**Project FP7-216923**

**SMARTMUSEUM**

**Cultural Heritage Knowledge Exchange Platform**

# Deliverable D2.2

# SMARTMUSEUM Report describing methods for dynamic user profile creation

Workpackage WP2 – Self adaptive user profile management

Task T2.3 - Developing theoretical solution for dynamic user profile creation and modification

| **Version** | 1.00 | **Date** | March 11, 2009 | **Classification** | Public | **Status** | Draft |
|---|---|---|---|---|---|---|---|


**Abstract**

SMARTMUSEUM (Cultural Heritage Knowledge Exchange Platform) is a Research and Development project sponsored under the Europeans Commission's 7th Framework. The overall objective of the project is to develop a platform for innovative services enhancing on-site personalized access to digital cultural heritage through adaptive and privacy preserving user profiling. Using on-site knowledge databases, global digital libraries and visitors' experiential knowledge, the platform makes possible the creation of innovative multilingual services for increasing interaction between visitors and cultural heritage objects in a future smart museum environment, taking full benefit of digitized cultural information.

The main objective of this deliverable is to describe a theoretical framework for management of dynamic user profiles.



**Authors**

SMARTMUSEUM Consortium (www.smartmuseum.eu)

Responsible Editors: KTH, ELIKO, TKK



**Grant Agreement Number:** FP7-216923
**Acronym:** SMARTMUSEUM
**Project title:** Cultural Heritage Knowledge Exchange Platform

## Executive Summary


The purpose of this deliverable D2.2 is introducing a theoretical framework for dynamic user profile management and its related operations specifically creation and modification of user profiles.

Existing methods for user behaviour monitoring and formalisation are studied and a state of art in broader picture of contextualization and personalization is given, taken into account the domain of the project, cultural heritage domain.

Within the proposed framework, we introduce two approaches, mainly contributing mechanisms used for learning and building user profiles as well as mechanisms for enabling personalized recommendation and filtering on behalf of museum users. Our approaches take into account the adaptivity and dynamisms of user profiles, as suggested user profile qualities. This enables us to measure how effective profiling framework works.

The introduced framework allowing us to create and learn dynamic user profiles for SMARTMUSUEM project. We have used the generic user profile structure introduced in D2.1 for generating generic and dynamic user profile structures, while by utilizing collaborative filtering we learn and use profiled prefernces of users of the platform. As collaborative filtering techniques are the main foundation of the work being done, we introduce recommendation as a substantial part of the proposed framework.

Introduction ........ 5
SMARTMUSEUM Project ........ 5
Deliverable purpose, scope and context ........ 5
Audience ........ 5
Background ........ 6
Introduction to personalization, user modelling and profiling ........ 6
Profiles Presentation ........ 7
Profiles Qualities ........ 8
Personalization, Recommendation and Contextualization in Museum Domain ........ 9
Overview of existing work on personalization in cultural heritage ........ 9
State-of-art approaches to user profile learning and construction ........ 10
Data Mining ........ 11
Conceptual Clustering ........ 12
Naive Bayes ........ 14
Bayesian Network ........ 15
Neighbourhood based methods ........ 16
Self-Organizing Map ........ 16
Nearest Neighbour ........ 17
Latent Semantic Indexing / Analysis ........ 17
State-of-art on recommenders and recommendation techniques ........ 18
Collaborative filtering ........ 19
User-based Approaches ........ 19
Item-based Approaches ........ 19
Model-based Approaches ........ 19
Similarity Measures ........ 20
Content-based filtering ........ 21
Hybrid filtering ........ 21
Dynamic Profile Operations ........ 22
Statistical Learning and Construction of User Profiles ........ 22
Modelling and Measuring User Preferences and Interests ........ 22
Selecting and Combining Initial Profile Learning Techniques ........ 23
Association rule mining ........ 23
Frequent closed item set mining ........ 24
Simple Collaborative Filtering: Utilizing Statistical User Profiles ........ 24
Dataset for the initial modelling ........ 24
Recalculation and incremental learning of user's preference ........ 25
Input data format for algorithms ........ 25
Output data buffering for algorithms ........ 26
Analysis ........ 26
Extended Collaborative Filtering: Utilizing Semantics and Social Trust ........ 28
Extending Recommender Systems with Semantics: A User-Item Ontological Model ........ 28
User Ontology ........ 28
Item Ontology ........ 29
Enriching Recommendations using Distributed Social Trust ........ 30

Recommendation prediction process......30
Rating a New Item Event......30
Generating a New Recommendation Event......31
Extending Recommender Systems with Socio-Semantic Trust: An Evaluation......32
Conclusion......33
General Conclusions......33
Detailed Conclusions: Selection of Algorithms and Mechanisms......34
Works Cited......35

# Introduction

The purpose of this section is to introduce the:

- SMARTMUSEUM Project
- Purpose, scope and context of this deliverable
- Intended audience for the deliverable

## SMARTMUSEUM Project

SMARTMUSEUM (Cultural Heritage Knowledge Exchange Platform) is a Research and Development project sponsored under the Europeans Commission's 7th Framework. The overall objective of the project is to develop a platform for innovative services enhancing on-site personalised access to digital cultural heritage through adaptive and privacy preserving user profiling. Using on-site knowledge databases, global digital libraries and visitors' experiential knowledge, the platform makes possible the creation of innovative multilingual services for increasing interaction between visitors and cultural heritage objects in a future smart museum environment, taking full benefit of digitized cultural information.

The SMARTMUSEUM project supports achieving the following general goals:

• Lowering costs of on-site access to digital cultural heritage content,
• Improving structured, user behaviour and preference dependent on-site access to the vast repository of cultural heritage,
• Improving the individual and shared experiences people receive from cultural and scientific resources,
• Bringing personalised cultural experience closer to non-expert communities,
• Making real reuse of personal experiences related to cultural heritage access for a variety of interest groups.

## Deliverable purpose, scope and context

The purpose of this deliverable D2.2 is to present: 1) A survey of existing approaches to dynamic user profile management and 2) A framework of implemented techniques for user profile management.

## Audience

The intended audience includes:

- Primarily SMARTMUSEUM Partners involved in developing the user profile-related operations
- Project partners involved in SMARTMUSEUM WP2

# Background

In this section a brief information about the main elements of this deliverable is given. We define personalization and usage profiling, and then the presentation aspects of profiles as well as qualities considered for profiling the usage in museum domain are discussed.

## Introduction to personalization, user modelling and profiling

User profiling takes its roots in human studies. A user profile is defined as gathering of raw personal material about the user, according to (Koch., May 8th, 2005). User profiles gather and present cognitive skills, abilities, preferences and interaction histories with the system (Gauch, et al., 2007). According to (Middleton, et al., 2004), User profiling is either knowledge-based or behavior-based. Knowledge-based approaches construct static models of users and match users to the closest model. Questionnaires and special forms are used to gather this user knowledge. Behavior-based methods consider the behavior as a modeling base, commonly by utilizing machine-learning techniques (Middleton, et al., 2004) (Bloedorn, et al., 1996) to discover useful patterns in the behavior. Behavioral gathering and logging is used in order to obtain the data necessary to detect and extract patterns, according to (Kobsa, 2001).

Personalization systems are based on user profiles. 45 personalization systems are listed by Pretschner (Pretschner, 1999), according to Gauch (Gauch, et al., 2007). Personalization (Sieg, et al., 2007) can be provided to user by customizing the content or the visualization of the system based on user’s profile (Weibelzahl, 2003).

Personalization techniques fall into two main categories (Dolog, et al., 2003).

First category are based mostly on adapting user interfaces and content selection and rendering based on the user's performance and behavior in a certain domain. These techniques are collectively referred to as Adaptive hypermedia techniques (Brusilovsky, 2001).

Another category of techniques is based on cognitive patterns (such as interests, preferences, likes, dislikes, and goals) a user have. This information is mostly stored as user profiles and stored at some kind of profiling or modeling server (Kobsa, 2007). These methods are known as filtering and recommendation techniques. They filter resources based on features (mostly metadata) extracted and gathered from a resource or according to ratings (generally weights) of a user of similar profile, according to Dolog and Nejdl (Dolog, et al., 2003).

## Profiles Presentation

Emergence of Semantic Web, created new possibilities for profile-driven personalization.

Ontologies, at the heart of Semantic Web technologies, are used for two major purposes (Dolog, et al., 2003). Mainly, they are utilized to formalize domain concepts which allow describing constraints for generation or selection of resource contents belonging the domain the user is keen towards, as well as being used to formalize the user model or profile ontology (Dokoohaki, et al., 2008) that helps making decision which resources to be adapted (for instance, shown or not shown) to the user. Ontologies along with reasoning create formalization that boost personalization decision making mechanisms, according to Dolog and Nejdl (Dolog, et al., 2007).

Ontological user profiles are becoming widely adopted. EU-project Spice constructs a multi-ontological approach, across the domain of mobile communications, from which one of the most important axis is user profile ontology (Sutterer, et al., 2008). Within the domain of digital cultural heritage, CHIP project[1] is definitely a significant stake holder. Considerable amount of research attention has been payed on semantically formalizing the user domain (Wang, et al., 2007) (Aroyo, et al., 2007), as well as personalization of information retrieval. Hybrid ontological user models are consumed to learn, gather, store and use personal user data, according to which semantically-enriched art works are recommended to, during both on-line and on-site visit to exhibition.

We have considered utilizing hybrid user models (Dokoohaki, et al., 2008), which incorporate a semantic presentation of personal information about users as well as incorporating notions of trust, privacy and ranking for items the user has interest towards. For a complete drescription of the formalization of SMARTMUSEUM user profile, reader is advised to read D2.1 .

1. [1]The CHIP (Cultural Heritage Information Presentation) project. http://www.chip-project.org/

## Profiles Qualities

The quality of user profiles is a key to success of profile-based systems. From the user's point of view, there are two potential problems, according to (Cetintemel, et al., 2000). The most important one is precision problem: If a large proportion of the items that the system sends to a user are irrelevant, then the precision of the system and its performance becomes a question to user. Qualities can be characterized and studied to measure the precision of the system. We have considered two main qualities for user profiles in SMARTMUSEUM scenario. First is self-adaptivity and second is dynamicity.

