Toward User Profile Representation in Adapted Mediation Systems

Sara Ouaftouh, Ahmed Zellou and Ali Idri
Mohammed V University In Rabat, Rabat, Morocco

Keywords: User Modeling, Mediation Systems, User Model Representation.

Abstract: The amount of information offered by different software systems is growing exponentially and the need of personalized approaches for information access increases. This personalization aims to offer the user the pertinent information corresponding to his needs basing on his profile. For the same purpose, mediation systems have to identify user preferences in order to offer him the most relevant information .In this work we discuss different representations of user profile models designed for providing personalized information access in order to make a comparison and identify the most appropriate for our context in mediation systems.

1 INTRODUCTION

Today, the amount of information available in the different information systems is increasing exponentially. This information can be in heterogeneous sources: relational or object sources, flat files, structured data, applications, web services, etc.

In order to integrate these heterogeneous information sources and offer an added value to its services, information systems in different domain are using data integration technologies. We basically distinguish between two principle integration methods: physical integration (J. Widom, 1995) and virtual integration. The virtual integration also called mediation allows combining a set of different information sources by allowing a real-time access to the sources while concealing their particularities (G. Wiederhold, 1992). The mediation system must be able to intercept user queries on the information system and return appropriate responses.

However, any system that doesn't know who is asking for information and for what purpose, will never be able to provide more than general answers. Therefore, we need a mechanism to adapt the behavior of information systems to user's preferences. When a system integrates this mechanism of adaptation, it is called an «Adaptive system» (K. Cheverst et al., 2002). These mechanisms must be able to provide the system with some "context awareness" by extracting from the user context, the information needed to identify his preferences. The system will then provide the user with personalized services. The personalization or adaptation in information system is based on the concept of user profile.

The user profile is defined as a set of information describing the user and simulating his preferences. The user profile is considered as a set of structured data describing the interaction environment between a user and a system (Y. Elallioui and O. El Beqqali, 2012). In the domain of Internet search engines (S. Calegari and G. Pasi, 2010), the user profile is used in order to have structured representation of user's interests.

The implementation of a user profile requires the creation of a user model (R. Guha et al., 2015). Adaptation via user modeling has started by the end of the 1970s before the introduction of the Web (A. Kobsa, 2001), recently it has become a main component of many web applications and information systems in general.

In this work we present a comparison between the different representations of a user profile in order to deduce the most appropriate to adapt in the context of mediation system personalization.

The remaining parts of this paper can be summarized as following: in section two, we present mediation systems. Section three describes the different factors that made the personalization of mediation system a necessity to satisfy user's expectations while section four is dedicated to discuss the different representations of user models that will be compared and discussed in section Five. We conclude then this paper and shares some future works.

Ouaftouh, S., Zellou, A. and Idri, A.
Toward User Profile Representation in Adapted Mediation Systems.
DOI: 10.5220/0006043000810087
In *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2016) - Volume 2: KEOD*, pages 81-87
ISBN: 978-989-758-203-5

2 MEDIATION SYSTEMS

Integration systems permit to link together information coming from sources that are often heterogeneous and distributed in order to provide a global view of information to users.

In our work, we focus on the mediator approach (G. Wiederhold, 1995) which consists of a mediation layer between the user and information sources in order to provide the user with a centralized and uniform view of information by hiding specific characteristics of their location, access methods and formats. An information source can be in different formats.

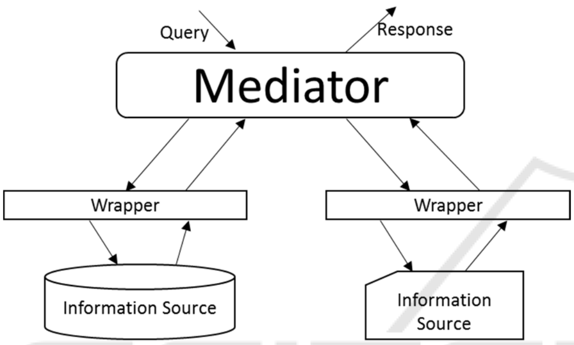

Figure 1: Mediation architecture.

