is concealed in culture- or region-specific user features.

Chapter 8 Implementation

8.1 Introduction

This chapter presents the implementation of our research framework and relevant techniques identified for spatial Web personalization. Without loss of generality, we consider an application scenario about a given Web urban space. We identify a spatial Web application scenario in the tourism domain, which is intended to provide personalized information about a variety of spatial entities in order to assist the user to travel in an Web urban space. In the travel and tourism domain many systems have been implemented to support personalized services on the Web. A wide range of heterogeneous information is available, and the complexity of product descriptions in the field of tourism is growing (Werthner and Klein 1999). Personalization techniques are intuitive and valuable extensions to, and meanwhile a common means for tourism information systems based on observations in the real world (Berka and Plößnig 2004).

We developed a Web-based prototype that illustrates the principles of our approaches and models. The prototype system can be used to assist the user to make a travel tourism plan to an urban space. It is capable of several functions: 1) elicit user's interests and preferences through observing user's behaviours; 2) personalized search for spatial entities; 3) dynamically recommend spatial entities with consideration of historical transactions and current user sessions. The user interface integrated with our prototype represents spatial entities using image schemata, and deliver personalized results to the user.

Our prototype provides an experimental validation of the BNAM-based user preference elicitation and personalized search algorithms. It is applied as a case study to the historical city of Kyoto, an urban context that possesses a high diversity of places. The spatial entities of interest are modelled as places that might present an interest to the user that wants to visit the city of Kyoto, reference locations as hotels from where the user will be able to act in the city. The set of sightseeing places are defined to be relevant to several semantic classes (Museum, Temple, Garden, Urban) with membership degrees in [0,1], and the hotels exclusively relevant to a semantic class Hotel.

The Web interface represents two sets of spatial entities, sightseeing places and hotels, and is enriched with image schemata for sightseeing places. The Web map interface is implemented within an interactive Web interface. Each spatial entity embedded on the

Web map interface is presented by a symbol or an image, and associated to additional textual information that allows the user to actively interact with the Web interface. Personalization results are presented to the user in various formats and an interactive map with hyperlinks to Web resources of interest.

The remainder of this chapter is organized as follows. Section 8.2 presents an experimental evaluation of user-centric conceptual maps, spatial proximity and similarity measures. Section 8.3 illustrates personalized search strategies and a hybrid personalization engine. Finally section 8.4 concludes this chapter.

8.2 User-centric spatial proximity and similarity measure experiments

This section introduces the principles of the implementation of our user-centric conceptual map and similarity model, which consists two phases. The first phase concerns the representation of user-centric conceptual map of spatial entities in a given urban space, according to a specific user's interests and preferences. The second phase evaluates spatial proximity and similarity measures applied in the context of user-centric conceptual map. The spatial proximity and similarity measures are used to measure "closeness" between spatial entities either in a same category, at a same level of abstraction (e.g. sightseeing places), or from different categories (e.g. a sightseeing place and a hotel). From a semantic perspective, the set of sightseeing places can be viewed to belong to a same category and at same level of abstraction (e.g. sightseeing entities in a given urban ontology). Each category refers to a set, collection, group, or configuration that contains members (spatial entities) regarded as having certain attributes or traits in common. In evaluation of spatial proximity and similarity measure, we focus on two collections from an urban ontology, namely Sightseeing places and Hotels. The semantic interrelationships between semantic categories Sightseeing places and Hotels can be computed from the terminological ontology illustrated in Figure 4.6.

8.2.1 User-centric conceptual map

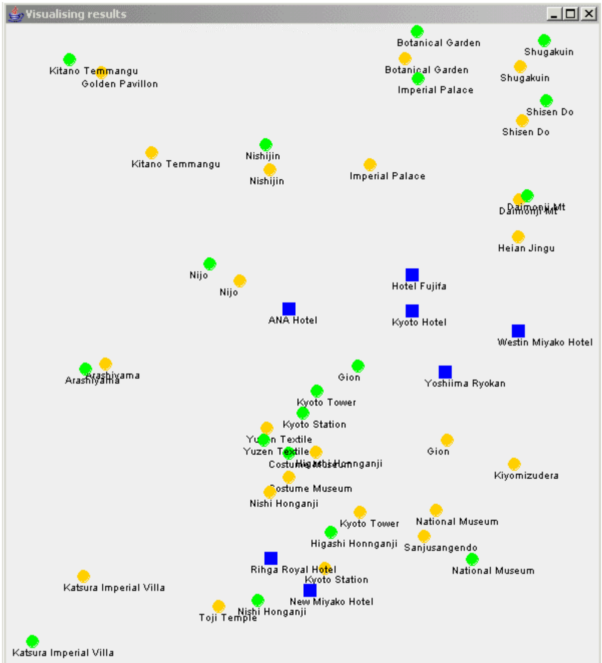
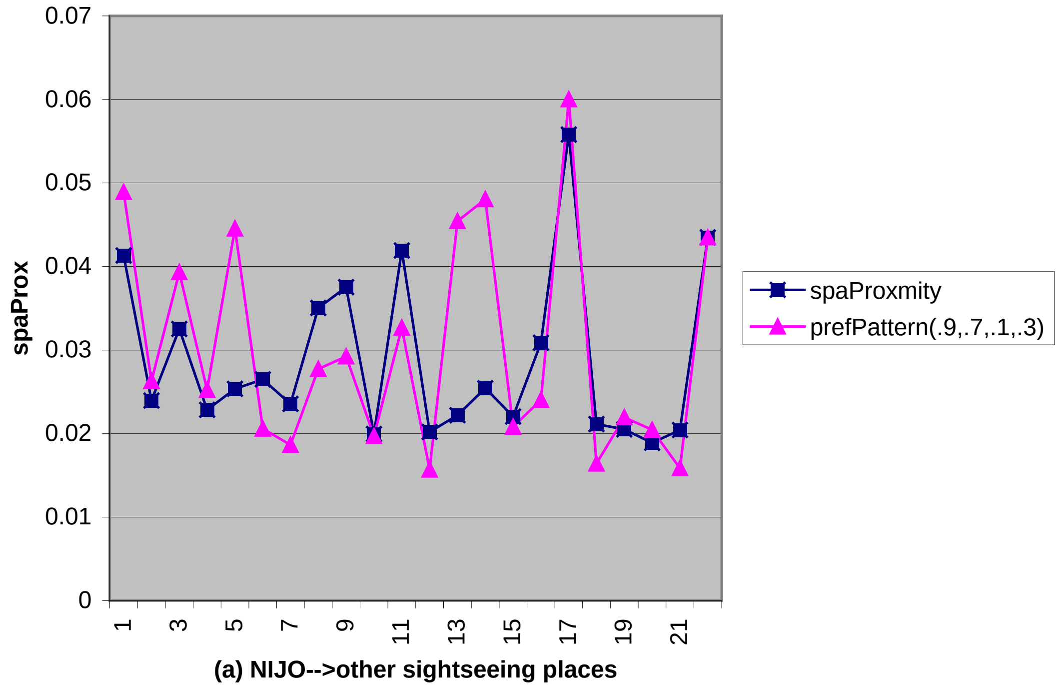


Figure 8.1 A user-centric conceptual map

User's interests and preferences are described with user preference patterns containing a set of preference indexes. Figure 8.1 illustrates an user-centric conceptual map based on a given user's interests and preferences, with Ana Hotel as her/his location and user preference pattern *prefPattern(0.9,0.7,0.1,0.3)* over the sightseeing places in Kyoto city. The grayish circles denote positions of these sightseeing places on the original map, and the dark gray circles, on the user-centric conceptual map. The sightseeing places "moves" to ANA Hotel by certain distance (either positively or negatively) dependent on her/his user preference pattern and their corresponding membership degrees. To some degree a conceptual map represents a given user's perception on the distribution of these spatial entities of interest, which is distorted version of the original map. For example, Gion appears nearer to the user since it fits her/his interests, and Nijo, on the contrary. The changes of the distribution of these spatial entities lead to different contextual knowledge influencing spatial proximity measures.

8.2.2 Spatial proximity and semantic similarity measures



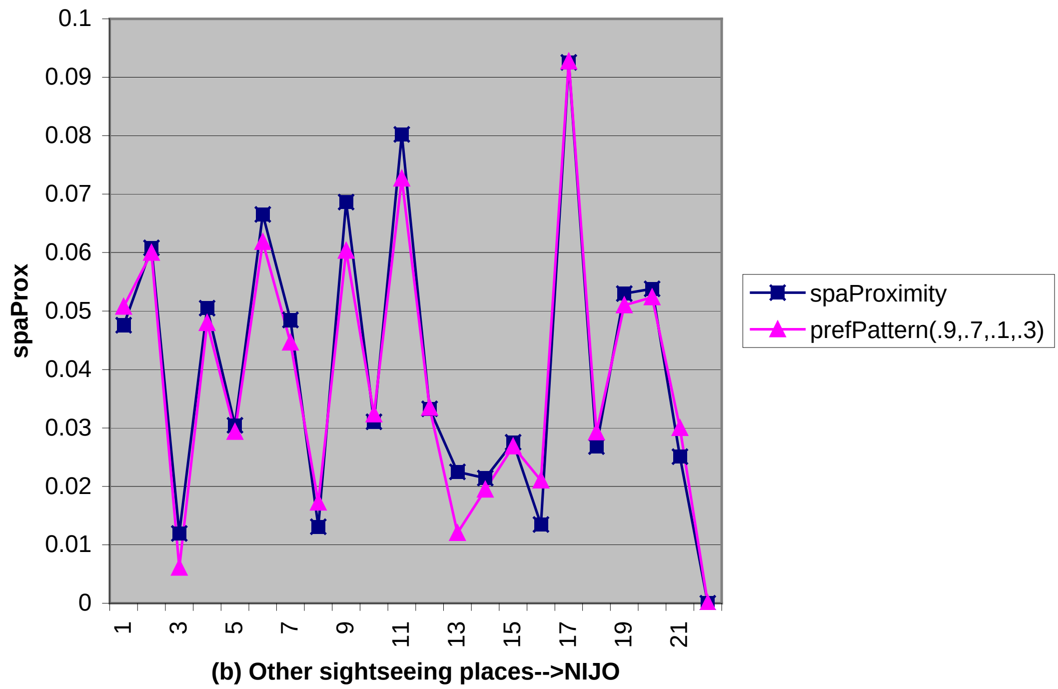


Figure 8.2 Spatial proximities between Nijo and the other sightseeing places

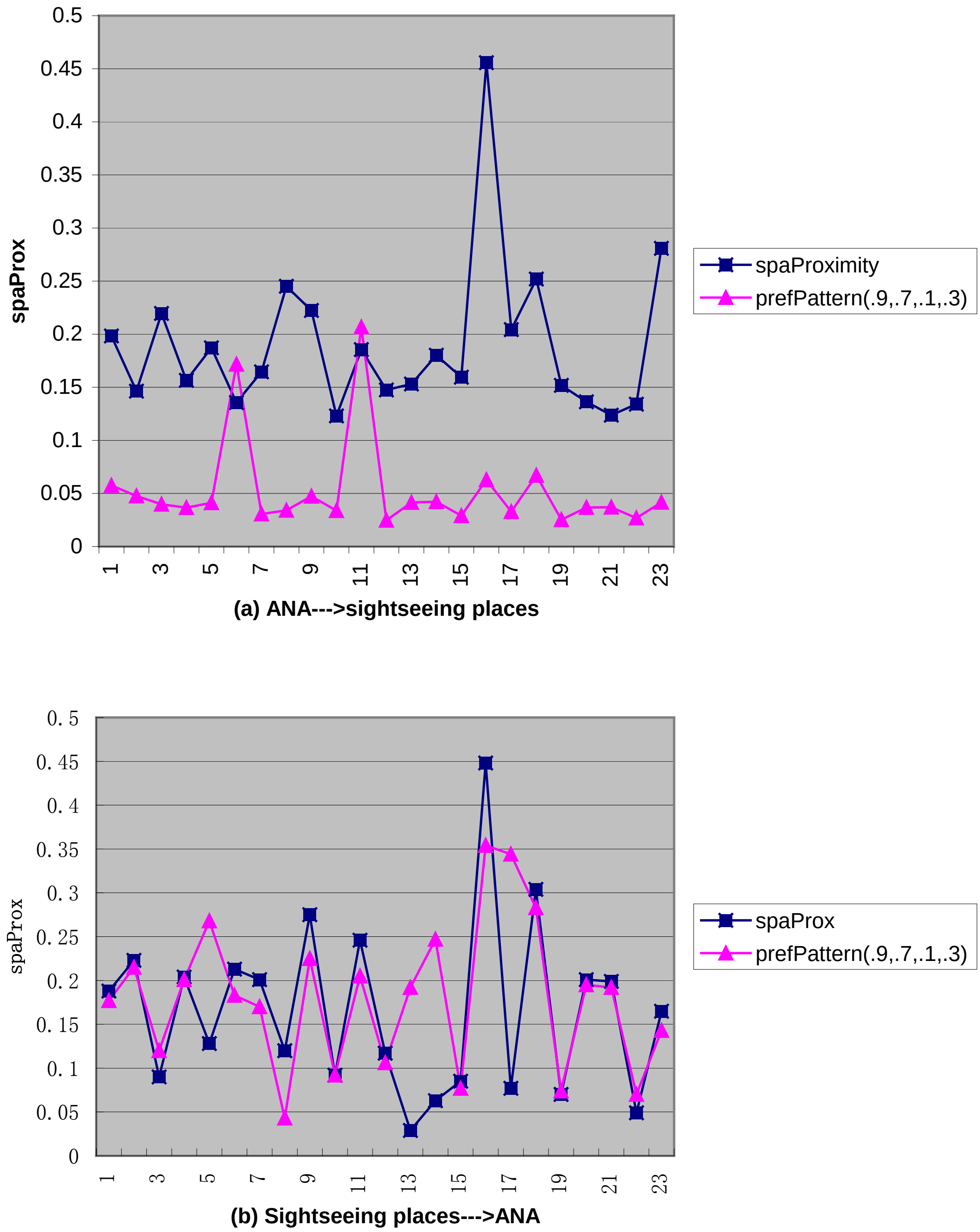


Figure 8.3 Spatial proximities between Ana Hotel and the sightseeing places

Figure 8.2 shows spatial proximities between Nijo and other sightseeing places, the left illustrates the ones from Nijo to the other places, and the right from the others to Nijo. The charts describe spatial proximities on the original map and those on the user-centric conceptual map. It appears that spatial proximities from Nijo to the other sightseeing places are more sensitive, than from others to Nijo, to user preference patterns. This also remains valid in the case of spatial proximities between the ANA Hotel (the user's position) and the sightseeing places (Figure 8.3). These trends result from the fact that Nijo is nearer to the user's location than the others. Comparing figures 8.2 and 8.3, spatial proximities between the ANA Hotel and the sightseeing places are quite different from those between Nijo and the other sightseeing places. This can be explained by the fact that Nijo belongs to the category of Sightseeing places, but the ANA Hotel to a different category.