We have proposed a self-adaptive profiling framework which represents user interests as a dynamic set of profile records. As pointed out previously, we presented the structure of these profiles in D2.1. The structure of the profiles were chosen to be generic, in order to adapt themselves to dynamic changes in environment, which in our case, these changes will come from chronical changes in users' interests as well as the number of records, due to CRUD (create, read, update and delete) operations that take place on user profiled material.

This creates a notion of flexibility. This flexibility enables our approach to trade off effectiveness and efficiency, which in turn, enables it to be tuned based on the requirements/characteristics of our target environment. As a matter of fact, effective profile management requires techniques for representing data items and profiles, assessing the relevance of the profiles to data items, and updating the profiles based on user feedback.

As stated previously, generic format of the records in which we store user data creates a felxible structure which allows dynamic information about users to be stored and retrieved. Dynamic aspect of the profiling framework allows preferences to be updated and changed regularly to create a more precise model of the user's cognitive patterns and interests.

The framework of dynamic operation, discussed later on point outs the dynamicity of the user profiles and it shows how effective they maintain the dynamicness of the content stored and retrieved from the user profiles.

# Personalization, Recommendation and Contextualization in Museum Domain

This section gives a thorough survey of state-of-art research in contextualization and personalization. The focus has been given to works within similar domain.

## Overview of existing work on personalization in cultural heritage

Within the domain of cultural heritage different approaches can be distinguished, some of them are more content orientated and some context oriented depending on the task to be solved.

Several projects are initiated to provide the best content in a certain context for users, like for providing personalized information in museums (Hyvönen, et al., 2005) (Hyvönen, 2007) (Sparacino, 2004), location-aware tourist guides (Abowd, et al.) (Cheverst, et al.) (Fink, et al., 2002) (Lam, et al., 2007) for handling personalized customer relationships (Kobsa, 2001), for providing a news program (Billsus, et al.) (Singh, et al.) for managing context information in mobile devices (Korpipää et al., 2003), etc.

As we could generalize the main directions of the research are the context and the content.

When the term 'context-aware' was first introduced (Schilit, et al., 1994) (Schilit, et al., 1994) then it defined the context as location, emphasizing nearby people and objects and changes related to those objects. (Dey, et al., 2001) define context more broadly and context refers to any information that characterizes a situation related to the interaction between humans, applications, and the surrounding environment. Context is typically the location, identity, and state of people, groups, and computational and physical objects. A survey of context-aware applications is given by (Chen, et al., 2000) and a solution for supporting context aware application prototyping by the Context Toolkit is presented (Day et al., 2001). To provide context specific services on the one hand context information can be used, but on the other hand user's previous activities or a profile must be taken into account. To provide information in a personalized manner, personalized systems observe a user' behaviour and, based thereon, make generalizations and predictions about them (Fink & Kobsa, 2002). Some methods and tools must be used to generalize and predict user's behaviour and also to offer personalized content.

Content on the web is available in different forms, as text documents, audio and video files, images, etc. Therefore the content can be defined content broadly as the stuff in your Web site (Rosenfeld, et al., 1998). on the web all kind of cultural content is available. Question is how to find out most relevant information sources and to provide them for the user in an appropriate manner.

The use of standardized approach makes the access to the cultural context easier and therefore some content models for semantic cultural portals have been developed (Hyvönen, 2007).

Several projects have been initiated using the semantic web that are oriented to provide personalized and context-aware information for the museum visitors. Like the project MUSEUMFINLAND (Hyvönen, 2007; Hyvönen et al., 2005). In this project by sharing a set of ontologies, it is possible to make collections semantically interoperable, and provide the museum visitors with intelligent contentbased search and browsing services to the global collection base.

Another example of a personalized cultural semantic portal is the CHIP project (Aroyo, et al., 2007) In the project an interactive approach is used to collect data about museum visitors in terms of their interests and preferences about artefacts from the Rijksmuseum collection. This data is stored in user profiles used further to recommend routes through the museum and to guide the users towards artefacts related to their interests and preferences. The overall goal of the project is to explore different users' characteristics and personalize users' museum experiences within the Rijksmuseum virtual and physical collections (Aroyo et al., 2007). And also a tourist context-aware guiding system, iJADE FreeWalker, which uses Semantic Web technologies, integrates GPS, ontology and agent technologies to support location awareness for providing the precise navigation and classify the tourist information for the users (Lam & Lee, 2007).

## State-of-art approaches to user profile learning and construction

In this section (Kirt, 2008) the methods used in different content and context based applications are introduced. Depending on the problem to be solved there are used a number of methods.

We have picked out five most referred methods that can also be used in our project.

The five groups of methods and their applications are as follows:

- **Data Mining**–mining association rules between sets of items in a large database of customer transactions (Agrawal, et al., 1993) (Agrawal, et al., 1994), data mining for web personalization (Mobasher, et al., 2000);

- **Conceptual Clustering**–a web document clustering algorithm-WebDCC (Godoy, et al., 2006), adaptive Web sites (Perkowitz, et al.), and user' interests estimator (Kim, et al., 2008).

- **Naive Bayes**–identifying interesting Web sites to a user (Billsus, et al., 1999), a model for news story classification, named News Dude (Billsus, et al., 1999), developing context-aware mobile application (Korpipää, et al., 2003) (Korpipää, et al., 2003);

- **Bayesian Network**– user-centered approach for computational storytelling (Sparacino, 2003; 2008);

Neighbourhood based methods–for example, Self-Organizing Map–a user activity detection (Laerhoven, et al., 2000) and location detection (Schmidt, et al.).

But usually all kind of methods are combined to get the best results. Like a project constructing an appropriate user profile uses three different machine learning algorithms in the model building process: Bayesian network, clustering and rule-based models (Sugiyama, et al., 2004).

### Data Mining

(Hand, et al., 2001) have defined data mining as follows: Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.

The relationships and summaries derived through a data mining exercise are often referred to as models or patterns.

In data mining, association rule learners are used to discover elements that co-occur frequently within a data set consisting of multiple independent selections of elements (such as purchasing transactions), and to discover rules, such as implication or correlation, which relate co-occurring elements.

Agrawal, Imielinski and Swami (1993) have introduced the problem of mining association rules between sets of items in a large database of customer transactions. Each transaction consists of items purchased by a customer in a visit. We are interested in finding those rules that have: Minimum transactional support s — the union of items in the consequent and antecedent of the rule is present in a minimum of s % of transactions in the database. Minimum confidence c — at least c % of transactions in the database that satisfy the antecedent of the rule also satisfies the consequent of the rule.

Commonly the confidence threshold value is chosen. But the fixed confidence threshold has little basis in statistics, since some sets may exceed it simply by random coincidence (thereby defeating the goal of finding meaningful correlations), and some meaningful associations may occur in the data without reaching the threshold (Zaki, 1999). However, in practice it does eliminate vast numbers of insignificant sets.

An example of data mining, often called the market basket analysis, relates to its use in retail sales (Agrawal, Imielinski, & Swami, 1993). If a clothing store records the purchases of customers, a data mining system could identify those customers who favour silk shirts over cotton ones. Although some explanations of relationships may be difficult, taking advantage of it is easier.

The ability to track user browsing behaviour down to individual mouse clicks has brought the vendor and end customer closer than ever before (Mobasher, Cooley, & Srivastava, 2000). It is now possible for vendors to personalize their product messages for individual customers on a massive scale, a phenomenon referred to as "mass customization." The usage mining tasks can involve the discovery of association rules.

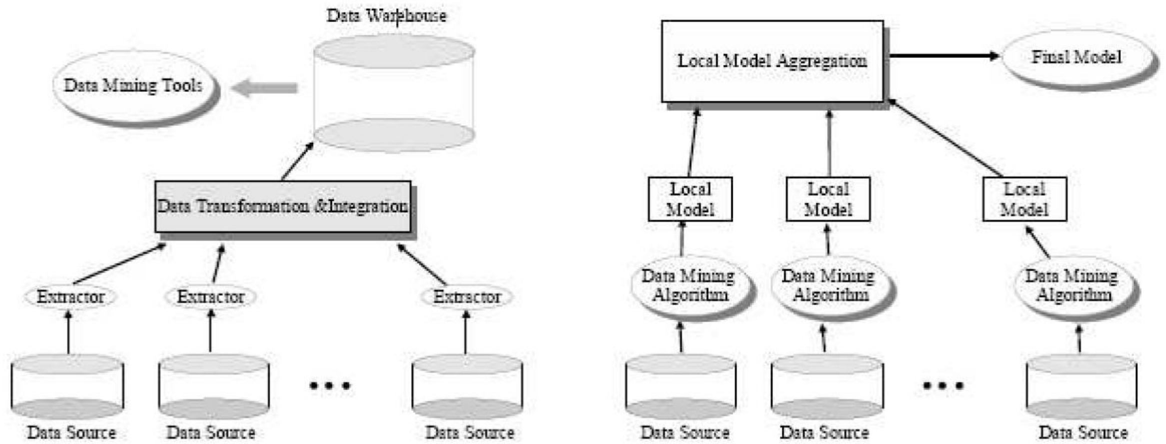


**Figure 1 - A data warehouse architecture and distributed data mining framework (Park, Kargupta, 2003)**

An overview of mining systems for distributed applications like data mining primitives in ad hoc wireless networks of mobile devices like PDAs, cell phones and wearable computers (Park, et al., 2003). Zaki (1999) surveys the state of the art in parallel and distributed association-rule-mining algorithms and concludes that such systems are still in their infancy, and a lot of exciting work remains to be done in system design, implementation, and deployment.

In the Smartmuseum project the association rules can be used to identify user preferences and to give him some advices which objects in the museum might be interesting. The generation of association rules is very fast and it can be applied to a large database. But it is noted that the selection of reasonable rules is rather difficult.

### Conceptual Clustering

Clustering is the classification of objects into groups by using sufficient number of variables. The classified objects can be connected to predefined categories that characterize the object in the best way. As (Rosch, 1978) pointed out the purpose of the categorization system of an organism is to provide maximum information about the environment with the least cognitive effort. Categorization must use features that maximize similarity between the samples in the cluster and at the same time separate different clusters as well as possible.