In a general manner, the user interacts with the system by querying a global schema (M. Lenzerini, 2002) which is a virtual representation of the data on the sources. In the one hand, the mediation system will perform the sources processing task to retrieve information satisfying the user query. On the other hand, the mediation system will hold an internal representation of information sources, called source or local schema, so that relations exists between global entities and the local schema: the mapping represents these relationships. Some software modules called wrappers hide source characteristics to facilitate the interaction between the mediation system and the sources.

Mediation systems are very useful in the presence of heterogeneous data sources, because they make the user feel like using a homogeneous system. Among the different categories of mediation systems applications, we can mention the information retrieval applications; online decision support systems and more generally, knowledge manage-ment applications.

However, mediation systems represent some limitations, including the fact that query rewriting in terms of data sources views requires the knowledge of the sources. There is also the risk of non-availability of all information sources needed to build a response at the time of issuing the request by the user. Furthermore, it is important to address the problem of adaptability of the mediation system to users' needs due to the large number of data sources, which may contain redundant information and varying quality. In our work, we are interested in how to personalize mediation systems in order to resolve this last problem.

3 NECESSITY TO PERSONALIZE MEDIATION SYSTEMS

As we presented in the previous section, despite the benefits of mediation systems, they suffer from some inconvenient. The one we are interested in our work is caused by the big number of information sources, which may contain redundant information in addition to vocabulary problems like polysemy and synonymy. In general, the evaluation of a user query is independent of the context and the needs of the user who issued it. Therefore, the same query submitted by two different users, produces the same results even if these users have different expectations which could make the search results of mediation systems not beneficial to the user. The user is then faced to a large number of information that doesn't correspond to his expectations when he submitted the request. However in personalized systems, the user preferences are included. Applying personalisation in an e-commerce context for example, if two different users send the same query to look for shoes, they could get different results. Considering that the first user is a girl who lives in Brazil; she will get as propositions, teenager sandals style corresponding to the warm weather in Brazil. The second user who is a 60 years old man living in Germany will get shoes corresponding to old people generation. In addition to adapted results, e-commerce web site implement the concept of recommender systems that permit the possibility to give the user suggestion about recommended content that could interest him according to his profile.

In order to offer the user the personalised answer adapted to his expectations, we need first to identify his needs and preferences. Many works in the literature discusses how information systems are being adapted to user's expectations. Adapted systems take into account the different characteristics of the user and his situations in different contexts basing on the concept of user profile.

We usually distinguish two phases in the modelling of the user profile: initialization and update. In fact, the profile is initialized on the first use of the system. The first elementary step to build the user profile is the collect of user's information (S. Schiaffino and A. Amandi, 2009). We separate here between three methods: Explicit, implicit and hybrid information gathering. To build a user profile different dimensions are considered: user's knowledge, preferences, habits, physical abilities, intentions, psychological states and geographical location (S. Ouaftouh et al., 2015). This list is not exhaustive, it could include, for example, personal data about the user, his professional or social role, etc. These dimensions can represent relatively stable characteristics over time or changing ones that are therefore updated over the time. In general, the nature of the information contained in the user profile strongly depends on the application and purpose of the system that implements it. The user profile is then evaluated, modelled and represented in a particular form. In this context and in a perspective of user profile integration in the conception of mediation systems, what is the most appropriate user profile representation?

4 USER MODEL REPRESENTATIONS

Adaptive systems are based on user model in order to have different behaviors for different users. The User Model is a representation of the information about a specific user. In literature, we find different categorizations for user models basing on different characteristics.

A. Keyword-based User Modeling:

Keyword-based user modeling was initiated in the domain of information retrieval and filtering (P. Brusilovsky and C. Tasso, 2004), where the content of a document is represented as a vector of terms called keywords, extracted from the text. The adaptive information retrieval and filtering applications combined for example a history of user's queries, accessed documents, e-mails, chat, etc. in a form of a keyword vector and use this vector for adapting a future retrieval or filtering process.

Many adaptive systems model users' information interests or needs as vectors of keywords extracted from the documents that the users have browsed or requested. Figure 2 is an example of a keyword vector representation that models user interest in a particular domain of interest, each keyword *«K_i»*, corresponds to a term found in the content of document consulted by the user. Each keyword *«K_i»* is associated with a numerical weight *«W_i»* representing its importance in the profile.