Semantic interrelationships between classes within hierarchical domain ontology can be used to refine semantic similarity measures. We extract a terminological ontology (Figure 4.6) from WordNet (Miller 1995). This ontology is mainly defined by the partial function of is-a relations between semantic classes. As for our application domain, semantic interrelationships among semantic classes Hotel, Temple, Museum, Garden, Urban are computed with Equation 4.11 (Table 8.1),

	Hotel	Temple	Museum	Garden	Urban
Hotel	1	0.833	0.833	0.048	0.048
Temple	0.833	1	0.833	0.048	0.048
Museum	0.833	0.833	1	0.048	0.048
Garden	0.035	0.035	0.035	1	0.800
Urban	0.035	0.035	0.035	0.800	1

Table 8.11 Semantic similarity between semantic classes in a hierarchical structure

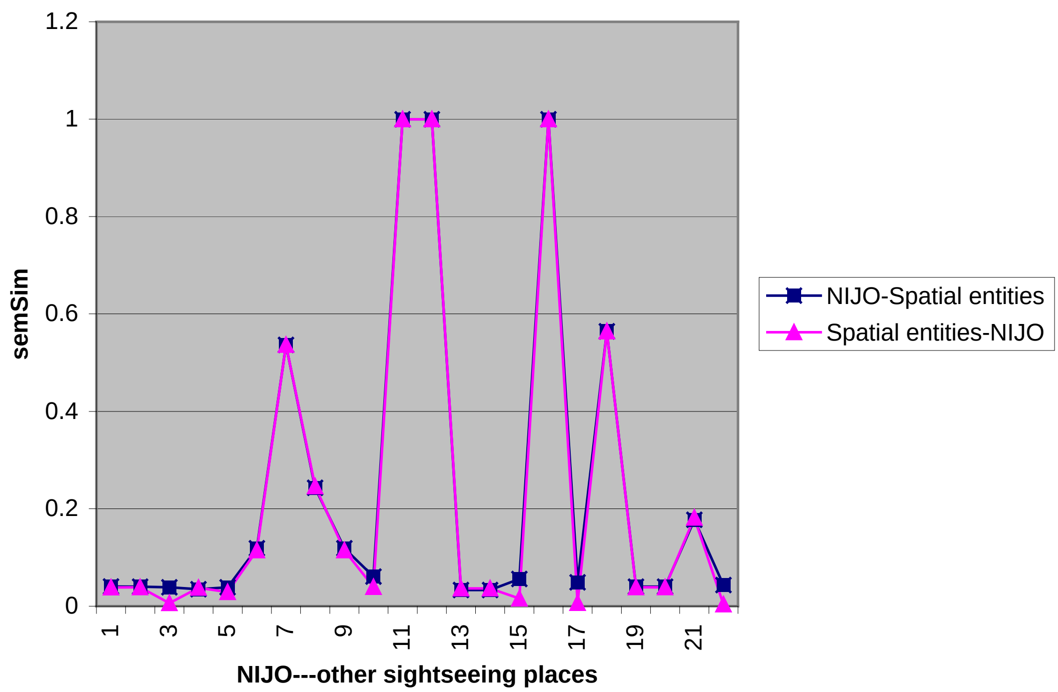


Figure 8.4 Semantic similarities with prefPattern(0.9, 0.7, 0.1, 0.3) (1)

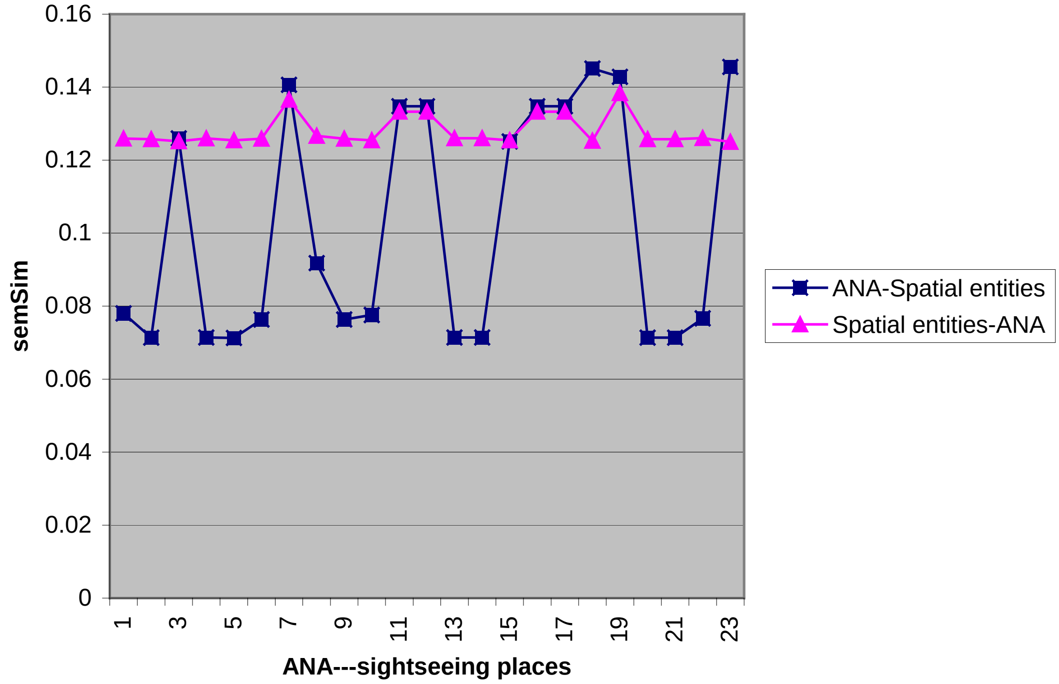


Figure 8.5 Semantic similarities with prefPattern(0.9, 0.7, 0.1, 0.3) (2)

Semantic similarity measures can be specialized into user preferable and non-preferable indexes. A measure of semantic similarity between entities that belong to a same category (Figure 8.4) and to different categories (Figure 8.5) are computed and illustrated. Figure 8.4 shows that semantic similarities from Nijo to the other sightseeing places are almost equivalent (with only four non-matching values) with those from the others to Nijo. This might be due to less semantic diversity and the small size of the data set. Regarding the set of sightseeing places, three places have the same semantic representation as the place NIJO. But, semantic similarities between ANA Hotel and the sightseeing places show explicitly asymmetric characteristics.

The experiment for user-centric spatial proximity and semantic similarity measure triggers the following conclusions:

(1) The user-centric conceptual map has substantial effects on the spatial proximity measure, especially between entities from different categories. Spatial proximities from an entity located nearby a given user's location to other entities are more sensitive to a user preference pattern, than the spatial proximities from the others to it.
(2) Spatial proximity and semantic similarity between spatial entities of same category shows less asymmetric characteristics, than those of different categories. This illustrates our previous assumptions that category and level of abstraction are two influential factors in spatial proximity and similarity

measure.

8.3 Spatial Web personalization experiments

8.3.1 Personalized search strategies

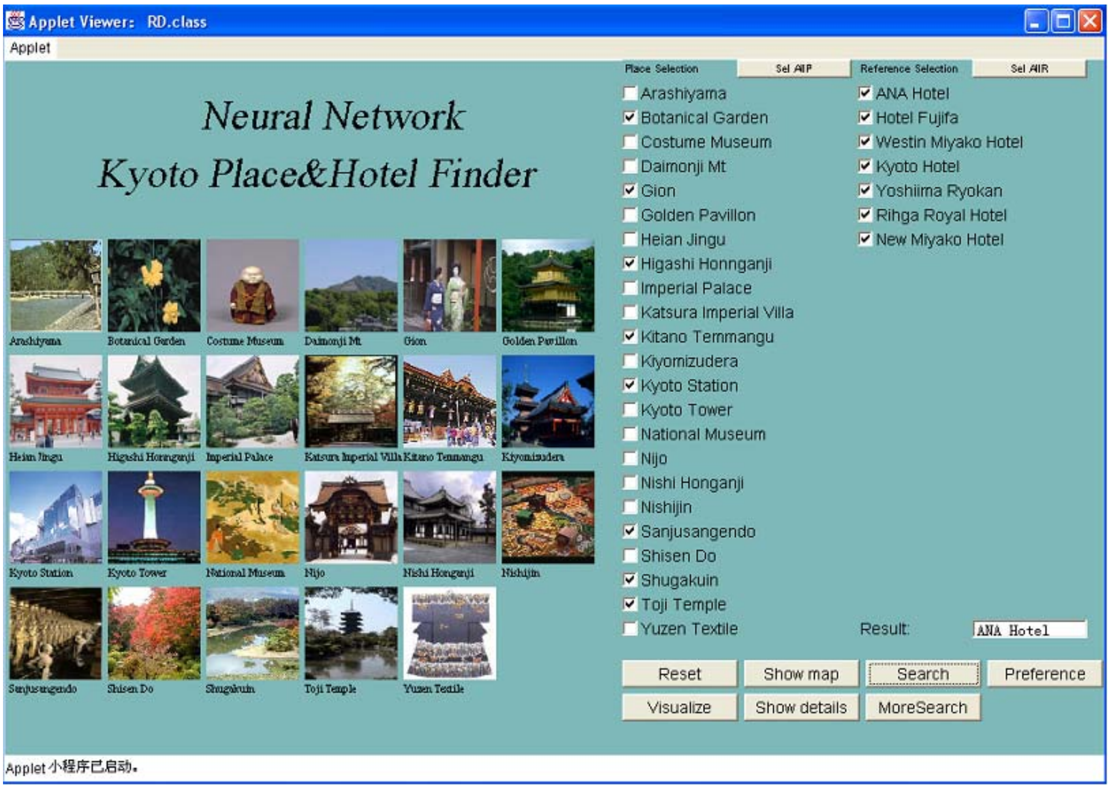


Figure 8.6 The user interface

The Java prototype implements a spectrum of algorithms (A and B are merged into an algorithm AB in the interface as they give similar results) plus a variation of algorithm A based on the absolute distance (denoted as algorithm A0 in the interface). The interface developed so far encodes two main levels of information inputs: places and reference locations. Several places of interest in the city of Kyoto have been pre-selected to give a sufficient range of preference opportunities to the user, and labelled using image schemata. Those places are encoded using fuzzy quantifiers according to predefined semantic classes (urban, temple, garden and museum) and geo-referenced. Reference locations are represented by a list of hotels that are also geo-referenced. Figure 8.6 presents the overall interface of the Kyoto finder. To the left are the image schemata of the places in the city of Kyoto offered for selection, to the top-right the list of seven hotels offered for selection. To the right-bottom is the functional and interaction part of the interface. The algorithm proposed by default is Case D, that is, the one based on an implicit elicitation of user's class preferences.

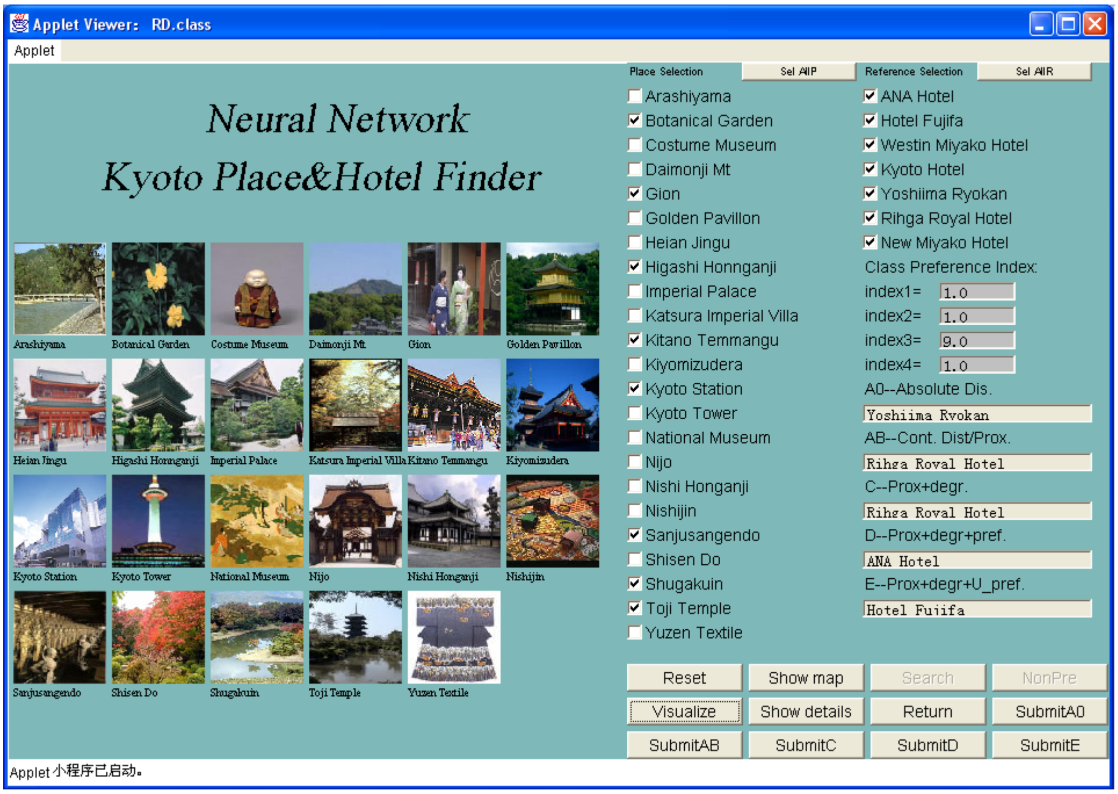


Figure 8.7 Algorithm output examples

The encoding and decoding parts of the algorithms are encapsulated within the Web-interface. The interface provides a selective access to the algorithms by making a distinction between options by default and advanced search facilities. The algorithm applied by default is the one given by the case D where the selection of image schemata is used to derive user's preferences (Figure 8.6). Advanced search facilities offer five algorithms to the user (namely A0, AB, C, D and E as illustrated in Figure 8.7). Figure 8.7 illustrates a case where the application of the algorithms gives different location/hotel winners (at the exception of algorithms AB and C that give a similar result). User's class preferences are explicitly valued by the user when the case E is chosen (index values illustrated at the right-middle interface presented in Figure 8.7).