There are a variety of theories specifying the best way to categorize objects. They range from the classical theory (Laurence, et al., 1999) or the prototype theory (Rosch, 1978) to a dynamic approach as the situated simulation theory (Barsalou, 2003). The classical theory holds that the category membership is defined as a set of necessary and sufficient features. If an object does not exhibit the necessary and sufficient features, then it does not belong to the category. Whether an object belongs to a category in the prototype theory is decided by using its similarity with a prototype (Rosch, 1978).

To measure similarity with the prototype, a set of features is needed. Some of the features have higher weight than others and the membership of the sample in a category is determined by measuring similarity between a sample and the category representation. The situated simulation theory states that conceptual representation and the category membership is highly contextualized and dynamical and is dependant on the previous experience and the current situation (Barsalou, 2003).

Conceptual clustering is a machine learning task for unsupervised classification of objects (Michalski, 1980). A most known example of the conceptual clustering approach is a conceptual clustering system COBWEB (Fisher, 1987). COBWEB is an incremental system for hierarchical conceptual clustering. The system carries out a hill-climbing search through a space of hierarchical classification schemes using operators that enable bidirectional travel through this space (Fisher, 1987). The COBWEB algorithm is an incremental method and is connected to the observation that most of human learning can be viewed as a gradual process of concept accusation and human ability for incorporating knowledge from new experiences into existent concept structures.

There are two ways of clustering (Kim Chan, 2008).

First, agglomerative (bottom-up) hierarchical clustering algorithms initially put every object in its own cluster and then repeatedly merge similar clusters together, resulting in a tree shape structure that contains clustering information on many different levels.

Second, divisive (top-down) hierarchical clustering algorithms are similar to agglomerative ones, except that initially all objects start in one cluster which is repeatedly split. The conceptual clustering belongs to the latter one.

There is a number of different approaches to the conceptual clustering.

(Godoy, et al., 2006) have presented a document clustering algorithm, named WebDCC (Web DocumentConceptual Clustering) that carries out incremental, unsupervised concept learning over Web documents in order to acquire user profiles. (Perkowitz, et al.) have built adaptive Web sites: sites that automatically improve their organization and presentation by learning from visitor access patterns.

(Kim, et al., 2008) are extracting a continuum of general (long-term) to specific (short-term) interests of a user. The proposed approach is to learn a user interest hierarchy (UIH) from a set of web pages visited by a user. They have developed a divisive hierarchical clustering (DHC) algorithm to group words (topics) into a hierarchy where more general interests are represented by a larger set of words.

Their divisive algorithm does not necessarily generate binary splits and uses a minimum cluster size as one of the stopping criteria. However, instead of using category utility to determine if child clusters are generated, we use a graph-based method and a different similarity function. (Kim, et al., 2008)

(Singh, et al.) are attempting to model the user's interests for ludic news reading behaviour, i.e., general reading of the news with basic themes of interest that may change slowly over time. It is adaptive approach in the sense that once the initial profile or interest hierarchy is built, the leaf categories of the hierarchy are updated after each session with the explicit feedback of the user. This adaptive phase continues the learning and can also model the "drift" in user's interests over time.

Another approach for building visual ontologies is to hierarchically cluster a visual training data set. Such clustering methods have been used earlier for ontology discovery in textual and numerical data (Clerkin, et al., 2001) The main advantages of the method of conceptual clustering are that it uses unsupervised learning method and it is incremental and can adapt new items into a conceptual structure. But there is not known how such methods behave in the case of large data sets and whether they are applicable if the number of variables is rather large end their mutual relations are complicated.

### Naive Bayes

A graphical approach to classification is called the naive Bayes model, in which conditional independence assumptions are used to simplify the model structure (Bishop, 2006). The structure of the naive Bayes classifier is a Directed Acyclic Graph (DAG) that represents the conditional probabilities with arrows between variables, and independence of variables if the arrow is missing (Korpipää, et al., 2003).

As the method is a very simple estimation of likelihood it is widely used in application of personalization and context awareness. For example, for identifying which Web sites would be interesting to a user a naïve Bayesian classifier is used (Pazzani, et al., 1997) and a model for news story classification, named News Dude (Billsus, et al., 1999). the naive Bayesian networks are applied to classify nine contexts of a mobile device user in their normal daily activities (Korinpää et al., 2003).

Naive Bayes framework is feasible for context recognition. In real world conditions, the recognition accuracy using leave-one-out cross validation was 87% of true positives and 95%

of true negatives, averaged over nine eight-minute scenarios containing 17 segments of different lengths and nine different contexts (Korinpää, al. 2003).

## Bayesian Network

Bayesian networks are graphical structures for representing the probabilistic relationships among a large number of variables and doing probabilistic inference with those variables (Neapolitan, 2003). The Bayesian network shares some common characteristics with the conceptual clustering; both of them are data driven methods and use unsupervised learning strategy.

Bayesian networks are directed acyclic graphs (DAG), where the nodes are random variables, and certain independence assumptions hold. To specify the probability distribution of a Bayesian network, one must give the prior probabilities of all root nodes (nodes with no predecessors) and the conditional probabilities of all nonroot nodes given all possible combinations of their direct predecessors. Bayesian networks allow one to calculate the conditional probabilities of the nodes in the network given that the values of some of the nodes have been observed (Charniak, 1991).

A good example of the use of the Bayesian network is given by Sparacino (2003). The Bayesian network is used to identify user's preferences for accessing for real-time sensor-driven multimedia audiovisual stories.

The network is used in the Museum Wearable (Sparacino 2003), a device which delivers an audiovisual narration interactively in time and space to the visitor as a function of the estimated visitor type.

This device is a museum guide which in real time evaluates the visitor's preferences by observing his/her path and length of stops along the museum's exhibit space, and selects content from a set of available movie clips, audio, and animations. The process consists of three phases–first modelling the user's type, then estimating his/her interest profile, and subsequently the selection of content.

The museum wearable uses a Bayesian network to provide a real time estimate of visitor types: a greedy type, who wants to know and see as much as possible, and does not have a time constraint, a busy type who just wants to get an overview of the principal items in the exhibit, and see little of everything, and the selective type, who wants to see and know in depth only about a few preferred items "short", "long" stop duration, or "skip" object (see Figure 2).

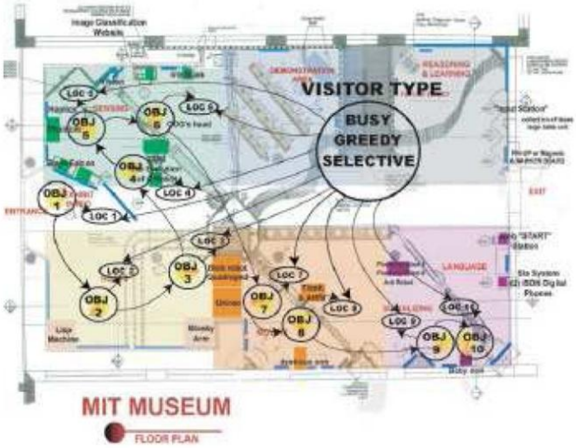


**Figure 2 - Three visitors type (Sparacino, 2003)**

For example, some objects can be very interesting or less interesting, according either to the opinion of the curator, or the public's preferences, and this information can be encoded in the Bayesian network. Moreover, it would be possible, theoretically, at the end of each week, to take the posterior probabilities for all objects and all visitors, and reintroduce them as priors of the Bayesian network the following week (Sparacino, 2008).

The Bayesian network is rather widely used in different applications. The unsupervised learning approach is the main advantage of this algorithm and also its way to handle missing variables. On the other hand the creation of Bayesian network is rather resource consuming and in case of large databases it might not to be applicable.

### Neighbourhood based methods

### Self-Organizing Map

The self-organizing map SOM (Kohonen, 1982) (Kohonen, 2000) is a neural network that uses an unsupervised learning algorithm. It means that there is no prior information presented to the algorithm on how input and output are connected. The SOM can be used for clustering and visualizing multidimensional data and for reducing the dimensionality of the data by plotting them in a two-dimensional output grid, also called a map.

For the visualization of the self-organizing map, a Unified distance matrix (U-matrix) is often used. A light colour corresponds to a small distance between two map units and a dark colour represents a bigger difference between the map units. The points on the output map that lie in the light area belong to the same group or cluster, while the dark area shows the borders between the clusters.

The SOM is widely used in several applications, because of its simplicity. An example of the use of the SOM in a user activity detection (Laerhoven, et al., 2000) and location detection (Schmidt et al., 1999).

### Nearest Neighbour

Even more close to the ideas of prototype theory, is the method called the nearest neighbour. It is similar to the SOM and quite often used to identify the cluster borders on the SOM. The method of nearest neighbour is for classifying objects based on closest training examples in the feature space. The k-nearest neighbour is used in a user activity detection (Laerhoven, et al., 2000).

(Dai, et al., 2003) have argued by using the k-Nearest-Neighbour (kNN) approach and compare the record with the historical records of other users in order to find the top k users who have similar tastes or interests. The mapping of a visitor record to its neighbourhood could be based on similarity in ratings of items, access to similar content or pages, or purchase of similar items. The identified neighbourhood is then used to recommend items not already accessed or purchased by the active user. The advantage of this approach over purely content-based approaches which rely on content similarity in item-to-item comparisons is that it can capture "pragmatic" relationships among items based on how their intended use or based on similar tastes of the users.

The nearest neighbours are used in the application of News Dude (Billsus, et al., 1999). For identification of stories that the user already knows the nearest neighbour algorithm (NN) is used. The NN algorithm simply stores all its training examples, in our case rated news stories, in memory. In order to classify a new, unlabeled instance, the algorithm compares it to all stored instances given some defined similarity measure, and determines the "nearest neighbour" or the k nearest neighbours. The class label assigned to the new instance can then be derived from the class labels of the nearest neighbours.

Advantages of the SOM are that it is unsupervised method and there is no need to predefine categorisation rules. The main disadvantage of the SOM is that it is rather resource consuming. The k-NN algorithm is a very simple one and can be used as a preliminary method to discover the most similar behaviour of the other visitors.