(W_i, K_i) Domain of interest

W_1	W_2	W_3	…	W_n
K_1	K_2	K_3	…	K_n

Figure 2: Keyword-based user model representation.

Each keyword can represent a topic of interest. Keywords can be grouped in categories to reflect a more standard representation of users' interests. Some systems model users' interests as networks of keywords instead of plain lists, where nodes represent keywords and arcs connect keywords co-occurring in the content.

Keyword-based modeling support only simple content data. To remedy problems like homonymy and synonymy, Natural Language (NL) technologies are required. Pure keyword based modeling is not able to represent the true meaning of the content. It relies on statistical regularities within the text and provides a framework for retrieving statistically close documents.

B. Overlay User Modeling:

This approach also called concept user modeling, was employed for the first time in 1988. It's about graining the domain knowledge into elementary components and using them to evaluate user's knowledge. The domain knowledge components have been named differently by different authors: topics, knowledge elements and – the most used – concepts.

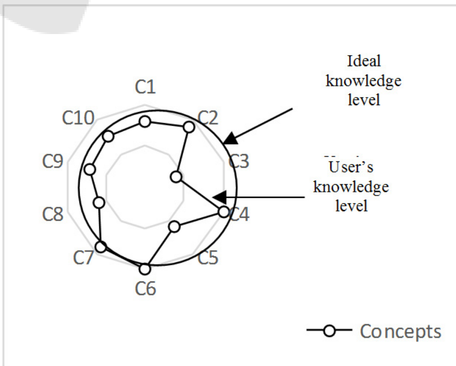

Figure 3: Overlay user model.

A concept represents an atomic piece of declarative domain knowledge, coherent and semantically complete. An aggregate of concepts forms the domain model. The overlay user model consists on a set of concept-value pairs, where the value represents an assessment of a particular concept. The user is characterized in terms of user's knowledge about these concepts in relation to the top level knowledge.

As shown in figure 3, «*Ci*», correspond to the different concepts from a domain knowledge, the user model represent user knowledge about these concepts in relation to the ideal knowledge level.

The benefit of the overlay user model is its precision and flexibility. An overlay model is capable to dynamically and precisely reflect the evolution of users' characteristics.

C. Stereotype User Modeling:

The main goal of adaptive systems is to adjust its behavior to each user's needs. However, for some contexts it is possible to identify typical categories of users that use the system the same way, expect from it similar reactions and can be described by similar characteristics. These categories are called stereotypes. Stereotype user modeling is one of the oldest approaches to user modeling. It was developed in the works of Elaine Rich (E. Rich, 1997). An adaptive system using stereotype-based modeling does not update every single feature of the user model directly; it uses a stock of preset stereotype profiles. Hence, the application can make expectations about a user even though there could be no data about that specific area, because studies have shown that other users in this stereotype have the same characteristics. Such categories are called stereotypes.

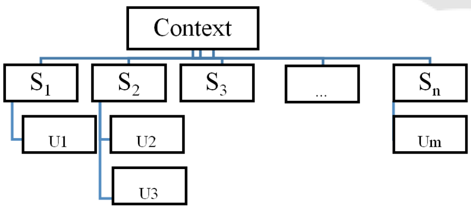

Figure 4: stereotype user model.

As shown in figure 4, an adaptive system in a specific context can identify a set of stereotypes *Si* and each user «U_i» of the system is assigned to a stereotype. A stereotype can correspond to one or many users.

As an example of stereotype-based user modeling is linear set of categories for typical levels of user proficiency**:** novice, beginner, intermediate, and expert.

Stereotype-based user modeling is advantageous when from a little evidence about a user the system should infer a great deal of modeling information. However, for modeling fine-grained characteristics about users as is the case in knowledge level of a particular concept, the overlay models should be employed.

D. Constraint-based user Modeling:

Constraint-Based model is a way to represent the domain knowledge as a set of constraints. It is mostly used for modeling users in Intelligent Tutoring Systems (A. Mitrovic, 2012). Constraint-based tutors are Intelligent Tutoring Systems that use Constraint-Based modeling to represent the user model as a set of information about abilities, knowledge and needs of the user. Constraint-based tutors are problem-solving environments; in order to provide personalized instruction; they diagnose users' actions, and maintain user models. These models are then used to provide appropriate examples and offer hints and help where the user is most likely to need them. In this approach, every constraint represents an acceptable set of equivalent problem states and a violated constraint indicates an error.