The whole personalized search process is implemented as two successive steps, namely an initial and a refinement search. The initial personalized search process recalls the best reference location and a set of top-n ranked spatial entities, whose names are displayed at the interface level. After the user's choice of spatial entities and a personalized search algorithm, user's class preferences are derived. In the example of algorithm D illustrated in Figure 4, the ordered list of class preferences is Temple with $p_2 = 0.31$, Garden with $p_3 = 0.27$, Museum with $p_1 = 0.25$, and Urban with $p_4 = 0.16$. When triggered, the neural network calculates the input values in the Y layer, and selects the hotel that best fits the user's preference patterns. Figure 8.8 summarises the results for the previous example (from the left to the right: hotels selected by the user, input values in the layer Y, normalised values in the layer Y). The winning reference location (Ana Hotel) is then propagated back to the X layer where places of interest are ranked according to the algorithm value function.

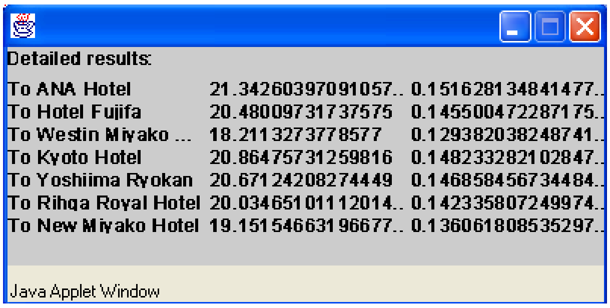


Figure 8.8 Place result examples

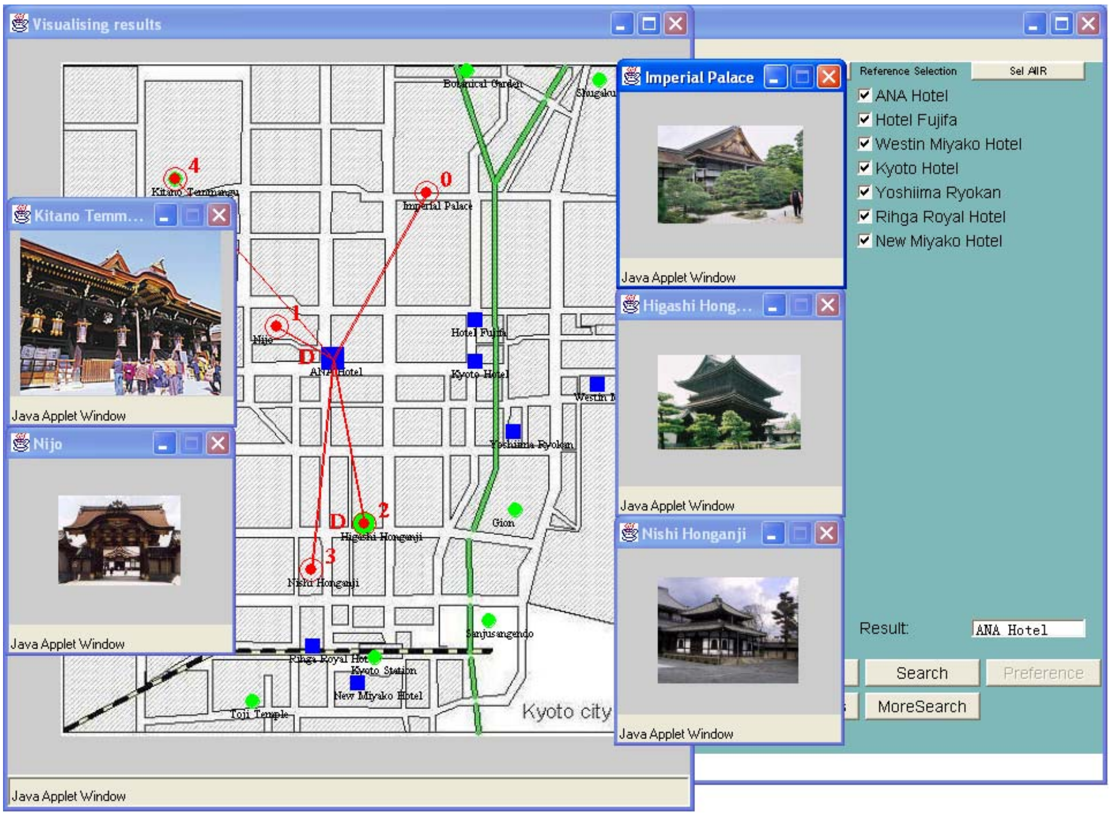


Figure 8.9 Initial results visualization (1)

Figure 8.10 Initial results visualization (2)

The results of the encoding and the decoding process can be presented to the user on the base map of the city of Kyoto. Figure 8.9 presents the map display of the previous example processed using algorithm D in the first stage. The wining hotel (i.e. Ana Hotel) is the best reference location with respect to the user's class preference pattern. The Ana Hotel is denoted by the central square symbol, the best-selected places are the circles connected by a line to that hotel. Each place can be selected by pointing in the interface in order to display the image schemata associated to them, other squares are the other hotels while the isolated circles are the initial selection of the user. The comparison of figures 8.9 and 8.10 shows an interesting pattern. In spite of a same reference location, but the ranking results are different. This results from the fact that the user preference patterns extracted are distinct in the two cases, although there are an equivalent set of spatial entities and a reference location.

If the user is not completely satisfied with the results of the first stage, the second step of the algorithm allows her/him to refine her/his preferences. Figure 8.11 shows the "refining interface", which represents the places selected by the user and the ones returned by the system in the previous stage. The results of the refining process are presented to the user (Figure 8.12). In the example presented, the winning reference place is still the ANA hotel together with several places recommended according to her/his preferences. One can remark that the places returned by the system as illustrated in Figure 8.12 are different from the ones first presented in Figure 8.9, since user preferences have been modified to some degree during the refinement process.

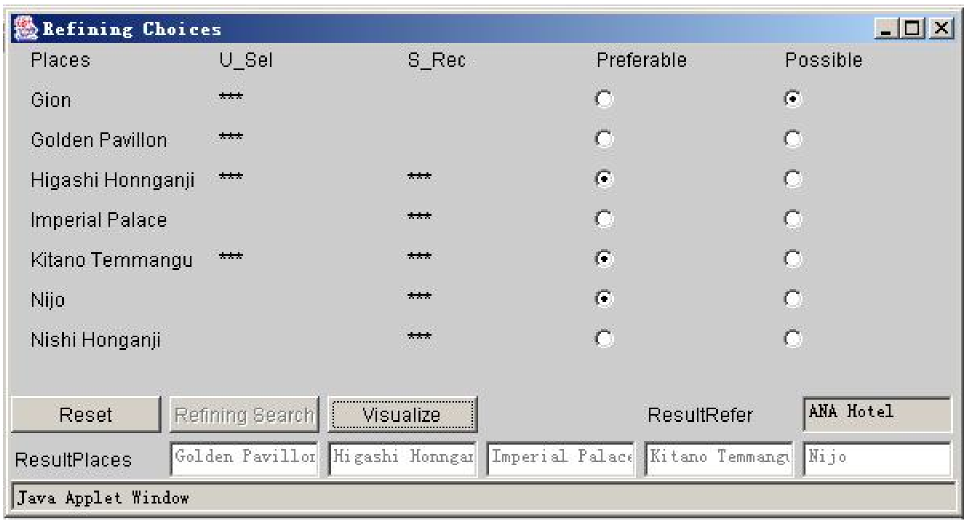
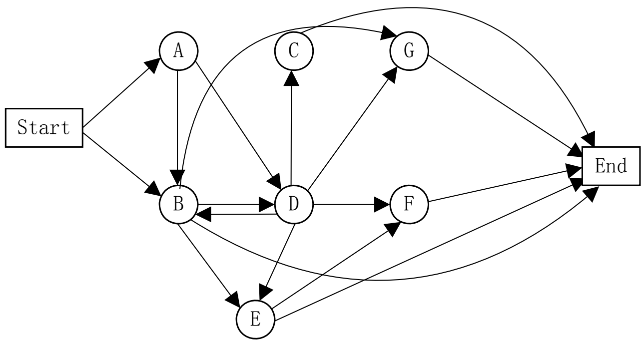


Figure 8.11 Refinement interface

Figure 8.12 Refinement results visualisation

8.3.2 Hybrid personalization engine

This section exemplifies how the proposed approach models Web transactions and personalizes navigational trails over spatial entities embedded on the Web.



Transactions	Frequency
A, B, D, C	10
B, D, B, E	10
A, B, D, F	8
A, B, D, B	4
A, D, B	8
B, D, B, C	8
B, D, E, F	4
A, B, D,B, G	2

Figure 8.13 Navigation trails and transactions

Transactions	Frenquency	Transitional probability
A, B, D --> C	10	5/12
A, B, D --> F	8	1/3
A, B, D --> B	6	1/4
B, D, B --> E	10	5/12
B, D, B --> C	8	1/3
B, D, B --> G	2	1/12
B, D, B --> END	4	1/6
D, E, F --> END	8	1.0
B, D, E --> F	4	1.0
D, B, C --> END	8	1.0
A, D, B --> END	8	1.0
B, D, F --> END	8	1.0
D, B, E --> END	10	1.0
B, D, C --> END	10	1.0
D, B, G --> END	2	1.0

Figure 8.14 Transaction, frequency and transitional probability for 3-order Markov chains

Without loss of generality, let us consider a set of Web transactions as presented in Figure 8.13 (right). The set of Web transactions records user's navigational trails involving several spatial entities that represent some historical and cultural interests. These entities include A (Kitano Temmangu), B (Nijo), C (Higashi Honnganji), D (Yuzen Textile), E (Arashiyama), F (Costume Museum) and G (Nishijin), represented as a labeled directed graph (the left of figure 8.13). We use a third-order Markov chain to model these transactions, i.e., k=3. Through some appropriate Web usage mining processes, the set of user's transactions is transformed to a set of transactions represented as 3-order Markov chains (Figure 8.14). The left part of these transactions forms the state space, while the transitional probabilities constitute the transitional probability matrix.

Suppose a specific user is browsing over the entity D after visiting entities A and B, successively. The Web personalization component takes the transaction $A \rightarrow B \rightarrow D$, and the semantic and spatial criteria into account to predict user's next visits. Possible candidate entities are C, F, B, and transitional probabilities from $A \rightarrow B \rightarrow D$ are 5/12, 1/3, 1/4, respectively. Computed results of *kSemSpa* are presented in Figure 8.15 with γ =2:

```
The computing results of kSemSpa:
kSemSpa(ABD-->B)=1.442
kSemSpa(ABD-->C)=0.900
kSemSpa(ABD-->F)=1.556
```

Figure 8.15 Computed results of *kSemSpa*

Computed results of *kSemSpaM* are (Figure 8.16)

```
The computing results of kSemSpaM:
kSemSpaM(ABD-->B)=0.600
kSemSpaM(ABD-->C)=0.612
kSemSpaM(ABD-->F)=0.720
```

Figure 8.16 Computed results of *kSemSpaM*

The Markov chain evaluation recommends the spatial entity which is most likely to be "visited", the entity F in the example above (Costume Museum). Personalization results are illustrated in Figure 8.17. Consequently, the transitional probability from the sequence <A, B, D> to F, that is, $p(F \mid A, B, D)$ is positively reinforced if the user follows the personalized result. Otherwise, as an example, if the user goes to visit the entity C, then $p(F \mid A, B, D)$ is reinforced negatively and $p(C \mid A, B, D)$ positively.

The hybrid Web personalization approach uses navigational knowledge and profiles extracted from users' historical trails on the Web, and then ameliorates them with the integration of spatial proximity and semantic similarity measures responsible for content-based filtering of spatial information entities. This avoids the defects of using each of the two individually. The reinforcement process is used to update the navigational knowledge through unobtrusively observing user's implicit feedbacks.

Figure 8.17 Personalization results visualization

8.4 Discussion

This chapter illustrates the experimental prototype evaluations of our research framework and relevant techniques identified and applied to spatial Web personalization. The prototype system developed so far illustrates personalization services: user-centric conceptual map, spatial proximity and similarity measures, several personalized search strategies, a hybrid personalization engine and a spatially enriched user interface. There still leaves some implementation issues related to full integration of a semantic user model, and its integration with personalization components.

The overall experimental evaluation addresses how effectively the spatial dimension can bootstrap user preference elicitation and Web personalization processes. The main contribution of the Web personalization processes presented here is the integration of the spatial dimension to facilitate personalization services in spatially-related Web applications.

The objective of our prototype is to act as an exploratory and illustrative solution of user preference elicitation and personalized search approaches for spatial information on the Web. The personalized search algorithms and the hybrid personalization engine presented offer several flexible solutions to the ranking of some reference locations with respect to places of interest in the city of Kyoto. But the principles of the approaches can be extended and applied to other spatial contexts (e.g. location-based mobile services) with some adaptations. The semantic and spatial criteria can be completed by additional semantic and spatial parameters with further consideration of the desired constraint of keeping user's inputs minimal. A second constraint imposed on the spatial Web personalization approaches, is to rely on an acceptable level of complexity in order to guaranty an easy comprehension of the algorithm results. The outputs given by the system are personalized suggestions tailored to the user. Those should allow her/him to actively and interactively explore the different options suggested and to further investigate the Web information space to complete the findings of the BNAM-based personalized search algorithms and the hybrid personalization engine.

Our work has shown that personalization techniques can be applied in the context of spatially related applications. Spatially related personalization aims to improve the qualities of services delivered to the user, thus providing a step towards assisting decision-making in wide-spreading applications such as navigation services (Baus *et al.* 2002), travel and tourism systems (Cheverst *et al.* 2000, Ricci *et al.* 2002), adaptive map services (Zipf 2002, Nafaa 2005), and emergency response and management (Erharuyi and Fairbairn 2003, Xie *et al.* 2005).