### Latent Semantic Indexing / Analysis

In this section the latent semantic indexing/ analysis is introduced. As the idea of the latent semantic indexing is very simple therefore it is widely used in information retrieval systems. Information retrieval is process of finding documents or other information that is meets users needs or interests (Meadow, et al., 2007).

Deerwester et. al. (1990) have proposed a method to deal with the vocabulary problem in humancomputer interaction using the singular value decomposition and called their approach Latent Semantic Indexing (LSI), known also as Latent Semantic Analysis (LSA).

The LSI is based on the singular value decomposition (SVD). A practical algorithm to solve the Singular Value Decomposition was introduced by G. Golub and W. Kahan in 1965. The SVD is as a decomposition technique for calculating the singular values, pseudo-inverse and rank of a matrix. The SVD is a linear algebra technique for reducing dimensionality of the data.The SVD is used to find out higher-order regularities in the original term-by-document matrix. The new dimensions in the reduced space are linear combinations of the original dimensions and are better representation of documents and queries, because noise is reduced.

## State-of-art on recommenders and recommendation techniques

Recommender systems are among the most popular personalization systems and solutions available. Recommendation problem is commonly defined as the problem of estimating ratings for the items that have not been seen by a user (Adomavicius, et al., 2005). Such estimation is usually based on the ratings given by this user to other items and on some other information. Once we can estimate ratings for the yet unrated items, we can recommend to the user the item(s) with the highest estimated rating(s).

Items can be of any type (Candillier, et al., 2009), such as films, music, books, web pages, online news, jokes, restaurants and in SMARTMUSEUM case, cultural heritage artifacts. Recommender systems help users to find such items of interest based on some information about their historical preferences.

An inventory of existing recommender systems and their application domain which have been developed, is listed by (Nageswara Rao & Talwar, 2008).

Recommender systems fall into three major categories:

- **Collaborative filtering**: The user will be recommended items similar to the ones the user preferred in the past.

- **Content-based filtering**: The user will be recommended items that people with similar tastes and preferences liked in the past.

- **Hybrid filtering**: These methods combine collaborative and content-based methods.

The recommendation system designer must select which strategy is most appropriate given a particular problem.

For example, if little item ranking is available then a collaborative filtering approach is unlikely to be well suited to the problem. If item descriptions are not available then content-based filtering approaches will be not suitable.

### Collaborative filtering

In collaborative filtering, the input to the system is a set of user ratings on items. Users can be compared based upon their shared appreciation of items, creating the notion of user neighbourhoods.

Similarly, items can be compared based upon the shared appreciation of users, rendering the notion of item neighbourhoods. The item rating for a given user can then be predicted based upon the ratings given in her user neighbourhood and the item neighbourhood. We can distinguish three main approaches: *user-based*, *item-based* and *model-based* approaches.

#### User-based Approaches

For user-based approaches (Resnick, et al., 1994) (Shardanand, et al., 1995), the prediction of a user rating on an item is based on the ratings, on that item, of the nearest neighbours.

So a similarity measure between users needs to be defined before a set of nearest neighbours is selected. Also, a method for combining the ratings of those neighbours on the target item needs to be chosen.

#### Item-based Approaches

Recently, there has been a rising interest in the use of item-based approaches (Sarwar, et al., 2001) (Karypis, 2001) (Linden, et al., 2003) (Deshpande, et al., 2004). Given a similarity measure between items, such approaches first define item neighbourhoods. The predicted rating for a user on an item is then derived by using the ratings of the user on the neighbours of the target item.

#### Model-based Approaches

The general idea in Model-based Approaches is to derive a model of the data off-line in order to predict on-line ratings as fast as possible.

The first type of models that have been proposed consist of grouping users using clustering and then predicting a user rating on an item using only the ratings of the users that belong to the same cluster.

Bayesian models have also been proposed to model dependencies between items. The clustering of items has been studied extensively (e.g. (Ungar, et al., 1998) (O'Conner, et al., 1999). Also, models based on association rules have been studied by (Sarwar et al., 2000) and (Lin et al., 2002).

Probabilistic clustering algorithms have also been used in order to allow users to belong, at some level, to different groups (Pennock, et al. 2000) (Kleinberg, et al., 2004). And hierarchies of clusters have been proposed, so that if a given cluster of users does not have an opinion on a particular item, then the super-cluster can be considered (Kelleher, et al., 2003).

A numerical example of the goal of collaborative filtering algorithms is given below (Liv,2008) .

Let us have the following table of people and ratings. Cells of the matrix should be interpreted as *Leo* rated *Pulp Fiction* with a score *2*. Let us have a matrix of such ratings, where several (often most) of the ratings are missing (people have not seen the movie, book, objects etc.) – our goal is to predict (within a continuous scale) how would (for example) Leo rate Tarzan (considering all the information about his and other past behaviors).

| | **Pulp fiction** | **Star Wars** | **Tarzan** | **Kevade** |
|---|---|---|---|---|
| **Leo** | 2 | 4 | **?** | 1 |
| **Andres** | 5 | **?** | 2 | 1 |
| **Ants** | 2 | 2 | 2 | 5 |

## Similarity Measures

The similarity defined between users or items is crucial in collaborative filtering. The first one proposed in (Resnick et al., 1994) is the Pearson correlation.

It corresponds to the Cosine of deviations from the mean. Simple Cosine or Manhattan similarities are also traditional ones. For these similarity measures, only the set of attributes in common between two vectors are considered. Thus two vectors may be completely similar even if they only share one appreciation on one attribute, according to (Candillier, et al., 2009).

### Content-based filtering

Content-based recommendation (Candillier, et al., 2009) systems recommend an item to a user based upon a description of the item and a profile of the user's interests.

Content-based recommendation systems share in common a means for describing the items that may be recommended, a means for creating a profile of the user that describes the types of items the user likes, and a means of comparing items to the user profile to determine what to recommend. Item descriptors can be the genre of a film or the location of a restaurant, depending upon the type of item being recommended. Finally, items that have a high degree of proximity to a given user's preferences would be recommended.

A User profile may be built implicitly from the user's preferences for items, by searching for commonalities in liked and disliked item descriptions, based upon her past actions or explicitly through questionnaires about her preferences for the item descriptions.

A User model may be learned implicitly by an automatic learning method, using item descriptions as input to a supervised learning algorithm, and producing user appreciations of items as output.

Preferences indicate a relationship between a given user and data. In recommender system research, a preference must be both machine codable and carry useful information for making recommendations. For example, in the field of cinematography, the monadic preference "I like Jackie Chan as an actor" can be coded as a high score for films with this actor.

### Hybrid filtering

In the case of hybrid filtering, both types of information, collaborative and content-based, are exploited (Candillier, et al., 2009). these technologies can be combined in various ways that make use of both user appreciations of items and their descriptor-based preferences. Other sources of data like social and demographic data about users can also be used.

# Dynamic Profile Operations

In this section the framework for dynamic profile operations are introduced.

We begin by learning and constructing statistical notions of user profiles and how to dynamically update the learned profiles. Based on the statistical learning methodology, a collaborative filtering engine is implemented and introduced, as a proof of concept prototype. Next, an extended approach is introduced which takes into account semantics and trust for extending a CF recommender system. The approach, being taken is introduced along with description of the system being implemented as a proof of concept. Finally, two approaches for learning and building user interests based on nearby geographical locations are described and introduced.

## Statistical Learning and Construction of User Profiles

In this section (Liv, 2008) we introduce the approach under taken for learning the preferences of the users using pure statistical methodologies.

First the modeling and measuring of the preferences of the user interests are introduced and then approaches chosen for composition of learning of interest indicators are introduced and the incremental learning of the user profiles are discussed.

### Modelling and Measuring User Preferences and Interests

The generic profile structure introduced, allows incorporation of rankings and ratings into the data about the items user has interest towards. As a numerical value is chosen for modeling and measuring interest degree of the users, the levels are taken from the range of [-1,1].

Existing suggestion, follows preference and behavior levels suggested in D1.1 for modeling usage behaviour:

- *1: strong like, manually input.*
- *0.7: user fetches a lot information about the object.*
- *0.3: user fetches basic information about the object.*
- *-0.01: user visits room with the object, shows no interest.*
- *-1: strong dislike, manual input.*

According to standard classification of scales, such data would be something between *ordinal* (ordered, but differences between values are strongly subjective) and *interval* (ordered, constant scale, no "natural" zero) data. Therefore, it would be reasonable to test several approaches to model the behaviour – both method-wise and data-wise.

For collaborative filtering approaches (based on previous knowledge about rating, the attempt is made to predect an unknown rating as a float data type) we will look at the same data as continous values. For data mining techniques we can consider the dataset as categorical (but ordered) data. Several algorithms prefer binary input data for better computational efficiency, therefore even categorical data should be preprocessed the following either way:

- *User's rating (float) should be derived into five binary attributes: rated_1= TRUE/FALSE, rated_0_7= TRUE/FALSE, rated_0_3= TRUE/FALSE, rated_m0_01= TRUE/FALSE, rated_m1=TRUE/FALSE or even (if they are treated as categorical data anyway) rated_strong_like= TRUE/FALSE, fetched_a_lot_info= TRUE/FALSE, fetched_basic_info=TRUE/FALSE etc. Additional data consistency check could be applied to such derived attributes (e.g. only one TRUE value per block);*

- *User's rating (float) should be derived into two blocks of binary attributes: rated_higher_than_0= TRUE/FALSE, rated_higher_than_0_01= TRUE/FALSE, rated_higher_than_0_3= TRUE/FALSE etc and rated_lower_than_1= TRUE/FALSE, rated_lower_than_0_7, rated_lower_than_0_7=TRUE_FALSE etc. Such derivation scheme allows several true values per block;*

- *Only one binary value can be chosen to present the "liking" according to some threshold.*

### Selecting and Combining Initial Profile Learning Techniques

If we are considering the problem of making recommendation based on the past behavior, we can only look at it as a unsupervised learning problem (there is no other training data than the input data itself and we are looking for connections between the input data in a completely unsupervised manner).

Mining for **association rules**, **frequent closed itemsets** (basically a two-mode clustering approach) and **collaborative filtering** approaches account for more than 80% of the solutions typically used for this kind of problems. There are several interesting methods to transform the "recommendation space" to a one or two-dimensional "proximity space".