Each constraint consists of an ordered pair (*Cr, Cs*), where *Cr* is the relevance condition and *Cs* is the satisfaction condition. The relevance condition checks whether the constraint is applicable to the user solution by testing the features of the solution. The satisfaction condition specifies additional test that must be met by correct solutions. If the relevance condition is met, but the satisfaction condition is not, then the user's solution is incorrect. Therefore, the general form of a constraint as presented in figure 5 is:

$$(Cr, Cs) \left| \begin{array}{l} \text{If } < Cr > \text{ is true,} \\ \text{Then } < Cs > \text{ had better also be true.} \end{array} \right.$$

Figure 5: Constraint-based user modelling.

E. Collaborative Filtering:

Being different from user modeling technologies cited before, this approach relies on modeling the user in terms of his relationships with other users. A typical collaborative user model is based on a vector of ratings that the user provided for particular items. The original implementation of this approach is recommender systems (N. Tintarev and J. Masthoff,, 2011), they recommend to the active user the items that other users with similar preferences liked in the past. The similarity in preference of two users is calculated based on the similarity in the rating history of the different users. The central hypothesis behind this method is that other users' opinions can be selected and combined in such a way to provide a reasonable prediction of the active user's preference. Intuitively, we assume that, if users agree about the quality or relevance of some items, then they will likely agree also about other items.

The information domain of a collaborative filtering system consists of users which have expressed preferences for various items. A preference expressed by a user for an item is called a rating and is frequently represented as a (User, Item, Rating) triple. These ratings can take many forms, depending on the system. Some systems use real or integer valued rating scales such as 0–5 stars, while others use binary (like/dislike) scales (J. B. Schafer et al., 2007). As represented in the example in table 1, the set of all rating triples forms a matrix referred to as the ratings matrix. The values of R_j correspond to the rating each user U_i gave to a specific item I_k. (User, Item) pairs where the user has not expressed a preference for the item are unknown values in this matrix and are marked with '?' to indicate unknown values i.e. the user has not rated that item.

(U_i, R_j, I_k)

	I_1	I_2	I_3	…	I_m
U_1	R_1	?	R_2	…	R_3
U_2	?	R_4	R_5	…	?
…	…	…	…	…	…
U_n	R_6	R_7	R_8	…	?

Figure 6: Collaborative filtering using ratings matrix.

Given a user and an item, what is the user's likely preference for the item? If the ratings matrix is viewed as a sampling of values from a complete user–item preference matrix, than the predict task for a recommender is equivalent to the matrix missing values problem.

Recommender technology, often based on collaborative filtering, has been integrated into many e-commerce and online systems. An important motivation for doing this is to increase sales volume; customers will likely buy an item if it is suggested to them but may not otherwise.

F. Bayesian Networks:

Bayesian Networks are probabilistic graphical models that consist of a qualitative and a quantitative part. The qualitative part is the structure of the network: a directed acyclic graph where nodes correspond to variables and arcs representing influences between variables. The quantitative part provides the conditional probability tables that make up the network settings.

More precisely, a Bayesian Network is a set consisting of a directed acyclic graph and n random variables
$(X_1, X_2, .. , X_n)$ such that there is a bijection between the set of vertices graph and the set of random variables and that:
$P(X1, X2, .. , Xn) = \prod_{i=1}^{n} P(Xi \mid pa(Xi))$ where $pa(Xi)$ is the set of parents of X_i in the graph.

Multiple systems used Bayesian Networks to model the relations between different components or dimensions of a user model, such as emotions, goals and knowledge (X. Zhou and C. Conati, 2003). Other systems used them to implement an overlay user model with internal inference capabilities, where every node represents a domain concept and links represents the concept relations (F. De Rosis, et al.,1992).

In the e-commerce applications, it is often useful to model a customer without developing any explicit modeling rules about him, but only by identifying certain statistical predictabilities that can be used for constructing an effective selling strategy. A user model in this case can contain a set of transactions matched against an association rule of items bought together or satisfying some conditions of buyers' behavior, or belonging to a cluster of similar buyers (S. Sosnovsky, 2010).