Chapter 9 Conclusion

9.1 Contribution summary

Web personalization attracts increasing research and application efforts to facilitate Web information retrieval and navigation. Personalizing information services on the Web implies to develop models, algorithms and procedures that tailor the structure and content of Web documents according to user profiles and expectations. Within the Web engineering community, a lot of search engines, tour guides, personalization agents and intelligent user interfaces have been developed to provide Web pages and information content to the user according to her/his intentions and preferences. However, there is still a need to explore personalization strategies for spatial information services on the Web. Spatial information on the Web is explicitly or implicitly presented, from the physical location of Web site to the spatial semantics contained in Web pages. Furthermore, a significant proportion of Web resources can be associated to some degree to geo-referenced entities. Statistics collected by search engines and systems on the Web show that spatial information is pervasive on the Web, and that many queries explicitly or implicitly contain spatial factors.

One the one hand, conventional Web personalization, user modelling and preference elicitation approaches do not take space as an essential component. On the other hand, spatial applications haven't fully integrated user's interests and preferences to improve information services at the interaction level. This might be explained by the fact that most spatial applications target specialists, instead of general users, as end users. There is still a need to design user models and to elicit user preferences in order to introduce personalization services in spatial applications. User model and preference elicitation in spatial application need to take into account user's personal information, interests and preferences in three complementary domains: space, time and semantic.

Spatial information retrieval and personalization demands appropriate approaches and criteria to manage and categorize spatial information entities. This implies the evaluation of spatial properties and relationships between spatial entities, which amounts to definitions of spatial proximity and semantic similarity concepts. These concepts in turn are relevant to perceptions and interactions between the user and a given environment. Spatial proximity and similarity measures are used in search strategies and relevance ranking of spatial entities. On the other hand, this requires the modeling of user's information and preference elicitation, that is, to represent and infer concepts and relationships about a variety of user features at a semantic level. A set of inference rules can be designed on the basis of a semantic user model, to derive useful user features as the main inputs of personalization functions. Spatial personalization functions get an explicit user query, or inferred from user profiles, and perform a series of personalization processes, e.g. semantic matching. Finally, the candidate entities are ranked according to their relevance to the user's query, and sent the personalized results to the interface. Spatial Web personalization processes are employed to provide personalized navigation assistance to the user who act in a given Web urban space.

This PhD research primarily addresses the issue of personalizing spatial information services on the Web. The proposed research models and manipulates spatial entities preferably at the Web design level. In order to deal with the complexity and interrelationships among a variety of information, spatial entities are manipulated at a finer level of the underlying semantics and domain ontology. Different kinds of spatial entities of interest (e.g. sightseeing places, hotels, universities) are embedded in a variety of multi-media Web documents, e.g., in textual or map forms.

This dissertation introduces an integrated framework for user modeling and preference elicitation and personalization on the spatial Web. It consists of a conceptual personalization service and a semantic user model. The two components communicate information about the user through inter-process communications, e.g., "*tell*" and "*ask*" operations. The personalization service framework unifies spatial and semantic criteria, underlying the three modeling principles:

- an approach for the design of user-centric conceptual maps, spatial proximity and similarity measures,
- image schemata and affordance concepts, and
- Bi-directional Neural Associative Memory (BNAM)-based search and learning mechanisms.

This supports personalized search strategies, a hybrid personalization engine, and a spatially enriched user interface. In order to provide more flexibility, a spectrum of personalized search algorithms is developed. These personalized search algorithms are based on recurrent associative memory to recall a set of spatial entities that best fit user's interests and preferences according to different criteria and functions. User's interests and preferences are inferred and refined in the selection of spatial entities of interest in a Web urban space. The retrieval results are ranked and presented to the user, based on a user-centric spatial proximity and similarity measure. A hybrid personalization approach and reinforcement processes are also introduced to facilitate navigations over and interactions with spatial entities by integrating user's navigational trails, considering that this can improve user preference elicitation and personalization processes. From user preference elicitation and personalization mechanisms, the personalized search strategies and the hybrid personalization engine are based on different principles. The former are based on static inferences, the latter, on dynamic inferences as the personalization engine takes into account user's current navigations. The former allows for active interactions, while the latter performs in a passive mode. Integration of these personalized search strategies and the personalization engine gives flexible mechanisms for supporting interactions between the user and spatial Web applications. Personalization services also supports a Web-based interface enriched with image schemata and affordance concepts that facilitate interactions between the user and the spatial Web, and user preference elicitation process.

The semantic user model employs expressive description logics to represent information and knowledge about the user, and to infer implicit user features from those explicitly available in a user modeling knowledge base. It provides domain-dependent user information as required by personalization components, and to tailor information services according to user's interests and preferences.

An application scenario in the tourism domain exemplifies our research framework. A Web-based Java prototype provides an experimental validation of the BNAM-based user preferences elicitation and personalized search algorithms. The main application scenarios within consideration include: Web-based travel planning and location-aware mobile tourism services. Our research work and the development of a prototype system is mainly oriented to the first application context, and attempts to provide a generic framework for Web-based travel planning in Web urban space. Based on this framework, appropriate inference rules are designed to identify user's interests and preferences, and then to personalize user's travel and experience on the spatial Web.

The research develops a valuable computational environment for personalization services in spatial applications. It allows interactions and multiple explorations of multi-modal data: visual, semantic, textual and cartographical. The fact that the Web is part of a large data repository allows further exploration in the information space. The modelling and computational principles of our approach are general enough to be extended to diverse spatially related application contexts, e.g., location-aware mobile environment. The main contributions of this thesis are:

- An integrated framework for user preference elicitation and personalization for spatial information on the Web.
- A spectrum of personalized search strategies flexibly combining spatial and semantic criteria, and user preference pattern.
- A user-centric spatial proximity and similarity measure.
- A hybrid personalization engine for assisting user's navigation on the spatial Web.
- A logic based semantic user model for describing various assumptions about the user and inferring relevant user features for spatial Web personalization.

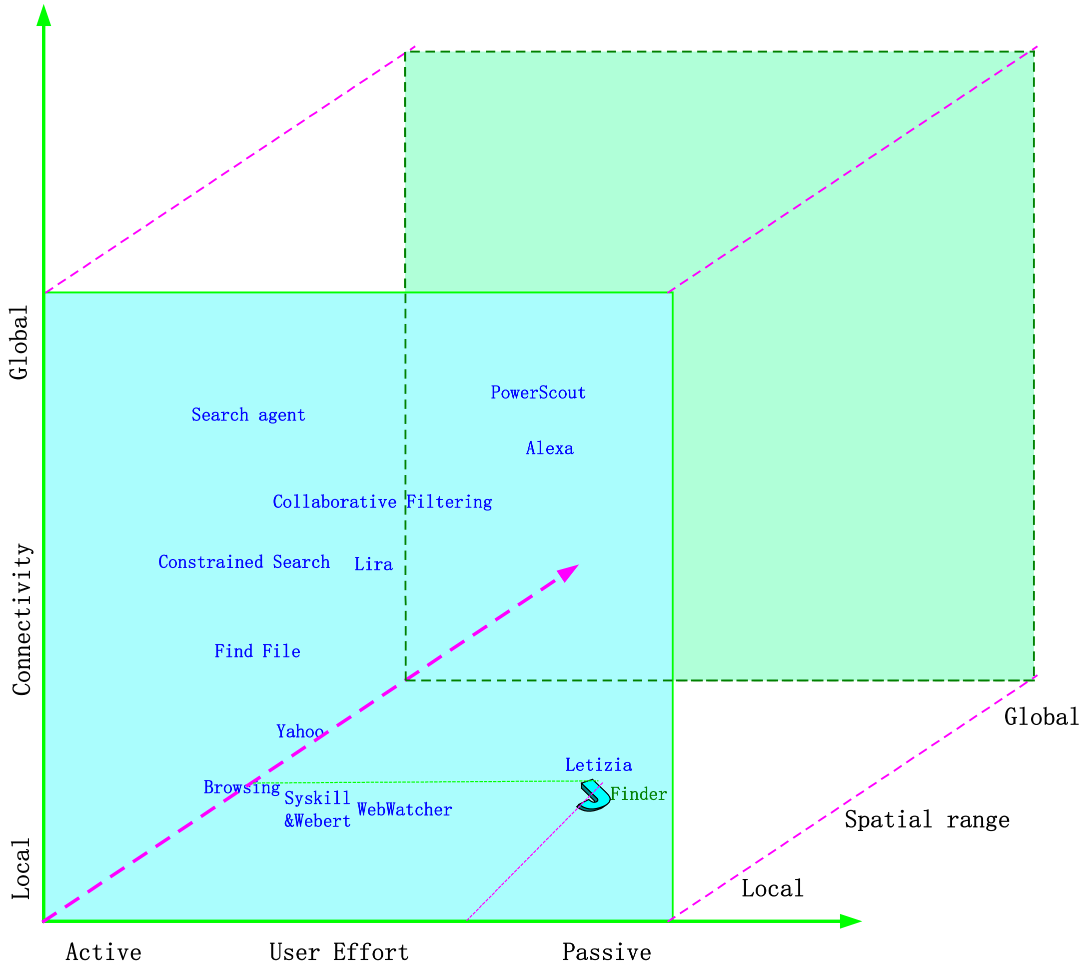


Figure 9.1 A taxonomical framework for Web personalization tools and agents[6]

Following Lieberman's taxonomic approach (Lieberman *et al.* 2001), the frame illustrated in Figure 9.1 describes a three-dimension space for both a-spatial and spatial Web personalization, in which the main tools and agents are plotted against to these attributes, namely user efforts (passive vs active), connectivity (local vs global), and spatial range (local vs global). Our prototype system, **JFinder** is developed at the Web designing level, and illustrated for tourism personalization services on the Web. It supports personalized search strategies and dynamic personalization services over spatial entities stored in local Web site. It can be considered to some degree local in terms of Web connectivity. Secondly, it is somewhat passive in that it implements a hybrid spatial Web personalization approach implemented in a way that doesn't need any explicit user inputs. It can also be viewed as local with respect to the spatial range dimension because of its local personalized search strategy.

[6] Most of Web personalization tools and agents are represented here at approximate locations in order to avoid overlaps.

9.2 Discussion

The spatial Web serves as a background and data repository for location-based mobile services and ubiquitous computing. The spatial Web research amounts to represent, manage and analyse spatial and semantic information on the Web. This implies to understand and model dynamic interactions between human and Web environment (virtual or physical). The objective of spatial Web personalization is to provide more relevant spatial information to the user according to her/his interests and preferences.

The research framework and relevant personalization techniques for spatial Web personalization are illustrated with an application case taken from the tourism domain. This PhD research is distinct at several aspects. (1) It provides a research framework that supports several personalized search strategies based on the Bi-directional Neural Associative Memory (BNAM), a hybrid personalization engine, and a spatially enriched user interface. (2) It proposes an approach for user-centric conceptual map, spatial proximity and similarity measures. (3) It introduces a semantic user model using expressive description logics to represent user information and knowledge about, and to infer relevant user features as required by the personalization services.

The research still deserves require additional considerations and efforts. (1) The approaches are developed preferably at the design level, this being time-consuming and relatively costly for implementation purposes. Another functional restriction results from the fact that these approaches are limited to a local Web site and data base. Thus there is still a need to extend this research to distributed Web environments using appropriate mechanisms such as semantic indexing and database retro-engineering techniques. (2) The research approaches are also partially applied, implemented and validated by the current prototype. Further application scenarios still requires further consideration in order to provide a complete implementation and relevant experimental evaluations.

9.3 Research perspectives

This research opens a wide scope of interesting and challenging issues on spatial Web personalization services.

Understanding the way human manipulate geographical information is a complex and non-deterministic task that implies to model how human perceive and interact with geo-referenced environment (both virtual and physical). This serves as the fundamental principles for user model and preference elicitation, personalization engine and intelligent search on the spatial Web. Whether the specific user is located and moves in spatial environment or not leads to different perceptions and processes. In spatial Web information retrieval and personalization, the spatial cognition requires further attentions. Most similarity measures applied in spatial information services consider spatial proximity as the basic Euclidean distance, and less efforts on possessing an explicit component for spatial proximity (Yao and Thill 2005). Spatial Web information retrieval and personalization require additional attentions to experimental models on the perceptions of spatial entities in a given environment. Experimental space models imply to understand and simulate the way that human beings percept and interact with spatial entities in a given environment. This shall be reflected as parts of spatial Web personalization functions, such as inference rules of user preferences and semantic similarity and spatial proximity measures. In order to make significant progress to more advanced and intuitive spatial Web personalization services, cognitive maps, concepts and relationships developed in geographical spaces are still to be adapted to Web environments.

Our research provides a preliminary User Model Knowledge Base (UMKB) as illustrated in the tourism domain. Its objective is to encode system's assumptions bout user's personal information, and to infer some domain-dependent user features, e.g., beliefs, tasks, interests and preferences. These user features are required by application systems to deliver personalized information services to the user. Further work should explore semantic user models in order to provide semantically enriched user profiles, and inference rules for the elicitation of implicit user features from explicit user information. It is also necessary to evaluate the effects of spatial and temporal properties in the processes of personalization, user model and preference elicitation. This concerns how to identify and qualify preference aspects relevant to the spatial and temporal dimensions.

We plan to implement a semantic user model to support semantic spatial Web personalization applications, and to apply it particularly to e-tourism personalization services. The first approach can be oriented to build a user modeling system based on a DL reasoner. The DL reasoner can be used to check consistency of user profiles, and infer relevant user features required by a personalization component. The second approach is to implement a user modeling system on the platform of an ontology editor integrated with a rule engine. The first approach may produce light-weight user modeling components that fit mobile devices, however it is hard to maintain and update. The latter provide a visual environment for developers to edit and maintain ontologies, and also a set of flexible interfaces to DL reasoners and rule engines. Semantic user modelling components can be used to support value-added Web services because of their capabilities for the representation and reasoning about a variety of user assumptions.

One of the promising directions in semantic Web domain is the semantic spatial Web (Egenhofer 2002). The semantic spatial Web attempts to integrate rich and formal spatial semantics toward machine-understandable Web content and structure. This should favour the generation of more precise and customized information services to the user on the Web. On the one hand, the creation of the semantic spatial Web needs the development of spatial ontologies supported by the formal semantic, qualitative representations of spatial knowledge on the Web, and corresponding retrieval and filtering processes. This implies to investigate spatial knowledge representations on the Web, combining spatial proximity and semantic similarity applied to spatial entities, and to facilitate user preference elicitation and personalization engine.