#### Association rule mining

Association rule mining (Agrawal, et al., 1993) (Agrawal, et al., 1994) is an accepted methodology, several companies with strong market cap (Microsoft, Oracle, IBM, SPSS) are already shipping their proprietary implementation of the technique in several off-the-shelf products, bringing it closer to end-user consumption of exploratory data analysis methods. A typical example of association rules is a book recommendation in Amazon e-store, where customers receive recommendations based on similarities in their and other users' behaviour in the past (Customers who like product X, tend to like also product Y). Typical measures of such rules are support (number of rows supporting such rule) and confidency (out of all rows liking product X, how many of those like also product Y).

A numerical example of an association rule:

*People, who like objects X and Y, tend to like object Z.*

### Frequent closed item set mining

Frequent closed itemset mining (Pasquier, et al., 1999) (Zaki, et al., 2002) is an enhancement of frequent itemset mining for pruding redundant patterns based on the closure property of the pattern combination lattice. Frequent closed itemset mining produces sets of objects which are commonly co-bought, co-chosen, co-used, e.g. products X,Y,Z are bought together with frequency F. Pruning using the property of closure would eliminate all subset combination of the same set of items, if they have an equal frequency. Counting the customers who bought together products X,Y,Z is trivial, but the set X,Y,Z is initially unknown and brute-forcing would result an exponential explosion of compexity (2(n-1) combinations should be tested).

A numerical example of a frequent closed itemset result:

*Objects X, Y and Z tend to be liked together.*

#### Simple Collaborative Filtering: Utilizing Statistical User Profiles

Approach taken here for filtering the items, is Slope one (Lemiere et al, 2005). Slope one is the simplest form of non-trivial item-based collaborative filtering based on ratings.

### Dataset for the initial modelling

As currently no data exist about objects in the museum exists, neither there are datasets available online that would fit exactly to our needs. A recommendation system dataset from GroupLens[2] was chosen for an algorithm testing, it assumably shares the distribution of

2 GroupLens is a research lab in the Department of Computer Science and Engineering at the University of Minnesota (http://www.grouplens.org/)

ratings (Power law[3]/Zipf[4]). For all three different approaches the dataset need a (different!) preprocessing. However this is no fallacy as finally only one approach will be probably chosen and recommendations are not recalculated real time. Ratings are originally on the scale of 1 to 10, which have to be mapped for the suggested rating values for this project.

### Recalculation and incremental learning of user's preference

Currently it is suggested to recalculate the recommendations once a week, which allows satisfactory computational time even with modest server configurations. In practice, interval of recalculations could be as low as an hour. Any recalculation scheme with shorter intervals should incorporate some incremental learning heuristics of decomposed parts of the system of ratings.

### Input data format for algorithms

Collaborative filtering algorithm takes data in (rows of) information triples <user,object,rating>, i.e. a list format for a data matrix, especially suitable for sparse matrices.

**Example:**

*1,1,0.7*
*1,2,-1*
*1,3,1*
*2,1,1*

Association rule and frequent closed itemset mining algorithms require the data to be differently preprocessed.

**Example.**

*(each row is a user, however the order of users is not important for rule generation)*
*(TRUE/FALSE information is not additionally provided, only true value attributes exist in the row)*
*(row /user 1) fetched_a_lot_of_information[1] strong_dislike[2] strong_like[3]*
*(row/user 2) strong_like[1]*

---

3 http://en.wikipedia.org/wiki/Power_law

4 http://en.wikipedia.org/wiki/Zipf%27s_law

## Output data buffering for algorithms

The output of algorithms should be stored for rapid real-time access – if an object is viewed, instantly a recommendation should be given to the user.

Association rules and frequent closed itemsets can be stored as triples in database *<object1,object2,pattern_importance/support/confidency>*, so the database query of "sort the recommendations for object1 based on the pattern importance value" could be performed on database server. Collaborative filtering algorithm could give top objects with best predicted value, but at current state should be tested more for convenient real-time access.

Recalculation procedures can take a lot of computational power, but the output should always be stored in a database to enable fast queries of the updated behavioral model at any time.

## Analysis

User profile summary described by D1.1 and D2.1, is sufficient to perform all the behavioral modeling of user profiles described in this document.

The choice of the previous algorithms comes down to total running time of the algorithm and value added. Mining of closed frequent itemsets was chosen for the following reasons:

- Fastest running time;
- Symmetric recommendation is quicker and more reasonable in practice. First intuition might suggest, that we might be interested in making the difference between “If person liked object X, he/she also likes object Y” and “If person liked object Y, he/she also likes object X”. However, with large datasets, the result could also be hundreds of thousands of combinations like “If person liked object X1 and X2 and X3 and X4 and X5, he/she also likes object X6”, “If person liked object X1 and X2 and X3 and X4 and X6, he/she also likes object X5”, “If person liked object X1 and X2 and X3 and X5 and X6, he/she also likes object X4” etc.
- Current collaboration filtering algorithm estimated well over 48 hours of calculation of scores and a possibility of a enormous object1-vs-object2 (entity-to-entity) distance matrix generation in the database.

A walkthrough from the use case and data storage point of view would be as follows.
A browsing and walking session has to be linked with specific **userID**. Therefore, if we visit a specific object (with **objectID**) and either rank it manually or tag with other scores as described by (Tammet, 2007), we have to store (at least) the following data to meet the requirements of our algorithms:

**<userID,objectID,ranking>**

(in a corresponding rating database table, which is described in data definition language in technical supplementary materials) Knowledge about the preferences of users is built up on such information periodically with the following scheme:

- Preprocessing of the database to meet the data mining algorithm's input (included in the technical supplementary material);
- Postprocessing of the found patterns to meet the operational level recommendations to be uploaded in the handheld interface or browsed via web browser (included in the technical supplementary material);

As a result, a database of patterns of objects which people tend to like (together) is stored and available to use in the following way:

1. If one **objectID** is known (e.g. at the point of interaction with the object), it is possible to query the patterns with strongest support by pattern frequency where that object is. The result of a pattern is a list of other objectID's this person should like.
2. Similarly, it is possible to query beforehand (e.g. at the point of entry in the museum) all the objects recommendations or for example the recommended route for visit – one needs to query all the patterns where all the objects in that museum exist (all the patterns are already precalculated).
3. With nested queries and database joins, it is possible (without extra algorithmic data analysis; natively with database queries) to link the **objectID** with museum and other preferences, therefore filtering the result in several ways, e.g.
    1. Exclude everything that the person has explicitly stated not liking or rank higher (currently the ranking is based on the patterns support by frequency) objects with the properties/attributes that the person has explicitly stated to like;
    2. It is possible to query and also to filter out cross-patterns between museums (and other places of interests). The **objectID** has to be linked with the museum and a query could show all the patterns, where there are objects that this person has liked, and filter out only objects which are in museum **X**.

A prototype of an item recommender and user profile matchmaker with aforementioned

description has been implemented and is available as a component (Liv, 2008).

## Extended Collaborative Filtering: Utilizing Semantics and Social Trust

Collaborative Filtering seems to be the dominant technique for Recommender systems which predicts new item's rating for a user based on the rating from other users. Traditional CF algorithms are based on similarity of users profile and use that as a weight to make recommendations. It is difficult to compute similarity measure between users when ratings are sparse. Therefore, the profile similarity on its own is not effective.

In the context of web-based social networks, where users are connected to their friends and share their opinions, we need to consider additional factors which can measure and interpret the similarity between user profiles. In social networks, users are able to express their feelings about their relationships like how much they trust others (Golbeck, 2007). Trust provides us with information about the people we should share content with and accept content from. In fact, users would prefer to receive recommendations from whom they trust. Therefore, we need trustful recommenders who have a "history of making reliable recommendations" (O'Donovan, et al., 2005). In fact, recommendations are made only based on the ratings given by users who are directly trusted by the current user or indirectly trusted by another trusted user (Massa, et al., 2004).

We (Zarghami, et al., 2009) propose a recommender system that can be extended with trustworthiness of the users. In order to model the trust, we create an ontology that extends FOAF (Friend-of-a-Friend) vocabulary (Brickley, et al., November 2007) and adopts the structure presented in (Dokoohaki, et al., 2007) (Dokoohaki, et al., 2008).

### Extending Recommender Systems with Semantics: A User-Item Ontological Model

We present a model with an ontological structure to build trust relationships between users considering all types of item, accessed by an unique URI in heterogeneous networks. Therefore, our model is composed of two ontologies for users and items. As a result, their corresponding concepts can be extended by other ontologies.

#### User Ontology

Trust ontology can be considered as a model for representing trust relationships in semantic web and social networks (Dokoohaki, et al., 2008) (Dokoohaki, et al., 2007).We take advantage of the trust model presented by (Dokoohaki, et al., 2007) to express the trust between users who extend FOAF Agent. In our model, trust value is computed based on users' ratings to different items possibly in different contexts. As shown in Figure 3, we create an instance of Relationship class between two users for whom trust value is computed. The users are specified as Truster and Trustee and their trust value and subject is assigned as a Main property (Dokoohaki, et al., 2008) to the instance defined earlier. In addition, we propose a measure called T-index, similar to H-index (Braun, et al., 2006) to show the minimum number of trust relationships with a trust value higher than the same minimum. T-index depicts how close a user is to the center of its cluster. We assign it as a Main property [5] for the Relationship instance. Therefore, the ontology is able to keep top-n trustee who has the highest trust value and T-index. In this work, T-index will range from 0 to 100 and trust value is in the rang of 0 to 1. For instance, we have a user whose T-index is 20 meaning that he has at least 20 trust relationships with trust value more than 0.20. We will show how T-index can be applied and we also define RankRelation class for associating a user to an item by a rank value. This class is used to keep the track of rated items by a user called user profile.