Interacting with a specific Web system is often only one part of the user's task. Her/his information needs might be met by browsing through several systems (or Websites). This thus requires cross-system personalization, specifically integration of different kinds of information from several heterogeneous sources to assist the user to find relevant information. A problem that still hampers the development of the spatial Web is the lack of a homogeneous standard for the representation and analysis of spatial information. Spatial information is still described in diverse formats, from various data sources, and represented at different levels of abstraction, due to a lack of a widely accepted set of terminology, conceptual model, and data format. Ontology-based approaches view ontologies as a general basis for knowledge sharing, identification and association of semantic-based concepts, to facilitate the integration of different kinds of information (Fonseca *et al.* 2002). Concerning the representation and analysis of spatial knowledge on the Web, this problem becomes unprecedentedly sharp, since significant portions of spatial information is implicitly embedded in Web documents. Therefore, it's laborious to identify, extract and integrate the underlying spatial semantics available on the Web.

The semantic spatial Web should support cross-Website, semantically rich, and more intuitive spatial information personalization services to facilitate interactions between the user and a given spatial environment. Ontology-based functions could to a big degree enhance user preference elicitation and personalization services on the spatial Web.

Spatial Web services essentially combine different sources of multi-dimensional and geographical-related information to represent spatial and semantic data of a given country, region or city on the Web. They are very expressive in terms of information content, and interaction opportunities offered to the users. Qualitative representation of spatial knowledge (Cohn 1997, Cohn and Mazarika 2001) underlines different aspects of a physical space and its representation, including ontological, geometrical and topographical aspects. It provides solid theory foundation but needs to be adapted to Web environments as the underlying reference frames are multi-dimensional, heterogeneous and to a high degree different from a convention physical space. Recently, several researches have been proposed to annotate spatial information on the Web in order to provide effective spatial information services (Hiramatsu and Reitsma 2004). However, there is still a need for appropriate solutions to automatically, or semi-automatically, identify and extract spatial information entities, and then index and annotate them on the Web. Furthermore, flexible representation models are still expected to adapt spatial knowledge representations according to user's tasks, interests and preferences.

Personalization techniques should offer different alternatives tailored to each individual who acts in a given information or physical environment. Potentially, user's activities during the interactions with the environment can be described as an instance of a workflow model. In other words, a sequence of activities is taken by a user with a specific task-oriented goal to favour her/his interests and preferences. Moreover, workflow-based personalization techniques could provide a personalized schedule rather than information recommendations. Ontology-based problem solving knowledge can be used to support the generation of a schedule with goals to achieve certain desired world state and to avoid undesired ones (Chandrasekaran *et al.* 1998). In the tourism domain, workflow-based personalization can be employed to generate a travel agenda with consideration of user's personal information, interests and preferences. A travel agenda can be dynamically updated with user features extracted from information about user's behaviours.

References

Aggarwal, C. C. J. L. Wolf, and P. S. Yu. A new method for similarity indexing for market data. In Proceedings of the 1999 ACM SIGMOD Conference, Philadelphia, PA, June 1999.

Albanese, M., Antonio Picariello, Carlo Sansone, Lucio Sansone: Web personalization based on static information and dynamic user behavior. In Proceedings of the 6th ACM international workshop on Web Information and Data Management (WIDM04). ACM Press: 80-87

Anderson, C.R., Pedro Domingos, and Daniel S. Weld. Personalizing Web Sites for Mobile Users. In Proceedings of the 10th World Wide Web Conference (WWW10). 2001.

Ardissono L. and A. Goy and G. Petrone and M. Segnan and P. Torasso, Tailoring the recommendation of tourist information to heterogeneous user groups, Lecture Notes in Computer Science n. 2266: ``Hypermedia: Openness, Structural Awareness, and Adaptivity'' Int. Workshops OHS-7, SC-3 and AH-3 , pp. 280-295, © Springer Verlag, 2001.

Armstrong, R., D. Freitag, T. Joachims, and T. Mitchell: 1995 "WebWatcher: A Learning Apprentice for the World Wide Web", AAAI 1995 Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments, Stanford, March 1995.

Baader F. and P. Hanschke. A schema for integrating concrete domains into concept languages. In Proc.of the 12th Int.Joint Conf.on Artificial Intelligence (IJCAI-91), pages 452--457, Sydney, 1991.

Baader F. and U. Sattler. An overview of tableau algorithms for description logics. Studia Logica, 69:5--40, 2001.

Baader, F., W. Nutt. Basic Description Logics. In the Description Logic Handbook, edited by F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider, Cambridge University Press, 2003, pages 47-100.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.). (2003a). The Description Logic Handbook. Theory, Implementation and Applications, Cambridge.

Baader, F., Horrocks, I., Sattler, U.: Description Logics as Ontology Languages for the Semantic Web. In: Hutter, D, Stephan, W. (eds.): Festschrift in honor of Jorg Siekmann. Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin Heidelberg New York (2003b)

Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval, ACM Press (1999)

Balabanovic, M., Shoham, Y., Yun, T., An adaptive agent for automated Web browsing", Stanford Univ. Digital Libraries Project, working paper DISL-WP 1995-0023 (1995).

Balabanovíc, M., and Shoham, Y. (1997). Fab: Content-Based, Collaborative Recommendation. Communications of the ACM 40 (3), pp. 66-72.

Baudisch. P. Joining Collaborative and Content-based Filtering. In Proceedings of the ACM CHI Workshop on Interacting with Recommender Systems, Pittsburgh, PA, May 1999. ACM Press.

Baudisch. P. Joining Collaborative and Content-based Filtering. In Proceedings of the ACM CHI Workshop on Interacting with Recommender Systems, Pittsburgh, PA, May 1999. ACM Press.

Baus, J., Krüger, A., Wahlster, W., (2002): A Resource-Adaptive Mobile Navigation System. In: Proceedings of the 2002 International Conference on Intelligent User Interfaces (IUI'02), ACM Press, pp. 15- 22, ISBN 1-58113-459-2.

Belkin, N.J., Croft, W.B., Information Retrieval and Filtering: Two Sides of the Same Coin, Communications of the ACM, 35(12), (1992) 29-38

Benelli, G., Caporali, M., Rizzo, A. et Rubegni, E. (2002) Design concepts for learning spatial relationships. *In : Proceeding of the 19th annual international conference on Computer documentation* , Sante Fe, New Mexico, USA, University of Siena. pp. 22-30.

Bera, R. and Claramunt, C., 2005, Connectivity inferences over the Web for the analysis of semantic networks, K.-J. Li and C. Vangenot (Eds.), Springer, Lecture Notes in Computer Science 3833, December, pp. 222 – 234.

Berendt, B., Hotho, A., & Stumme, G. (2002). Towards Semantic Web Mining. In I. Horrocks & J. Hendler (Eds.), The Semantic Web - ISWC 2002 (Proceedings of the 1st International Semantic Web Conference, June 9-12th, 2002, Sardinia, Italy) (pp. 264-278). LNCS 2342, Heidelberg, Germany: Springer

Berghell, M., Andrea Capata, Tiziana Catarci, Paolo Cerrocchi, Paolo Masi, Marco Oppedisano, Amiliano Trevisani, Andrea Vitaletti, Andrea Calì. 2005 Smartdate:User Adaptation in Location-based Mobile matchmaking. International Workshop on Plastic Services for Mobile Devices, Rome, Italy, 12 September 2005

Berka, T. and Manuela Plößnig (2004): Designing Recommender Systems for Tourism. In: ENTER 2004, Kairo.

Berners-Lee, J., Hendler, J. and O. Lassila, The Semantic Web, Scientific American, vol. 184, no. 5, pp. 34-43, 2001

Bishop L. 2005 Mannix Marketing SEO Newsletter, Retrieved Sep. 20 from

http://www.mannixmarketing.com/articles-explosion-of-local-search.htm

Bodner, R. and M. Chignell. Dynamic Hypertext: Querying and Linking. ACM Computing Surveys, Vol. 31(4es), December 1999. Article No. 15. Available online:

http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/39.html

Borgida, A., Brachman, R.J.: Conceptual Modelling with Description Logics. In Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P., eds.:Description Logic Handbook, Cambridge University Press (2002) 349 – 37

Borgman., C. The User's Mental Model of an Information Retrieval System: An Experiment on a Prototype On-line Catalogue. International Journal of Man-Machine Studies, Vol. 24(1):pages 47–64, 1986.

Breese, J.S. David Heckerman and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, pages 43--52, July 1998.

Brin S. and L. Page.: 1998 The anatomy of a large-scale hypertextual Web search engine, Computer Networks and ISDN Systems, 30:107--117

Brown, B. and Chalmers, M, 2003, Tourism and mobile technology. In Proceedings of the 8th European Conference on Computer Supported Cooperative Work, 14th-18th September, Helsinki, Finland: Kluwer Academic Publishers.

Burke, R., Hammond, K. & Cooper, E. Knowledge-based Information Retrieval from Semi-structured Text. In AAAI Workshop on Internet-based Information Systems, page 9-15. AAAI, 1996.

Burke, R., Hammond, K., and Young, B. The FindMe Approach to Assisted Browsing. IEEE Expert, 12(4), pages 32-40, 1997

R. Burke. Integrating Knowledge-based and Collaborative- ltering Recommender Systems. In Proceedings of the Workshop on Arti cial Intelligence and Electronic Commerce (AAAI99). AAAI Press/MIT Press, 1999.

Burke, R. Knowledge-based Recommender Systems. In A. Kent (ed.), Encyclopedia of Library and Information Systems. Vol. 69, Supplement 32. New York: Marcel Dekker, 2000

Buyukkokten O., Cho J., Garcia-Molina H., Gravano L. and Shivakumar N. (1999). Exploiting Geographical Location Information of Web Pages. In Proc. of the ACM SIGMOD Workshop on the Web and Databases (WebDB'99), Philadelphia, USA.

Bush, V. (1945) "As we may think" Atlantic monthly, 176 (1), 101-108. http://www.ps.uni-sb.de/%7Educhier/pub/vbush/vbush-all.shtml

Cahuzac, Hubert, Leblanc, Benoît (2004), From intuitive mapping to concept mapping : an application within an anthropological urban field IN Canas, Alberto et Novak, Joseph et Gonzalez, Firmin (ed), Concept Maps : Theory, Methodology, Technology, university publication of Navarra, .

Calì, A., Diego Calvanese, Simona Colucci, Tommaso Di Noia, Francesco M. Donini: A Logic-Based Approach for Matching User Profiles. Mircea Gh. Negoita, Robert J. Howlett, in Lakhmi C. Jain (Eds.): Knowledge-Based Intelligent Information and Engineering Systems, Proceedings of 8th International Conference, KES 2004, Wellington, New Zealand, September 20-25, 2004., Springer 2004 187-195

Callan, J., Smeaton, A.: Personalization and recommender systems in digital libraries. Technical report, DELOS-NSFWorkshop on Personalization and Recommender Systems in

Digital Libraries (2003) Further Contributors: Beaulieu M., Borlund P., Brusilovsky P., Chalmers M., Lynch C., Riedl J., Smyth B., Straccia U., Toms E.

Calvanese, D., Giuseppe De Giacomo, and Maurizio Lenzerini, Conjunctive query containment in Description Logics with n-ary relations, In Proc. of the 1997 Description Logic Workshop (DL'97), pages 5-9, 1997

Calvanese, D., Maurizio Lenzerini, and Daniele Nardi, Description logics for conceptual data modeling, In Jan Chomicki and Günter Saake, editors, *Logics for Databases and Information Systems*, pages 229-263. Kluwer Academic Publisher, 1998a.

Calvanese, D., Giuseppe De Giacomo, and Maurizio Lenzerini, On the decidability of query containment under constraints, In Proc. of the 17th ACM SIGACT SIGMOD SIGART Sym. on Principles of Database Systems (PODS'98), pages 149-158, 1998b

Calvanese, D., Maurizio Lenzerini, and Daniele Nardi, Unifying class-based representation formalisms, J. of Artificial Intelligence Research, 11:199-240, 1999.

Carmagnola, F., Federica Cena, Cristina Gena, Ilaria Torre, A Multidimensional Approach for the Semantic Representation of Taxonomies and Rules in Adaptive Hypermedia Systems. In the proceedings of the PerSWeb'05 workshop (UM 2005).

Cecconi A., Galanda M. (2002): Adaptive Zooming in Web Cartography. In Proceedings of SVG Open 2002 (Zurich, Switzerland).

Cecconi, A., Shenton, C., and Weibel, R., (1999): Tools for Cartographic Visualization of Statistical Data on the Internet. In Proceedings of the 19th International Cartographic Conference (Ottawa, Canada).

Chandrasekaran, B., J.R. Josephson, and R. Benjamins. The Ontology of Tasks and Methods. In Proceedings of the 11th Knowledge Acquisition Modeling and Management Workshop, KAW'98, Banff, Canada, April 1998.

Chen, L. and Katia Sycara. WebMate: A personal agent for browsing and searching. In Katia P. Sycara and Michael Wooldridge, editors, Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98), pages 132--139, New York, 1998. ACM Press.

Cheverst, K., Nigel Davies, Keith Mitchell, Adrian Friday and Christos Efstratiou. Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences Proceedings of CHI 2000, Netherlands, April 2000, pp 17-24.

Chiclana, F., Herrera, F. and Herrera-Viedma, E., 1998, Integrating three representation models in multipurpose decision making based on preference relations. Fuzzy Sets and Systems, 97, 33-48.

Chrisment C., DOUSSET B., KAROUACH S., MOTHE J. (2004) "Information mining : extracting, exploring and visualising geo-referenced information". Workshop on Geographic Information Retrieval, SIGIR 2004

Cingil, I., Dogac, A. & Azgin, A. 2000. "A broader approach to personalization." Communications of the ACM, Aug 2000; Vol. 43(8) 136-142.

Claypool M., Anuja Gokhale, Tim Miranda, Pavel Murnikov, Dmitry Netes, and Matthew Sartin. Combining Content-Based and Collaborative Filters in an Online Newspaper, ACM SIGIR Workshop on Recommender Systems, Berkeley, CA, August 19, 1999.

Cohn, A. (1997). Qualitative spatial representation and reasoning techniques. In Brewka, G., Habel, C., and Nebel, B., editors, KI-97: Advances in Artificial Intelligence, Lecture Notes in Artificial Intelligence, pages 1--30. Springer-Verlag, Berlin.

Cohn A. G. and S. M. Hazarika. Qualitative spatial representation and reasoning: an overview. Fundamenta Informaticae, 46(1-2):1--29, 2001.

Cooley, R., B. Mobasher, and J. Srivastava. Web mining: Information and pattern discovery on the World Wide Web. In International Conference on Tools withArtificial Intelligence, pages 558 – 567, Newport Beach, CA, 1997.

Cooley, R., B. Mobasher, and J. Srivastava. Data preparation for mining world wide Web browsing patterns. Knowledge and Information Systems, 1(1):5--32, 1999.

Cooper, W., F. Gey, and D. Dabney. Probabilistic retrieval based on staged logistic regression. In N. J. Belkin, P. Ingwersen, and A. M. Pejtersen, editors, Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Copenhagen, Denmark, June 21-24, 1992.

Cullot, N., Christine Parent, Stefano Spaccapietra, and Christelle Vangenot Ontologies: A Contribution to the DL/DB debate In Proc. of the 1st International Workshop on the Semantic Web and Databases, 29th International Conference on Very Large Data Bases, Berlin Germany, September 7-8 2003

Dai H. and B. Mobasher. Using ontologies to discover domain-level Web usage profiles. In 2nd Semantic Web Mining Workshop at ECML/PKDD-2002, 2002.

Dai, H., B. Mobasher, 2005 Integrating Semantic Knowledge with Web Usage Mining for Personalization (with). In Web Mining: Applications and Techniques, Anthony Scime (ed.), IRM Press, Idea Group Publishing. 2005.

De Giacomo G. and M. Lenzerini. Concept languages with number restrictions and fixpoints, and its relationship with -calculus. In Proc. 11th European Conference on Artificial Intelligence (ECAI-94), pages 411--415. John Wiley and Sons, 1994.

De Rijke, M. Modal logics and description logics. In E. Franconi, G. De Giacomo, R.M. MacGregor, W. Nutt, and C.A. Welty, editors, Proc. DL'98, pages 1--3, 1998.

Der Hoek, V., W., & de Rijke, M. (1995). Counting objects. Journal of Logic and Computation, 5 (3), 325--345.

Deshpande, M., Karypis G.: 2004 Selective Markov models for predicting Web page accesses, ACM Trans. Internet Techn. 4(2): 163-184

Dolog, P., Henze, N., Nejdl, W. and Sintek, M. Personalization in distributed e-learning environments. In Proc. of WWW2004 | The Thirteenth International World Wide Web Conference, May 2004.

Eirinaki, M., M. Vazirgianis. "Web Mining for Web Personalization", ACM Transactions on Internet Tehnology (TOIT), volume 3, issue 1, 2003

Eirinaki, M., M. Vazirgiannis, H. Lampos, S. Pavlakis, Web Personalization Integrating Content Semantics and Navigational Patterns, submitted for revision at WIDM 2004

Doyle, J., Q. Han, J. Weakliam, M. Bertolotto, D. Wilson, Developing Non-Proprietary Personalized Maps for Web and Mobile Environments, Proceedings W2GIS'04 (4th International Workshop on Web and Wireless Geographical Information Systems), Koyang, Korea, November 2004.

Egenhofer, M.J. (2002) Toward the Semantic Geospatial Web. In Proceedings of the Tenth ACM International Symposium on Advances in Geographic Information Systems, McLean, Virginia

Eirinaki, M., M. Vazirgianis. "Web Mining for Web Personalization", ACM Transactions on Internet Tehnology (TOIT), volume 3, issue 1, 2003

Erharuyi, N., D. Fairbairn (2003) Mobile geographic information handing technologies to support disaster management. Geography 88(4): 312-319.

Fensel D. and M. Musen, The Semantic Web: A New Brain for Humanity, IEEE Intelligent Systems, vol. 16, no. 1, pp. 24-25, 2001

Fink, J. and A. Kobsa (2000): A Review and Analysis of Commercial User Modeling Servers for Personalization on the World Wide Web. User Modeling and User-Adapted Interaction 10(3-4), Special Issue on Deployed User Modeling, 209-249

Fink J. and A. Kobsa. User modeling for personalized city tours. Artificial Intelligence Review, 18(1):33--74, 2002.

Fischer, G.: User modeling in human-computer interaction. User Modeling and User-Adapted Interaction, 11(1&2):65-86, 2001.

Fonseca, F. T., Egenhofer, M. J., Davis, C. A., & Borges, K. A. V. (2000). *Ontologies and Knowledge Sharing in Urban GIS*. Computer, Environment and Urban Systems, 24(3), 232-251.

Fonseca, F. T., Egenhofer, M. J., Agouris, P., & Camara, G. (2002). Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS, 6(3), 231-257.

Franklin, N., & Tversky, B. (1990). Searching imagined environments. Journal of Experimental Psychology: General, 119(1), 63-76.

Franklin, N., Tversky, B, & Coon, V. (1992). Switching points of view in spatial mental models. Memory & Cognition, 20, 507-518.

Freeman, JA., Skapura, DM.: Neural Networks: Algorithms, Applications and Programming Techniques, Addison-Wesley, MA (1991)

Fu, G., C.B.Jones, A.I.Abdelmoty, Ontology-Based Spatial Query Expansion in Information Retrieval, to appear in Proceedings of the OTM Confederated International Conferences

CoopIS, DOA, and OOBASE, 2005a.

Fu, G., C. B. Jones and A. I. Abdelmoty, Building a Geographical Ontology for Intelligent Spatial Search on the Web, in Proceedings of IASTED International Conference on Databases and Applications, pp167-172, 2005b.

Gallagher, A. and Golde, C. (1999) "The Challenges of Conducting Interdisciplinary Research in Traditional Doctoral Programs." Ecosystems, 2 (June/July): 281-285.

Gauch, S., Chaffee, J., Pretschner, A.:Ontology-Based Personalized Search and Browsing J. Web Intelligence and Agent Systems 1(3-4):219-234, 2003

Gaye, L., Maz, R., and Holmquist, L. E. Sonic City: The Urban Environment as a Musical Interface. In Proceedings of New Interfaces for Musical Expression (NIME'03) (Montral, Canada, May 22-24, 2003)

Gibson, James. J. The Ecological Approach to Visual Perception. Lawrence Eribaum. 1979.

Goldberg, D., David Nichols, Brian M. Oki, and Douglas Terry. Using collaborative ltering to weave an information tapestry. Communications of the ACM, 35(12):61-70, 1992.

Goldstone, R. L, & Son, J. (2005). Similarity. In K. Holyoak & R. Morrison (Eds.). Cambridge Handbook of Thinking and Reasoning. Cambridge: Cambridge University Press. (pp. 13-36)

Gu, T., Xiao Hang Wang, Hung Keng Pung, Da Qing Zhang. A Middleware for Context-Aware Mobile Services. IEEE Vehicular Technology Conference. Milan, Italy, 2004

Goldberg, D., D. Nichols, B. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. CACM, 35(12):61--70, 1992.

Goy A. and D. Magro, "STAR: a Smart Tourist Agenda Recommender, ", Proc. Configuration Workshop at ECAI 2004, Valencia, Spain, 2004

Greco, G, Greco, S, Zumpano, E: A Probabilistic Approach for Distillation and Ranking of Web Pages. World Wide Web: Internet and Information Systems, 4(3), (2001) 189-208

Gruber, T.R. 1995. Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, 43(5-6):907--928.

Haarslev V. and R. Möller. RACER System Description. In R. Goré, A. Leitsch, and T. Nipkow, editors, International Joint Conference on Automated Reasoning, IJCAR'2001, June 18-23, Siena, Italy, pages 701–705. Springer-Verlag, 2001a.

Haarslev V. and R. Möller. Description of the RACER System and its Applications. In Proceedings International Workshop on Description Logics (DL-2001), Stanford, USA, 1.-3. August, pages 131–141, 2001b.

Haddawy, P., Ha, V., Restificar, A., Geisler, B. and Miyamoto, J., 2003, Preference elicitation via theory refinement. Journal of Machine Learning Research, 4, 317-337.

Haffernick, J. J., Messerschmitt, D. S. and Vandrick, S. (1997), Collaborative research: Why and how? Educational researcher, December, 31-35.

Hammond, K. (1989). Case-Based Planning: Viewing Planning as a Memory Task. Academic Press, San Diego.

Heckmann D. and Antonio Krüger. A User Modeling Markup Language (UserML) for Ubiquitous Computing ,User Modelling 2003, Springer, 2003.

Heckmann, D., Tim Schwartz, Boris Brandherm, Michael Schmitz, Margeritta von Wilamowitz-Moellendorff GUMO - the General User Model Ontology Proceedings of the 10th International Conference on User Modeling (UM'2005), Edinburgh, UK, 2005a

Heckmann, D., Tim Schwartz, Boris Brandherm, Alexander Decentralized User Modeling with UserML and GUMO Kröner Proceedings of the Workshop on Decentralized, Agent Based and Social Approaches to User Modelling (DASUM 2005), Edinburgh, UK, 2005b

Hiramatsu, K., and Reitsma, F.: GeoReferencing the Semantic Web: ontology based markup of geographically referenced information, Joint EuroSDR/EuroGeographics workshop, 2004.

Hirtle, S.C., Jonides, J.: Evidence of Hierarchies in Cognitive Maps. Memory & Cognition, 13 (3), 208-217 (1985)

Hörnig, R., K. Eyferth & H. Gärtner (2000). Orienting and Reorienting in Egocentric Mental Models. In C. Freksa, W. Brauer, C. Habel & K. Wender (Eds.) Spatial Cognition II -- Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications. Springer Verlag, LNCS 1849.

Horrocks, I. The FaCT system. In H. de Swart, editor, Automated Reasoning with Analytic Tableaux and Related Methods: International Conference Tableaux'98, number 1397 in Lecture Notes in Artificial Intelligence, pages 307-312. Springer-Verlag, May 1998.

Horrocks, I. FaCT and iFaCT. In P. Lambrix, A. Borgida, M. Lenzerini, R. Möller, and P. Patel-Schneider, editors, Proceedings of the International Workshop on Description Logics (DL'99), pages 133-135, 1999.

Horrocks, I., U. Sattler, and S. Tobies. Practical reasoning for very expressive description logics. Logic Journal of the IGPL, 8(3):239-263, 2000a.

Horrocks, I., U. Sattler, and S. Tobies. Reasoning with individuals for the description logic SHIQ. In David MacAllester, editor, Proc. of the 17th Int. Conf. on Automated Deduction (CADE 2000), number 1831 in Lecture Notes in Artificial Intelligence, pages 482-496. Springer-Verlag, 2000b.

Horrocks, I., P. F. Patel-Schneider, and F. van Harmelen. From SHIQ and RDF to OWL: The making of a Web ontology language. Journal of Web Semantics, 1(1):7--26, 2003.

Hunolstein von and S Zipf, A.. (2003): Task oriented map-based mobile tourist guides.International Workshop: "HCI in Mobile Guides". at Mobile HCI 2003. Fifth International Symposium on Human Computer Interaction with Mobile Devices and Services. 8.-11.Sept. 2003. Undine. Itlay.

Jansen, Bernard J. and Udo W. Pooch: A review of Web searching studies and a framework for future research. Journal of the American Society for Information Science and Technology (JASIST), (3): 235-246 (2001)

Jin X. and B. Mobasher, Using Semantic Similarity to Enhance Item-Based Collaborative Filtering, in Proceedings of The 2nd IASTED International Conference on Information and Knowledge Sharing, Scottsdale, Arizona, November 2003.

Joachims, T., Freitag, D., Mitchell, T.: WebWatcher: A Tour Guide for the World Wide Web. Proceedings of International Joint Conference on Artificial Intelligence (IJCAI), Morgan Kaufmann, (1997) 770-777

Joachims, T.: Optimizing Search Engines Using Clickthrough Data. In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-02), ACM Press, (2002)133--142

Johnson, M. The body in the mind: the bodily basis of meaning, imagination, and reason, University of Chicago Press, (1987).

Jones C.B., Alani H and Tudhope D - Geographical information retrieval with ontologies of place Spatial Information Theory Foundations of Geographic Information Science COSIT 2001 LNCS 2205 Springer, Berlin (2001) 323-335 ISSN 0302-9743 ISBN 3-540-42613-2

Jones C.B., R. Purves, A. Ruas, M. Sanderson, M. Sester, M.J. van Kreveld, R. Weibel (2002). ‘Spatial information retrieval and geographical ontologies an overview of the SPIRIT project’. SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland, 387-388.

Kacprzyk, J., 1986, Group decision making with a fuzzy linguistic majority. Fuzzy Sets and Systems, 18, 105-118.

Keates, S., LANGDON P, CLARKSON PJ, ROBINSON P (2002) 'User models and user physical capability. ' Journal of User Modeling and User-Adapted Interaction (UMUAI), Special issue 1: Empirical evaluation of User Models, Wolters Kluwer, 12, 139-169, 2002.

Keita A., LAURINI R., ROUSSEY C., ZIMMERMAN M. (2004)"Towards an Ontology for Urban Planning: The Towntology Project". Accepted for the 24th UDMS Symposium, Chioggia, October 27-29, 2004.

Kelly, D., Teevan, J.: Implicit Feedback for Inferring User Preference: a Bibliography. SIGIR Forum 37(2), (2003) 18-28

Kerschberg, L., Kim, W. and Scime, A., WebSifter II: A Personalizable Meta-Search Agent Based on Semantic Weighted Taxonomy Tree. in The 2001 International Conference on Internet Computing, IC 2001, (Las Vegas, Nevada, 25-28 June 2001).

Kim, D., Il Im, Vijayalakshmi Atluri, Michael Bieber, Nabil R. Adam, Yelena Yesha: 2004, A clickstream-based collaborative filtering personalization model: towards a better performance. In Proceedings of the 6th ACM international workshop on Web Information and Data Management (WIDM04). ACM Press: 88-95

Kleinberg, JL.: Authoritative Sources in an Hyperlinked Environment. Journal of the ACM, 46(5): (1999) 604-632

Kleinberg, J., and S. Lawrence. The Structure of the Web. Science 294 (2001), 1849-1850.