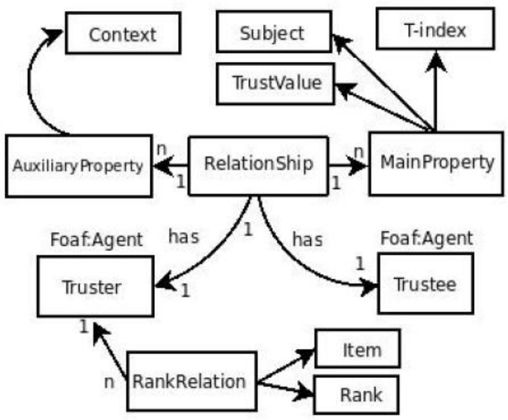


**Figure 3 - User Ontology Model**

## Item Ontology

We develop an ontology for item's knowledge domain which can be extended by all other ontologies. We introduce a new class called TopTrusteeList to store users who have no trust relationship yet but might have similar interests. We assign it to an individual item in order to achieve list of users who rate the item. The list of raters is sorted by their T-Index. In a real scenario, these TopTrusteeList can be implemented by Distributed Hash Tables (DHT) with unique URI as their keys.

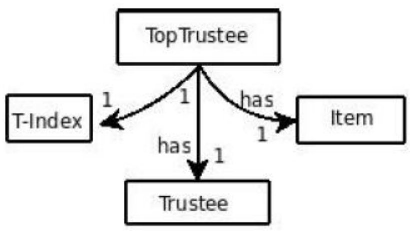


**Figure 4 - Item Ontology Model**

Similarity between two users and how much this similarity is related to the trust in between them, is discussed in (Ziegler, et al., 2005). Golbeck also shows through the analysis of data in FilmTrust (Golbeck, 2006) that there is a correlation between similarity and trust. They define a recommenders rating to be correct if the difference between it and the users rating is less than a threshold value. Two computational models of trust are proposed by O'Donovan and Smyth (O'Donovan, et al., 2005) as profile-level and profile-item-level based on the past behavior of user profiles They set a producer's rating as correct if the difference between it and consumer's rating is less than a predefined threshold. (Lathia, et al., 2008) propose a trust formula based on difference of a user's rating and its recommender's rating to their common item(s). Hence, as the distance between their rating value increases, trust decreases linearly. Assuming we have two users $u_a$ and $u_b$, who rated items $i_1$ and $i_2$ by value $r_1$ and $r_2$, respectively. Thus, trust between $u_a$ and $u_b$ is formalized, according to (Lathia, et al., 2008):

$$T(u_a, u_b) = 1 - \frac{\sum_{i=1}^{n}(r_{u_a,i_i} - r_{u_b,i_i})}{r_{max} * n}$$

Formula calculates the total differences between user's and its recommender's rating values over n historical ratings multiplied by the maximum value in each rating scale (i.e., 5). In order to improve prediction accuracy, suggest to transpose the rating values of truster based on its past experience with trustee considering number of user's ratings which are the same, lower or higher than its recommender's ratings, according to (Lathia, et al., 2008):

$$tr(r) = \frac{(r - 1 * lower_r) + (r * same_r) + (r + 1 * higher_r)}{lower_r + same_r + higher_r}$$

The transpose of rating values, $tr(r(u_b,i_i))$ can be used instead of $r(u_b,i_i)$ to guarantee that a user and its recommenders have the same scale for rating.

### Recommendation prediction process

As mentioned earlier, we keep top-n neighbors list of a user in an ontological structure based on their mutual trust value. The list is updated upon two events: rating a new item and making a new recommendation. If the events lead up to some changes in top-n neighbors list of a user, then T-index value is recalculated and updated in all TopTrusteeLists which contain the same user.

#### Rating a New Item Event

When a user rates a new item, we compute its trust with all users in the item's TopTrusteeList who do not exist in its current top-n neighbors list but might be trustworthy users potentially. We also update trust values between the user and its top-n neighbors. Ultimately, we form a new top-n neighbors list by selecting the most trustworthy users from the union of its preceding neighbors and the potential trustees. Algorithm 1 describes the process of rating a new item. T-index and TopTrusteeList of rated item are required to be updated along with the process. More details are described later.

### Generating a New Recommendation Event

In order to recommend a new item to a user, it explores its neighborhood to find the users who rate the item, with a mutual trust value more than a predefined threshold and a maximum distance value from the user. The details of user's traversals through the trust graph will be discussed. As a result, if a user discovers another user out of its top-n neighbors list, who has more mutual trust value than its current neighbors, it will be added to top-n neighbors list. In case the list exceeds the maximum number of trustees called n, the neighbor with the lowest trust value will be removed.

**Algorithm 1** Rating a new item

```
upon event ⟨RATE NEW ITEM | Truster,Item⟩
    UpdateTrust(neighbours)
    pointers ← calculateTrustwithTrustees(Item,m)
    neighbours ← selectTopnTrustee(neighbours,pointers)
    T − index ← updateTindex(neighbours)
    updateTopTrusteeList(Item) //if it exceeds, remove the last one
    for all ⟨ratedItems⟩ in Truster do
        updateTindex(rateedItems)
    end for
end event
```

As mentioned above, when a user rate a new item, as shown in algorithm 1 , it calls updateTopTrusteeList() function. The function compares the new user's T-index with other users of the item's TopTrusteeList. Then if it exists in top m T-index, it will be added to the TopTrusteeList. In this case, if TopTrusteeList's length exceeds the maximum possible number of pointers, the pointer with the lowest T-index will be removed. If its T-index value does not exist in top-m T-index, the TopTrusteeList remains unchanged. In order to add a user to TopTrusteeList, we first check its trust with current users in TopTrusteeList with higher T-index, to make sure that there is no other users with direct trust relationship in the TopTrusteeLists.

By using TopTrusteeList, we aim to make other users accessible across the web for a particular user, which is not in their neighborhood even within t paths. For this purpose, TopTrusteeList should point to users from different clusters, each consist of similar users. It looks like selecting a representative from any kind of preferences model which might be interested in a particular item. Therefore, we define a measure called T-index as mentioned before, to select the most trustworthy nodes from all kinds of interests shown as clusters of users. The T-index value increases as a user becomes closer to the center of a cluster.

When a user rate a new item, it computes mutual trust between the user and top-m users in the item's TopTrusteeList. If it finds more trustworthy users than its current neighbors, it will update its trustee list with the new users.

Defining the number of traversing leads to a trade off between accuracy and performance. Therefore, as the number of parallel traversals and their maximum length increases, we achieve better prediction accuracy and coverage. However, it takes more resources of bandwidth and computation. We define a minimum trust and maximum distance threshold that any traverse can proceed until it reaches specified the thresholds. The distance is defined by the number of edges from the user where the traversing originates. For computing trust between two indirect users, we simply multiply the trust value of the edges which builds a path between them.

$$T_{u_a,u_c} = T_{u_a,u_b} * T_{u_b,u_c}$$

This trust computation is done repeatedly during each traversal to achieve the trust value between the source and the target user. Traversing continues until it reaches the minimum trust threshold.

In order to recommend a specific item $i_i$ to user us, we originate two traversing, one directly from us's neighbors and the other one from the $i_i$'s TopTrusteeList. If a user can find a trustworthy user among its trustees , it can directly starts its inquiry directly from a user located in the center of the cluster which its users more likely rated the item before and are trustworthy as well. We show that this approach increases both accuracy and coverage rather than when the inquiry is done only among the neighbors. After collecting the rating from all the originated traversing, firstly we adjust their rating based on the formula 2 and then we predict its ranking for the us as the following:

$$p(u_s, i_i) = \frac{\sum_x (tr(r_{u_x,i_i}) * T(u_s, u_x) * T_{index}(u_x))}{\sum_x (T(u_s, u_x) * T_{index}(u_x))}$$

## Extending Recommender Systems with Socio-Semantic Trust: An Evaluation

A prototype of the recommender system has been implemented. Reader is advised to refer to (Zarghami, et al., 2009) for further information.

We  have used the MovieLens dataset (Riedl, et al.) for evaluating our method which consists of 943 ontological user profiles. Ratings are based on five point scale.The profiles are divided into training and test sets including 80% and 20% of ratings, respectively.

## Conclusion

This section discusses the conclusions of the work gathered and disseminated under deliverable 2.2.

First a general preface is given, while a more detailed description of the selection of algorithms and approaches is given in the following section.

### General Conclusions

The merit of WP2 has been introducing the formalization of semantics and content of user profiles for the project. In this deliverable we have presented a framework for extending the user profiles presentation with dyamic operations allowing creation and modification of the profiles of users using the platform.

In order to justify the approaches undergone to implement introduced operations within the framework, a very thorough state-of-art of research survery is given. This survey gives a thorough knowledge about personalization, contextualization and recommendation in the context of cultural heritage. Majority of the focus of this survery is given to learning and building user profiles while recommender systems have been studied briefly to pave the way for introducing the dynamic profile operations at the top of them.

Following the survey, dynamic profile operations are introduced as a framework for extending the profile framework introduced in D2.1. We have suggested collaborative filtering recommenders for enabling dynamic profile operations for learning, modifying as well as allowing profile-driven value-adding services such as personalized information retrieval based on user preferences or trust.

General conclusion is that within the framework of this WP, we have introduced two CF-based solutions.

First approach, can incrementally learn and build user profiles, in an statistical way (simple mode), and can recommed items based on statistical notion of user's preferences, while second approach can consider social trust between users of the platform and improve the recommendation based on the social trust (extended mode).

**Detailed Conclusions: Selection of Algorithms and Mechanisms**

For *mining and learning profiles* of the users, we have selected unsupervised learning approach, using *association rules*, *frequent closed item sets* (two-mode clustering approach) (Liv, 2008) (Pasquier, et al., 1999) (Zaki, et al., 2002).

For example, in implementation of closed item set mining, *Apriori and Eclat* (Borgelt., 2003) are utilized and implemented.

For *matching user profiles* and *generating/predicting recommendation for users* we strongly suggest utilizing *Collaborative Filtering* approaches, taken as general.

*Similarity measures* used in Collaborative Filtering approaches is used for matching user/item attributes based on approaches under taken.

*Item-based family* of Collaborative Filtering approaches (Sarwar, et al., 2001) (Karypis, 2001) (Linden, et al., 2003) (Deshpande, et al., 2004) is suggested for simple case.

In this case *Slope one* (Lemiere et al., 2005) is considered and implemented along with .

While *User-based family* of Collaborative Filtering approaches (Resnick, et al., 1994) (Shardanand, et al., 1995) is suggested for extended case, where weights (privacy and trust) are utilized with user profiles.