Kobsa, A. (1993): User Modeling: Recent Work, Prospects and Hazards. In: M. Schneider-Hufschmidt, T. Kühme and U. Malinowski, eds. (1993): Adaptive User Interfaces: Principles and Practise. Amsterdam: North Holland Elsevier.

Kobsa, A. and W. Pohl (1995): The User Modeling Shell System BGP-MS. User Modeling and User-Adapted Interaction 4(2), 59-106.

Kobsa, A. 2000, User Modeling as a Key Factor in System Personalization, Retrieved Sep. 16 2005 from

http://www.zurich.ibm.com/~mrs/chi2000/contributions/kobsa.html

Kobsa, A. Generic User Modeling Systems. User Modeling and User Adapted Interaction, Ten Year Anniversary Issue, Kobsa (Ed.), 11 (1/2), pp. 49-63, Kluwer Academic Publishers, 2001.

Kolodner, J. 1993. Case-based reasoning. San Mateo, CA: Morgan Kaufmann.

Konstan, J. A. Miller, B. N. Maltz, D. Herlocker, J. L. & Gordon, L. R. Riedl, J. `GroupLens: Applying Collaborative Filtering to Usenet News' in `special section: recommendation systems' in CACM March 1997, Vol. 40, No. 3, pp77-87

Kosko, B. Adaptive bi-directional associative memories, Appl. Opt. 26 (23) (1987) 4947-4960.

Kosko, B. (1988), "Bidirectional Associative Memories", IEEE Transaction on Systems, Man and Cybernetics, vol. SMC, no. 18, pp. 49-60.

Kosko, B. Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence, Prentice-Hall, 1992.

Kuhn, W.(1996): 'Handling Data Spatially: Spatializing User Interfaces'. In Proceedings of 7th International Symposium on Spatial Data Handling, SDH'96, Advances in GIS Research II, (Kraak, M.-J., & Molenaar, M., eds.), in Delft, The Netherlands (August 12-16, 1996), Published by IGU, Vol. 2, pp: 13B.1 - 13B.23.

Lang, K. Newsweeder: Learning to filter news. In Proceedings of the 12th International Conference on Machine Learning, pages 331--339, 1995.

Larson R.R. and P. Frontiera. “Ranking and Representation for Geographic Information Retrieval”. Extended abstract in SIGIR 2004 Workshop on Geographic Information Retrieval. (Sheffield, UK, July 29, 2004a)

Larson R.R. and P. Frontiera. 2004b Spatial Ranking Methods for Geographic Information Retrieval (GIR) in Digital Libraries. In Research and Advanced Technology for Digital Libraries Rachel Heery and Liz Lyon (Eds), Proceedings 8th European Conference, ECDL 2004, Bath, UK, September 2004 Proceedings. pp. 45-57, LNCS 3232, 2004.

Lawrence, S. (2000). Context in Web search. IEEE Data Engineering Bulletin, 23(3):25--32.

Lesk, M., 1997, Going digital. Scientific American, 276 (3), 58-60

Leuski A. and James Allan. 2004. "Interactive information retrieval using clustering and spatial proximity". User Modeling and User Adapted Interaction (UMUAI), 14(2-3), p. 259 - 288

Levene, M. and R. Wheeldon. Navigating the world-wideWeb. In M. Levene and A. Poulovassilis, editors, Web Dynamics. Springer-Verlag, 2003.

Lieberman, H.: Letizia: An Agent that Assists Web Browsing. In Chris S. Mellish, editor, Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95), Montreal, Quebec, Canada, Morgan Kaufmann publishers Inc.: San Mateo, CA, USA (1995) 924-929

Lieberman, H., Neil W. Van Dyke, Adriana Santarosa Vivacqua: Let's browse: a collaborative browsing agent. Knowl.-Based Syst. 12(8): 427-431 (1999)

Lieberman, H., Fry, C., and Weitzman, L., Exploring the Web with reconnaissance agents. Communications of the ACM, 2001.44(8): p. 69-75.

Linden, G., Hanks, S. and Lesh, N., 1997, Interactive assessment of user preference models: The automated travel assistant. User Modeling, June.

Lutz, C. Adding Numbers to the SHIQ Description Logic-First Results. In Proceedings of the Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR2002). Morgan Kaufman, 2002.

Lutz. C. Description Logics with Concrete Domains - A Survey. In Philippe Balbiani, Nobu-Yuki Suzuki, Frank Wolter, and Michael Zakharyaschev, editors, Advances in Modal Logics Volume 4. King's College Publications, 2003.

Lynch, K.: The Image of the City, MIT Press, Cambridge, MA. (1960)

Lynch, C., 1997, Searching the internet. Scientific American, 276 (3), 52-56

MacEachren A.M., Kraak, M.J. (2001): Research challenges in geovisualization. In: Cartography and Geographic Information Science, 28 (2001) 1, pages 3-12.

Madria, SK., Bhowmick, SS., Ng, WK., Lim, EP.: Research Issues in Web Data Mining. In: Proceedings of the 1st International Conference on Data Warehousing and Knowledge Discovery, (1999) 303-312

Malaka, R., R. Porzel, A. Zipf, V. Chandrasekhara (2000). Integration of smart components for building your personal mobile guide. In: Malaka, R. (ed.), Artificial Intelligence in Mobile Systems – AIMS2000, Workshop in conjunction with ECAI 2000, Berlin, Germany, August 22, 2000, Workshop Notes, pp. 22-26.

Malaka, R. and Zipf, A. (2000): DEEP MAP - Challenging IT research in the framework of a tourist information system. In: Fesenmaier, D. Klein, S. and Buhalis, D. (Eds.): Information and Communication Technologies in Tourism 2000. Proceedings of ENTER 2000, Barcelona.

Spain. Springer Computer Science, Wien, New York. 15-27.

Malczewski, J. GIS and Multicriteria Decision Analysis, 1999, 392 pages, ISBN 0-471-32944-4. New York

Mehta, B., Claudia Niederée and Avaré Stewart, 2005 Towards Cross-system Personalization in the Parallel session entitled "Future directions in personalization and context-aware access",3rd International Conference on Universal Access in Human-Computer Interaction (UAHCI 2005), USA, 22-27 July 2005

Melville P., Raymond J. Mooney, and Ramadass Nagarajan. Content-boosted collaborative filtering for improved recommendations. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02), July 2002.

Milic-Frayling, N., Rachel Jones , Kerry Rodden , Gavin Smyth , Alan Blackwell , Ralph Sommerer, Smartback: supporting users in back navigation, Proceedings of the 13th conference on World Wide Web, May 17-20, 2004, New York, NY, USA

Miller, George A. ``WordNet: a lexical database for English." In: Communications of the ACM 38 (11), November 1995, pp. 39 - 41.

Mitchell, T. 1997. Machine Learning. McGraw-Hill.

Mladenic, D., "Personal Webwatcher: Design and implementation, " Technical report ijs-dp-7472, School of Computer Science, Carnegie-Mellon University, Pittsburgh, USA, 1996.

Mladenic, D., (1999) Machine learning used by Personal WebWatcher in Proceedings of ACAI-99 Workshop on Machine Learning and Intelligent Agents, Chania, Crete, July 5-16, 1999.

Mobasher, B. WebPersonalizer A Server-Side Recommender System Based on Web Usage Mining Technical Report #01-004, DePaul University, School of CTI. This is an extended version of an earlier paper that appeared in the Proceedings of the 9th Workshop on Information Technologies and Systems (WITS'99), Charlotte, NC, Dec. 1999.

B. Mobasher, R. Cooley, and J. Srivastava, "Automatic Personalization Based on Web Usage Mining," Comm. ACM, vol. 43, no. 8, Aug. 2000a, pp. 142-151.

Mobasher, B., H. Dai, T. Luo, Y. Sung, and J. Zhu, 2000b Integrating Web Usage and Content Mining for More Effective Personalization. In Proceedings of the International Conference on E-Commerce and Web Technologies (ECWeb2000), September 2000, Greenwich, UK.

Mobasher, B., H. Dai, T. Luo, and M. Nakagawa. Using sequential and non-sequential patterns in predictive Web usage mining tasks. In 2002 IEEE International Conference on Data Mining (ICDM'02), 2002.

Mobasher, B. (2004). Web usage mining and personalization. Chapter in Practical Handbook of Internet Computing, Munindar P. Singh (ed.), CRC Press

Moloney T, Lea A C, Kowalchuk C, 1993, Manufacturing and packaged goods, in Profiting from a Geographical Information System, GIS World Books Inc, Fort Collins pp. 105-129:

Mountrakis, G., P. Agouris, A. Stefanidis, (2005). Adaptable User Profiles for Intelligent Geospatial Queries. Transactions in GIS, Vol. 9, No. 4, pp. 561-583.

Mulvenna, M., S. Anand, and A. Buchner. Personalization on the net using Web mining. CACM, 43(8):122--125, 2000.

Nafaâ J. 2005 A Multi-Agent System for On-The-Fly Web Map Generation and Spatial Conflict Resolution, Université Laval

Niederée, C.J., Stewart, A., Mehta, B., Hemmje, M.: A multi-dimensional, unified user model for cross system personalization. In: Proceedings of Advanced Visual Interfaces International Working Conference (AVI 2004) - Workshop on Environments for Personalized Information Access, Gallipoli (Lecce), Italy, May 2004. (2004)

Novak, Joseph D. The Theory Underlying Concept Maps and How To Construct Them. Cornell University, 2003, available at http://cmap.coginst.uwf.edu/info/

Nyce, J.M. and P. Kahn, editors. From Memex to Hypertext: Vannevar Bush and the Mind's Machine. Academic Press, San Diego, Ca., 1991.

Oard, D.W., Kim, J.: Modelling Information Content Using Observable Behaviour. In Proceedings of the 64 Annual Meeting of the American Society for Information Science and Technology, USA, (2001) 38-45.

Oberle, D., Bettina Berendt, Andreas Hotho, Jorge Gonzalez Conceptual User Tracking In Ernestina Menasalvas Ruiz and Javier Segovia and Piotr S. Szczepaniak, Advances in Web Intelligence, First International Atlantic Web Intelligence Conference, AWIC 2003, Madrid, Spain, May 5-6, 2003, Proceedings, volume 2663 of Lecture Notes in Artificial Intelligence, pp. 142-154. Springer, 2003. ISBN: 3-540-40124-5

O'Conner M.. and J. Herlocker. Clustering items for collaborative filtering. In Proceedings of the ACM SIGIR Workshop on Recommender Systems, Berkeley, CA, August 1999

Open GIS Consortium, Inc. (OGC) (2004). Open GIS Web Map Server Interface Implementation Specification (Revision 1.3.0). Wayland, Massachusetts: Open GIS Consortium, Inc. http://www.opengeospatial.org/specs/

Open GIS Consortium, Inc. (OGC) (2005). Open GIS Web Feature Service Implementation Specification (Revision 1.1.0). Wayland, Massachusetts: Open GIS Consortium, Inc. http://www.opengeospatial.org/specs/

Padmanabhan V. and J. Mogul.: 1996 Using Predictive Prefetching to Improve World Wide Web Latency. In Proc. of the ACM SIGCOMM '96 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communication, pages 22-36, July 1996.

Pazzani, M., Muramatsu, J. and Billsus,. D. Syskill Webert: Identifying Interesting Web Sites. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, pp 54-61, Portland, 1996

Pazzani, M., Billsus, D.: Learning and Revising User Profiles: The identification of interesting Web sites. Machine Learning 27, (1997) 313-331

Pazzani, M. and Billsus, D. (2002). Adaptive Web Site Agents. Journal of Agents and Multiagent systems, 5(2). 205-218

Perkowitz M. and O. Etzioni: Adaptive Web Sites: an AI Challenge. In Proceedings of the 15th International Joint Conference on AI (IJCAI-97), Nagoya, Japan, August 23-29, 1997

Perkowitz M. and O. Etzioni: "Adaptive Web Pages: Automatically Synthesizing Web Pages", Proc. of 16th Natl' Conf. on Artificial Intelligence and 10th Conf. on Innovative Applications of Artificial Intelligence (AAAI/IAAI '98), Madison July 1998.

Perkowitz M. and O. Etzioni, Towards adaptive Web sites: conceptual framework and case study, Artificial Intelligence 118(1–2) (2000) 245–275

Perugini, S. and Ramakrishnan, N. 2002 Personalizing Interactions with Information Systems, in Advances in Computers, edited by M. V. Zelkowitz, Academic Press, 9,57, 323--382.

Perugini S., Marcos André Gonçalves, Edward A. Fox: A Connection-Centric Survey of Recommender Systems Research. The Computing Research Repository (CoRR) CoRR cs.IR/0205059: (2002) Available (verified 01/09/2004) at http://arxiv.org/PS_cache/cs/pdf/0205/0205059.pdf

Pierrakos, D. , G. Paliouras, C. Papatheodorou, and C.D. Spyropoulos. Web usage mining as a tool for personalization: A survey. User Modeling and User-Adapted Interaction, 13:311-372, 2003.

Piaget, J. & Inhelder, B. (1969). The Psychology of the Child. NY: Basic Books.

Pirolli, P. L., Pitkow, J. E.: 1999 Distribution of surfers' paths through the World Wide Web: empirical characterization. World Wide Web. 1999, 2 (1-2); 29-45.

Pitkow J. and Peter Pirolli.: 1999 Mining longest repeating subsequence to predict wolrd wide Web surfing. In Second USENIX Symposium on Internet Technologies and Systems, Boulder, C0.

Pitkow, J., H. Schutze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, and T. Breuel. Personalized search. Communications of the ACM, 45(9):50--55, 2002.

Pohl W. (1999) Logic-Based Representation and Reasoning for User Modeling Shell Systems. In User Modeling and User-Adapted Interaction 9(3): 217-283.

Poslad, S., Laamanen, H., Malaka, R., Nick, A., Buckle, P. and Zipf, A. (2001): CRUMPET: Creation of User-Friendly Mobile Services Personalised for Tourism. In: Proceedings of: 3G 2001 - Second Int. Conf. on 3G Mobile Communication Technologies. 26-29.03.2001. London. UK.

Putz, S.: 1993 World Map Viewer, HTTP Service at Xerox PARC.

Putz, S.: 1994 Interactive information services using World-Wide Web hypertext. The First International Conference on the World-Wide Web (WWW'95), May 25-27, 1994, CERN, Geneva. Published in Computer Networks and ISDN Systems Vol. 27 (1994), Number 2, November 1994, pp 273-280.

Resnick, P. and H.R. Varian. Recommender systems. Communications of the ACM, 40(3):56--58, 1997

RDF Core Working Group, W3C Semantic Web Activity, W3C. 2004 Resource Description Framework (RDF), W3C Recommendation, 10 February 2004, http://www.w3.org/TR/rdf-primer/

Ricci, F., D. Blaas, N. Mirzadeh, A. Venturini, H. Werthner "Intelligent Query Management for Travel Products Selection" ENTER 2002 Conference, Innsbruck; 22-25 January, 2002.

Rich E. (1979). "User Modeling via Stereotypes". International Journal of Cognitive Science, 3, 329-354

Rich, E. (1989). Stereotypes and User Modeling. In: Kobsa, A. & Wahlster, W. (eds.) User Models in Dialog Systems, 35-51. Berlin, Heidelberg, Springer.

Riecken D.: Introduction: personalized views of personalization. Commun. ACM 43(8): 27-28 (2000)

Riesbeck, C.K. and R.C. Schank, Inside Case-Based Reasoning, Lawrence Erlbaum, Hillsdale, NJ, 1989

Rinner C. and M. Raubal (2004) Personalized Multi-Criteria Decision Strategies in Location-Based Decision Support. Journal of Geographic Information Sciences 10(2): 149-156.

Rodriguez, A., Egenhofer, M. and Rugg, R. (1999) Assessing Semantic Similarity Among Geospatial Feature Class Definitions. in: A. Vckovski, K. Brassel, and H.-J. Schek, (Eds.), Interoperating Geographic Information Systems - Second International Conference, INTEROP'99, Zurich, Switzerland. Lecture Notes in Computer Science 1580, pp. 1-16, Springer-Verlag, Berlin.

Rocchio, J. 1971. Relevance feedback in information retrieval. Prentice Hall, Ing. Englewood Cliffs, New Jersey.

Roure, D.D., W. Hall, S. Reich, G. Hill, A. Pikrakis, and M. Stairmand. MEMOIR -- an open framework for enhanced navigation of distributed information. Information Processing and Management, 37:53 -- 74, 2001.

Rucker, J. and J.P. Marcos, "Siteseer: Personalized Navigation for the Web," Comm. ACM, Vol. 40, No. 3, Mar. 1997, pp. 73-- 75.

Saaty, T. L., 1980, The Analytic Hierarchy Process, McGraw-Hill, New-York.

Sadalla, E.K., Burroughs, W.J., Staplin, L.J.: Reference Points in Spatial Cognition. Journal of Experimental Psychology: Human Learning and Memory, 5, (1980) 516-528

Sarukkai, R.: 2000 Link Prediction and Path Analysis Using Markov Chains, Computer Networks, vol. 33, nos. 1-6, pp. 377-386, June 2000.

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. Analysis of recommender algorithms for e-commerce. In Proceedings of the 2nd ACM E-Commerce Conference (EC'00), Minneapolis, MN, October 2000a.

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. Application of dimensionality reduction in recommender systems-a case study. In Proceedings of the WebKDD 2000 Workshop at the ACM-SIGKDD Conference on Knowledge Discovery in Databases (KDD'00), August 2000b.

Schafer, J. B., Konstan, J., and Riedl, J., 1999, Recommender systems in e-commerce. In Proceedings of the ACM Conference on Electronic Commerce, pp. 158-166.

Schild, K. A correspondence theory for terminological logics: preliminary report. In Proceedings of IJCAI-91, 12th International Joint Conference on Artificial Intelligence, pages 466--471, Sidney, Australia, 1991.

Schild, K. Terminological cycles and the propositional ¯-calculus. In Jon Doyle, Erik Sandewall, and Piero Torasso, editors, Proc. 4th International Conference on Principles of Knowledge Representation and Reasoning, Bonn (Germany), May 1994. Morgan Kaufmann Publishers.

Schiller, J. and Voisard, A. Location-based services, Morgan Kaufmann, San Francisco, California, 2004. 250p.

Schmidt-Belz, B., Stefan P., Nick, A. and Zipf, A. (2002): Personalized and Location-based Mobile Tourism Services. Workshop on "Mobile Tourism Support Systems". 17.09.2002. Pisa. in conjunction with Mobile HCI '02 (Fourth International Symposium on Human Computer Interaction with Mobil Devices). 18.-20.09.2002.

Schmidt-Belz, B., Zipf, A., Poslad, S., Laamen, H. (2003): Location-based mobile tourist services - first user experiences. ENTER 2003. Int. Congress on Tourism and Communications Technologies. Helsinki. Finland. (Springer Computer Science. Heidelberg, Berlin).

Scime, A. and Kerschberg, L., WebSifter: An Ontological Web-Mining Agent for E-Business. in Proceedings of the 9th IFIP 2.6 Working Conference on Database Semantics (DS-9): Semantic Issues in e-Commerce Systems, (Hong Kong, 25-28 April 2001), IFIP, 200-214.

Shahabi, C., Chen, Y.: Web Information Personalization: Challenges and Approaches, The 3nd International Workshop on Databases in Networked Information Systems (DNIS 2003), Aizu-Wakamatsu, Japan, September (2003) 5-15

Shavlik, J. and Towell, G., 1989, An approach to combining explanation-based and neural learning algorithms, Connection Science, 1(3), 233-255.

Shearin, S. and Lieberman, H., 2001, Intelligent profiling by example. In Proceedings of the International Conference on Intelligent User Interfaces (IUI 2001), Santa Fe, NM, pp. 145-152.

Silva, M. J. Bruno Martins, Marcirio Chaves, Nuno Cardoso, "Adding Geographic Scopes to Web Resources" Ana Paula Afonso ACM SIGIR 2004 Workshop on Geographic Information

Retrieval, Sheffield, UK, June 2004.

Sinner, A., Thomas Kleemann, Andreas von Hessling: Semantic User Profiles and their Applications in a Mobile Environment. In Artificial Intelligence in Mobile Systems 2004 (AIMS 2004), September 2004.

Sirin E. and Bijan Parsia. Pellet: An OWL DL Reasoner. Position Paper, In 2004 International Workshop on Description Logics (DL2004), Whistler, Canada, 2004.

Spaccapietra, S., C. Parent, N. Cullot, C. Vangenot On Using Conceptual Modeling for Ontologies In Proceedings of the International Workshop on the Intelligent Networked and Mobile Systems, Ontologies for Networked Systems (ONS) track, located at the 5th International Conference on Web Information Systems Engineering WISE 2004, Brisbane, Australia, 22nd-24th November 2004

Spaccapietra, S., Shijun Yu, Lina Al-Jadir. Somebody, sometime, somewhere, something International Workshop on Ubiquitous Data Management (UDM2005), In Memoriam Prof. Yahiko Kambayashi, April 4, 2005, Tokyo, Japan, In conjunction with IEEE ICDE 2005, IEEE proceedings UDM'05 (to appear)

Srikant, R., Agrawal, R.: Mining generalized association rules. Future Generation Computer Systems 13 (1997) 161—180

Steel, N., Huppert, F.A., McWilliams, B., and Melzer D., 2004. Physical and cognitive function. Avialable at: www.ifs.org.uk/elsa/report03/ch7.pdf

Stewart, A., Claudia Niederée, Bhaskar Mehta 2004 State of the Art in User Modelling for Personalization in Content, Service and Interaction.Technical Report, NSF/DELOS Report on Personalization, 2004.

Suchman, L. A. Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge University Press, 1987.

Sugiyama, K., Hatano, K., Yoshikawa, M.: Adaptive Web Search Based on User Profile Constructed without Any Effort from Users, WWW 2004, New York, USA (2004) 675-684

Tan S., Pearl, J.: 1994 Qualitative decision theory, Proceedings of the twelfth national conference on Artificial intelligence (vol. 2), p.928-933, October 1994, Seattle, Washington, United States

Terveen, L., Hill, W., Amento, B., McDonald, D., Creter, J., PHOAKS: A System for Sharing Recommendations, Communications of the ACM, pp. 59-62, Vol. 40, no. 3, March 1997.

Terveen, L.G and Hill, W. Human-Computer Collaboration in Recommender Systems, in Carroll, J. (ed.), HCI in the New Millennium (2001) , Addison Wesley

Tezuka, T., Ryong Lee, Hiroki Takakura, Yahiko Kambayashi, “Web-Based Inference Rules for Processing Conceptual Geographical Relationships”, The 1st International Workshop on Web Geographical Information Systems, in conjunction with the 2nd International Conference on Web Information Systems Engineering, IEEE Computer Society Press, pp.375-384, 2001.

Tezuka, T., Lee, R., Takakura H., and Kambayashi, Y., Integrated Model for a Region-Specific Search Systems and Its Implementation, Proceedings of International Conference on Internet Information Retrieval 2003, pp. 243-248, Koyang, Korea, Oct. 2003.

Tezuka, T. and Tanaka, K., 2004, Temporal and spatial attribute extraction from Web documents and time-specific regional Web search system. In Proceedings of the 4th International Workshop on Web and Wireless Geographical Information Systems, Internet Information Retrieval Research Center, Hankuk Aviation University, South Korea, pp. 78-89.

Thill, J.C., 1999, Spatial Multicriteria Decision Making and Analysis: A Geographic Information Sciences Approach, Ashgate, Brookfield, VT, 1999, 377 pp.

Thomson, L. 2005 A Standard Framework for Web Personalization, 1st International Workshop on Innovations In Web Infrastructure (IWI 2005) Co-Located with the 14th World-Wide Web Conference (WWW2005), Japan, May 2005.

Thorndyke, P.W., Hayes-Roth, B.: Differences in Spatial Knowledge Acquired from Maps and Navigation, Cognitive Psychology, 14, (1982) 560-589.

Tobler, W.R., 1970, A computer model simulating urban growth in the Detroit Region. Economic Geography, 46, 234-240

Towle, B., & Quinn, C. (2000). Knowledge-based recommender systems using explicit user models. Knowledge-based Electronic Markets: Paper from AAAI, Technical Report WS-00-04, 74-77.

Tversky, B.: Cognitive Maps, Cognitive Collages, and Spatial Mental Models. In Frank, A.U. and Campari, I. (eds.), Spatial Information Theory, Proceedings of COSIT'93, Lecture Notes in Computer Science 716. Springer-Verlag, Berlin, Germany, (1993) 14-24

Ungar, L.H., Foster, D.P. (1998). Clustering Methods for Collaborative Filtering. Proceedings of the 1998 Workshop on Recommender Systems, pp. 114-129.

Wache, H., T. Voegele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann and S. Huebner (2001) Ontology-Based Integration of Information - A Survey of Existing Approaches Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing Seattle, USA, August 4-5, 2001. (108-118)

Watkins,C,1989.Learning from Delayed Rewards, Thesis,University of Cambidge, England.

Weakliam, J., D. Lynch, J. Doyle, M. Zhou, E. MacAoidh, M. Bertolotto, D. Wilson, 2005a Managing Spatial Knowledge for Mobile Personalized Applications, Proceedings KES 2005 (9th International Conference on Knowledge-based Intelligent Information and Engineering Systems), Melbourne, Australia, September 2005, Lecture Notes in Artificial Intelligence 3684, Springer-Verlag, pp. 329-335.

Weakliam, J., M. Bertolotto, D. Wilson, 2005b Implicit Interaction Profiling for Recommending Spatial Content, Proceedings ACMGIS'05, Bremen, Germany, November 4-5, 2005, ACM Press (in press).

Web Ontology Working Group, W3C Semantic Web Activity, W3C. 2004 OWL Web

Ontology Language, W3C Recommendation, 10 February 2004, http://www.w3.org/2001/sw/WebOnt/

Webb, G., Pazzani M.J., Billsus, D.: Machine Learning for User Modelling. User Modelling and User–Adapted Interaction, (11): (2001) 19–29

Weißenberg, N., Agnès Voisard, Rüdiger Gartmann: Using ontologies in personalized mobile applications. GIS 2004: 2-11

Weiser, M. The computer for the 21st century. Scientific American, 265(3):66--75, February 1991.

Wellman, B., Computer Networks As Social Networks Science 293, 14, Sept 2001: 2031-2034

Werthner, H. and S. Klein. 1999. Information Technology and Tourism - A Challenging Relationship (Springer Computer Science.) Springer.

Winter, S.; Tomko, M., in press: Translating the Web Semantics of Georeferences. In: Taniar, D.; Rahayu, W. (Eds.), Web Semantics and Ontology, Idea Publishing, Hershey, PA 2005

Worboys, M., 1996, Metrics and topologies for geographic space. In Advances in GIS Research II, Kraak, M.J., Molenaar, M. (eds.), Taylor and Francis, pp. 365-375

Worboys, M. F.: Nearness relations in environmental space. International Journal of Geographical Information Science 15(7): 633-651 (2001)

Xie, C. Bo Huang, Richard Tay. A Graph Theory Based Algorithm for Blocking a Fleeing Vehicle with Probabilistic Behavior 13rd International Conference on Geoinformatics, August 17-19, 2005, Toronto, Canada

Yager, R.R. On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making, IEEE Transaction on Systems, Man and Cybernetics 18 (1988) 183--190.

Yao, X. and Thill, J.C. 2005. How Far Is Too Far - A Statistical Approach To Proximity Modeling. Transactions in GIS. Vol. 9(2): 157-178.

Yao, Y.: Measuring Retrieval Effectiveness based on User Preference of Documents, Journal of the American Society for Information Science, 46(2), (1995)133-145

Yu, S., Stefano Spaccapietra, Nadine Cullot, Marie-Aude Aufaure User Profiles in Location-based Services: Make Humans More Nomadic and Personalised In Proc. of the International Workshop on Next Generation Geospatial Information, NG2I 2003, Cambridge, USA, October 2003,

Zhang, B., Seo, Y.: Personalized Web-Document Filtering Using Reinforcement Learning. Applied Artificial Intelligence 15(6), (2001) 665-685

Zipf, A. (2002): User-Adaptive Maps for Location-Based Services (LBS) for Tourism. In: K. Woeber, A. Frew, M. Hitz (eds.), Proc. of the 9th Int. Conf. for Information and Communication Technologies in Tourism, ENTER 2002. Innsbruck, Austria. Springer Computer Science. Heidelberg, Berlin.