For instance, in extended approach (*T-index mechanism*) (Zarghami et al., 2009) User-based approach is used where It computes the difference of user rating to common items considering their transposed values (skewed rating). Based on these differences, each user keeps N high trustworthy as its direct neighbors and propagate the trust with other indirect neighbors maybe who has not any common items with itself.

## Works Cited

**Abowd G. D. [et al.]** Cyberguide: A mobile context-aware tour guide [Journal]. - [s.l.] : Wireless Networks. - Vol. 3.

**Adomavicius G and Tuzhilin A** Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions [Journal] // IEEE transactions on knowledge and data engineering. - [s.l.] : IEEE, 2005. - 6 : Vol. 17. - p. 734. - 1041-4347.

**Agrawal R and Srikant R.** Fast algorithms for mining association rules [Conference] // International Conference on Very Large Data Bases. - Santiago, Chile : [s.n.], 1994. - pp. 487–499.

**Agrawal R., Imielinski T. and Swami A.** Mining Association Rules Between Sets of Items in Large Databases [Conference] // ACM SIGMOD Record. - 1993. - Vol. 22 (2). - pp. 207–216.

**Aroyo L [et al.]** Personalized Museum Experience: The Rijksmuseum Use Case [Conference]. - [s.l.] : Proceedings of museums and the Web 2007, 2007.

**Auer S. [et al.]** Triplify â˘A ¸S lightweight linked data publication from relational databases. [Conference] // Proceedings of the 18th World Wide Web Conference. - April 20th-24th 2009.

**Barsalou L. W.** Situated simulation in the human conceptual system [Article] // Language and Cognitive Process. - 2003. - 5/6 : Vol. 18.

**Billsus D. and Pazzani M.** A hybrid user model for news story classification [Conference] // UM99 User Modeling: Proceedings of the Seventh International Conference / ed. Kay J.. - New York : Springer-Verlag, 1999.

**Bishop C. M.** Pattern Recognition and Machine Learning (Information Science and Statistics) [Journal]. - New York : Springer-Verlag, 2006.

**Bloedorn E. Mason, G. Mani I. and MacMillan T. R.** Machine Learning of User Profiles: Representational Issues. [Conference] // Proceedings of the national conference on artificial intelligence. - [s.l.] : John Wiley & Sons ltd, 1996. - Vol. 13. - pp. 433-438.

**Braun T., Glänzel W and Schubert A. S.** A Hirsch-type index for journals [Journal] // SCIENTOMETRICS. - Netherlands : Springer Science + Business Media, 2006. - 1 : Vol. 69. - pp. 169-173 . - 0138-9130.

**Breese J., Heckerman D. and Kadie C.** Empirical analysis of predictive algorithms for collaborative filtering. [Conference] // 14th Conference on Uncertainty in Artificial Intelligence. - [s.l.] : Morgan Kaufman, 1998. - pp. 43–52.

**Brickley Dan and Miller Libby** FOAF Vocabulary Specification 0.91 [Online] // FOAF Project. - November 2007. - http://xmlns.com/foaf/spec/.

**Brusilovsky P.** Adaptive hypermedia [Article] // User Modeling and User-Adapted Interaction. - 2001. - 11 : Vols. 1-2. - pp. 87-100.

**Candillier L [et al.]** State-of-the-Art Recommender Systems [Book Section] // Collaborative and Social Information Retrieval and Access. - [s.l.] : IGI, 2009.

**Charniak E.** Bayesian Networks Without Tears [Journal]. - [s.l.] : AI Magazine, 1991. - 4 : Vol. 12.

**Chen G. and Kotz D.** A Survey of Context-Aware Mobile Computing Research [Report]. - [s.l.] : Dept. of Computer Science, Dartmouth College, 2000. - TR2000-381.
**Cheverst K. [et al.]** Developing a context context-aware electronic tourist guide: some issues and experiences [Conference] // SIGCHI - conference on Human factors in computing systems. - The Hague, The Netherlands : [s.n.].
**Clerkin P, Cunningham P and Hayes C** Ontology discovery for the Semantic Web using hierarchical clustering [Conference] // Proceedings of Semantic Web Mining Workshop. - Freiburg, Germany : ECML/PKDD, 2001.
**Borgelt C**., Efficient implementations of apriori and eclat. Workshop of Frequent Item Set Mining Implementations, 2003.
**Dai H., and Mobasher B.** A road map to more effective Web personalization:Integrating domain knowledge with Web usage mining [Conference] // International Conference on Internet Computing. - Las Vegas, USA : [s.n.], 2003. - pp. 58–64.
**Deshpande M. and Karypis G.** Item-based top-N recommendation algorithms [Journal]. - [s.l.] : ACM Transactions on Information Systems, 2004. - 1 : Vol. 22.
**Dey A.K., Abowd G.D., and Salber D.** A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications [Conference] // Human-computer interaction. - 2001. - Vols. 16(2-4). - pp. 97–166.
**Dokoohaki N., Matskin M.** Effective Design of Trust Ontologies for Improvement in the Structure of Socio-Semantic Trust Networks [Journal] // International Journal On Advances in Intelligent Systems,. - [s.l.] : IEEE Computer Society, 2008. - 1 : Vol. 1. - publication in progress. - 1942-2679.
**Dokoohaki N., Matskin M.** Personalizing Human Interaction through Hybrid Ontological Profiling: Cultural Heritage Case Study [Conference] // 1st Workshop on Semantic Web Applications and Human Aspects, (SWAHA08), / ed. Ronchetti Marco. - Bangkok, Thailand : AIT e-press, 2008. - pp. 133-140. - In conjunction with Asian Semantic Web Conference 2008.
**Dokoohaki N., Matskin M.** Structural Determination of Ontology-Driven Trust Networks in Semantic Social Institutions and Ecosystems [Conference] // International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM'07) and the International Conference on Advances in Semantic Processing (SEMAPRO 2007). - Papeete, French Polynesia : IEEE Computer Society, Los Alamitos, CA, USA,, 2007. - pp. 263-268. – 0-7695-2993-3.
**Dolog P. and Nejdl W.** Challenges and Benefits of the Semantic Web for User Modelling [Conference]. - [s.l.] : Proceedings of Adaptive Hypermedia, 2003.
**Dolog Peter and Nejdl Wolfgang** Semantic Web Technologies for the Adaptive We [Book Section] // The Adaptive Web. - [s.l.] : Springer Berlin / Heidelberg, 2007.
**Fink J. and Kobsa A.** User Modeling for Personalized City Tours [Article] // Artificial Intelligence Review. - 2002. - Vol. 18.
**Fisher D. H.** Knowledge Acquisition Via Incremental Conceptual Clustering [Article] // Machine Learning. - 1987. - Vol. 2. - pp. 139–172.
**Frank A.** Qualitative spatial reasoning about distances and directions in geographic space [Journal]. - [s.l.] : Journal of Visual Languages and Computing, 1992. - Vol. 3. - pp. 343–371.
**Gauch Susan [et al.]** User Profiles for Personalized Information Access [Book Section] // The AdaptiveWeb / ed. Brusilovsky Peter, Kobsa Alfred and Nejdl Wolfgang. - [s.l.] :

Springer Berlin / Heidelberg, 2007. - Vol. 4321/2007. - ISSN: 0302-9743 (Print) 1611-3349 (Online). - 978-3-540-72078-2.

**Godoy D. and Amandi A.** Modeling user interests by conceptual clustering [Article] // Information Systems. - 2006. - 31. - pp. 247–265.

**Golbeck J.** FilmTrust: Movie Recommendations from Semantic Web-based Social Networks [Conference]. - [s.l.] : IEEE Consumer Communications and Networking Conference, 2006.

**Golbeck J.** Trust and nuanced profile similarity in online social networks [Report]. - [s.l.] : University of Maryland, College Park, 2007.

**Haines E.** Point in polygon strategies [Article] // Graphics Gems IV / ed. Heckbert P.. - [s.l.] : Academic Press, 1994. - pp. 24–46.

**Hand D., Mannila H. and Smyth P.** Principles of data mining. [Book]. - Cambridge, MA : MIT Press, 2001.

**Hyvönen E** Semantic Portals for Cultural Heritage, [Book Section]. - 2007. - Manuscript draft for a book chapter.

**Hyvönen E. [et al.]** Culturesampo – finnish cultural heritage collections on the semantic web 2.0. [Conference] // Proceedings of the 1st International Symposium on Digital humanities for Japanese Arts and Cultures (DH-JAC-2009). - Ritsumeikan, Kyoto, Japan : [s.n.], March 2009.

**Hyvönen E. [et al.]** Culturesampo – finnish culture on the semantic web 2.0. thematic perspectives for the end-user [Conference] // Proceedings, Museums and the Web 2009. - Indianapolis, USA : [s.n.], April 15-18 2009.

**Hyvönen E. [et al.]** MuseumFinland—Finnish museums on the semantic web. [Journal] // Journal of Web Semantics. - 2005. - 2 : Vol. 3. - pp. 224–241.

**Karypis G.** Evaluation of item-based top-N recommendation algorithms [Conference] // 10th International Conference on Information and Knowledge Management. - 2001. - pp. 247–254.

**Kauppinen T. [et al.]** Ontology-based disambiguation of spatiotemporal locations [Conference] // Proceedings of the 1st international workshop on Identity and Reference on the Semantic (IRSW2008). - Tenerife, Spain : 5th European Semantic Web Conference 2008 (ESWC 2008),, June 1-5 2008.

**Kauppinen Tomi** Methods and Tools for Fetching Knowledge about Nearby Points and Coordinates of Interest [Report]. - Espoo, Finland : TKK, 2009.

**Kelleher J. and Bridge D.** Rectree centroid:An accurate, scalable collaborative recommender [Conference] // 14th Irish Conference on Artificial Intelligence and Cognitive Science / ed. Cunningham P., Fernando, T., & Vogel, C.. - 2003. - pp. 89–94.

**Kim H.-R. and Chan P. K.** Learning implicit user interest hierarchy for context in personalization. [Article] // Applied Intelligence. - 2008. - 2 : Vol. 28. - pp. 153-166.

**KIRT TOOMAS** PERSONALIZATION AND CONTEXTUALIZATION IN A SMARTMUSEUM ENVIRONMENT: A SURVEY [Report]. - Tallinn : ELIKO, 2008.

**Kleinberg J. and Sandler M.** Using mixture models for collaborative filtering [Conference] // ACM Symposium on Theory of Computing. - [s.l.] : ACM, 2004. - pp. 569–578.

**Kobsa A.** Generic User Modeling Systems [Journal]. - [s.l.] : User Modeling and User-Adapted Interaction Journal, 2001. - Vol. 11. - pp. 49-63.

**Kobsa A.** User Modeling: Recent work, prospects and Hazards [Book Section] / book auth. SCHNEIDER-HUFSCHMIDT M.. - [s.l.] : SCHNEIDER-HUFSCHMIDT, M. , 1993.

**Koch. Nora** Software Engineering for Adaptive Hypermedia Systems [Report]. - Munich, Germany : PhD thesis, Ludwig-Maximilians University, May 8th, 2005.

**Kohonen T.** Self-organized formation of topologically correct feature maps [Journal]. - [s.l.] : Biological Cybernetics, 1982. - 43. - pp. 59–69.

**Kohonen T.** Self-organizing maps [Journal]. - Berlin : Springer, 2000.

**Korpipää P. [et al.]** Bayesian Approach to Sensor-Based Context Awareness [Journal]. - [s.l.] : Personal and Ubiquitous Computing Journal, 2003. - 4 : Vol. 7.

**Korpipää P. [et al.]** Managing context information in mobile devices [Conference] // IEEE pervasive computing. - 2003. - Vol. 2 (3).

**Laerhoven K., and Cakmakci O.** What Shall We Teach Our Pants? [Conference] // IEEE International Symposium on Wearable Computers. - [s.l.] : IEEE, 2000.

**Lam T. H. W., and Lee R. S. T.** iJADE FreeWalker – An Intelligent Ontology Agent-based Tourist Guiding System, [Article] // Studies in Computational Intelligence (SCI),. - 2007. - Vol. 72. - pp. 103-125.

**Lathia N., Hailes S. and Capra L.** Trust-based collaborative filtering [Conference] // IFIPTM 2008: Joint iTrust and PST Conferences on Privacy, Trust management and Security. - London, UK : University College London, 2008. - Vol. 14.

**Laurence S. and Margolis E.** Concepts and Cognitive Science [Book Section] // Concepts: Core readings / book auth. Margolis E. and Laurence S.. - Cambridge, MA : MIT Press, 1999.

**Lemire D. and Maclachlan A.**, Slope One Predictors for Online Rating-Based Collaborative Filtering, In SIAM Data Mining (SDM'05), Newport Beach, California, April 21-23, 2005

**Linden G., Smith B. and York J.** Amazon.com recommendations: Item-to-item collaborative filtering [Conference] // IEEE Internet Computing. - 2003. - Vol. 7(1).

**Liv Innar** Cultural Heritage Object Recommendation System and User Profile Matching [Report] : Technical Report. - [s.l.] : ELIKO, 2008. - Rev. 2.

**Massa P. and Avesani P.** Trust-aware collaborative filtering for recommender systems [Conference]. - [s.l.] : Springer, 2004.

**Meadow C. T. [et al.]** Text Information Retrieval Systems [Conference]. - London : Academic Press, 2007.

**Michalski R. S.** Knowledge acquisition through conceptual clustering: A theoretical framework and algorithm for partitioning data into conjunctive concepts [Journal] // International Journal of Policy Analysis and Information Systems. - 1980. - Vol. 4(3). - pp. 219-243.

**Middleton Stuart E., Shadbolt Nigel R. and De Roure David C.** Ontological User Profiling in Recommender Systems [Article] // ACM Transactions on Information Systems. - [s.l.] : ACM Transactions on Information Systems, 2004.

**Mobasher B., Cooley R. and Srivastava J.** Automatic personalization based on Web usage mining [Conference] // Communications of the ACM. - 2000. - Vol. 43(8). - pp. 142–151.

**Neapolitan R. E.** Learning Bayesian Networks [Book]. - Upper Saddle River, NJ : Prentice-Hall, 2003.

**O'Conner M. and Herlocker J.** Clustering items for collaborative filtering [Conference] // ACM SIGIR Workshop on Recommender Systems. - 1999.

**O'Donovan J. and Smyth B.** Trust in recommender systems [Conference] // 10th international conference on Intelligent user interfaces (IUI '05). - New York, NY, USA : ACM, 2005. - pp. 167–174.

**Park B. and Kargupta H.** Distributed Data Mining: Algorithms, Systems, and Applications [Book Section] // The Handbook of Data Mining / book auth. Ye N.. - New York : Lawrence Erlbaum, 2003.

**Pasquier N. [et al.]** Discovering Frequent Closed Itemsets for Association Rules [Conference]. - [s.l.] : SPRINGER VERLAG KG, 1999. - Vol. 1540. - pp. 398-416 . - ISSN 0302-9743.

**Pazzani M. and Billsus D.** Learning and Revising User Profiles The Identification of Interesting Web Sites [Article] // Machine Learning. - 1997. - 3 : Vol. 27.

**Pennock D. [et al.]** Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach [Conference] // 16th Conference on Uncertainty in Artificial Intelligence. - 2000. - pp. 473–480.

**Perkowitz M. and Etzioni O.** Towards adaptive Web sites: Conceptual framework and case study [Article] // Artificial Intelligence. - Vol. 118. - pp. 245–275.

**Pretschner A.** Ontology Based Personalized Search [Report]. - Kansas : University of Kansas, 1999.

**Resnick P. [et al.]** Grouplens: An open architecture for collaborative filtering of netnews. [Conference] // Computer Supported Cooperative Work. - [s.l.] : ACM, 1994. - pp. 175-186.

**Riedl J and Konstan J** Movielens dataset.

**Rosch E.** Principles of categorization [Book Section] // Cognition and Categorization / book auth. Lloyd E. Rosch & B. B.. - Hillsdale, NJ : Erlbaum, 1978.

**Rosenfeld L. and Morville P.** Information Architecture for the World Wide Web Designing Large-scale Web Sites [Book]. - Cambridge, MA : O'Reilly, 1998.

**Sarwar B. M. [et al.]** Item-based collaborative filtering recommendation algorithms [Conference] // 10th International World Wide Web Conference.. - 2001.

**Schilit B. N. and Theimer M. M.** Disseminating Active Map Information to Mobile Hosts [Article] // IEEE Network. - [s.l.] : IEEE, 1994. - 5 : Vol. 8.

**Schilit B. N., Adams N., and Want R.** Context-aware computing applications [Conference] // IEEE Workshop on Mobile Computing Systems and Applications. - Santa Cruz, USA : IEEE press, 1994.

**Schmidt A. [et al.]** Advanced Interaction in Context [Conference] // 1st international symposium on Handheld and Ubiquitous Computing. - Karlsruhe, Germany : [s.n.].

**Shardanand U. and Maes P.** Social information filtering: Algorithms for automating “word o fmouth” [Conference] // Conference on Human Factors in Computing Systems. - [s.l.] : ACM, 1995.

**Sieg A., Mobasher M. and R. Burke** Web Search Personalization with Ontological User Profiles [Conference] // CIKM’07. - Lisboa, Portugal : ACM, 2007.

**Singh S. [et al.]** An Adaptive User Profile for Filtering News Based on a User Interest Hierarchy. [Conference] // ASIS&T Annual Meeting..

**Sparacino F.** Natural Interaction in Intelligent Spaces: Designing for architecture and entertainment [Journal]. - [s.l.] : Multimedia Tools and Applications, Springer, 2008.

**Sparacino Flavia** MUSEUM INTELLIGENCE: USING INTERACTIVE TECHNOLOGIES FOR EFFECTIVE COMMUNICATION AND STORYTELLING [Conference]. - Berlin : ICHIM, 2004.

**Sugiyama K., Hatano K. and Yoshikawa M.** Adaptive web search based on user profile constructed without any effort from users [Conference] // 13th international conference on World Wide Web. - New York : ACM, 2004.

**Sutterer M, Droegehorn O and David K** UPOS: User Profile Ontology with Situation-Dependent Preferences Support [Conference]. - [s.l.] : First International Conference on Advances in Computer-Human Interaction, 2008. - 0-7695-3086-9.

**Tammet, Tanel.** Deliverable1.1 [Report] . SMARTMUSEUM, 2007.

**U. Cetintemel, Franklin M. and Giles C.** Self-Adaptive User Profiles for Large-Scale Data Delivery [Conference] // INTERNATIONAL CONFERENCE ON DATA ENGINEERING. - [s.l.] : IEEE Computer Society Press, 2000.

**Ungar L. and Foster D.** Clustering methods for collaborative filtering [Conference] // Workshop on Recommendation Systems. - [s.l.] : AAAI Press, 1998.

**Viljanen K., Tuominen J. and Hyvönen E.** Publishing and using ontologies as mash-up services [Conference] // Proceedings of the 4th Workshop on Scripting for the Semantic Web (SFSW2008). - [s.l.] : 5th European Semantic Web Conference 2008 (ESWC2008), June 1-5, 2008.

**Wang Y. [et al.]** Interactive User Modeling for Personalized Access to Museum Collections: The Rijksmuseum Case Study [Conference]. - [s.l.] : 11th User Modeling Conference, 2007.

**Weibelzahl Stephan** Evaluation of Adaptive Systems [Report] : PhD thesis. - [s.l.] : University of Trier, 2003.

**Zaki M. J.** Parallel and distributed association mining: A survey [Article] // IEEE Concurrency. - [s.l.] : IEEE, 1999. - 7. - Vol. 4.

**Zaki MJ and Hsiao CJ** CHARM: An efficient algorithm for closed itemset mining [Conference] // 2nd SIAM International Conference on Data Mining. - 2002.

**Zarghami Alireza [et al.]** An Ontological Trust-Based Recommender System using T-index [Report]. - Stockholm : KTH ( Royal Institute of Technology) , 2009.

**Ziegler C.N. [et al.]** Improving recommendation lists through topic diversification [Conference] // WWW '05: Proceedings of the 14th international conference on World Wide Web. - New York, NY : ACM, 2005.