5 COMPARISON AND DISCUSSION

A. Comparison:

After having presented the different categories of user models and the different representations, we constructed a table summarizing each type of user model representation to help us make a comparison. Table I represent for each category, the definition, representation, explication, principle domain of use, advantages and disadvantages.

After having presented the different categories of user models and the different representations, we constructed a table summarizing each type of user model representation to help us make a comparison.

The keyword based modeling approach can be considered lightly similar to overlay user modeling because it uses elements of domain representation as a reference to express user characteristics. Overlay user modeling is most used to model student knowledge in e-learning systems while keyword based models are used to model user interests for example in the domain of information retrieval and filtering. Constraint-based models are also used in

Table 1: Comparison table between different user model representations.

Name	**Overlay**	**Keyword**	**Stereotype**	**Constraint-based**	**Collaborative Filtering**	**Bayesian Network**
Principe	Measure how well a user knows a concept.	Measure user's interests on a specific keyword	Define typical categories of users.	Define a set of constraints representing domain knowledge.	Model user in terms of his relationships with other users.	Model the relations between the components of a user model
Representation	Vector of Concept-value pairs.	Vector of keywords-value.	Set of preset profiles.	Set of constraints.	Vector of ratings : (User, Item, Rating)	Directed acyclic graph
Explication	Concepts are subset of domain knowledge.	Keyword extracted from the text consulted by the user.	Users that belong to the same stereotype described by similar characteristics.	Diagnose of user's actions to provide personalized instructions and decisions.	Predict user interest basing similar user's feedback.	Nodes of the graph represent variables and arcs represent influences between variables.
Principle Domain of use	E-learning: Modeling user's knowledge.	Information retrieval and filtering. Model user interests.	Modelling groups of users.	Modeling user's knowledge in Intelligent Tutoring Systems	Recommender system: predicting user's interests and preferences.	Modeling emotions, goals and knowledge.
Advantages	Automatic modeling of content.	Faster results	Make expectations about a user even though there is no data about him	Encodes correct domain knowledge	No need to additional information about user except ratings.	Model a user only by identifying certain statistical predictabilities
Disadvantages	Support only simple content.	Support only simple content data. Lack of semantics and Polysemy.	Use of preset profiles, no update to users' features	Overspecificity: a highly detailed model of user's knowledge	Cold-start problem : no rating available in the start of the system	Difficult reaching agreement on the Bayesian network structure with experts

the domain of e-learning especially for intelligent tutoring systems to model students' knowledge basing on a set of constraint.

For its part stereotype user modelling can be used to model groups of users in the case the categories of system's user are predefined.

Collaborative filtering modeling is mainly used in recommender systems to predict user's interests and preferences. This method utilizes only ratings and do not require any additional information about users or items. The principal disadvantage of Collaborative filtering systems is the Cold-Start problem which cannot produce recommendations if there are no ratings available.

To model emotions, goals and also knowledge, the Bayesian network approach can be used with the advantage of modeling the user only by identifying certain statistical predictabilities.

B. Discussion:

Roughly, the user dimension considered in a particular system depends on the intended field of application. For example, in the domain of e-learning we are mainly interested to model user's knowledge, skills and interests. In e-commerce context, it's more interesting to know user's preferences. Among the deductions of our comparison between the different user models representation, each representation is satisfying the particularities of a certain application domain.

In order to select the suitable user model representation for our context, which is mediation systems, we have to specify our system's characteristics. To personalize a mediation system, we are interested to model the most of user's dimensions. Mediation systems are characterized with a set of exchanges between the mediator and sources (couple of requests and responses). We can then recommend the use of keyword based user modeling or collaborative filtering as the most appropriate to be applied in a mediation system context.

6 CONCLUSIONS AND PERSPECTIVES

A mediation system is a powerful tool allowing easy access to different information collected from distributed data sources that can be heterogeneous. It must integrate diverse information in order to provide the user with a centralized and uniform view of data by masking the specific characteristics of their location, access methods and formats. In a perspective of mediation system improvement, it has been necessary to adapt system's responses to user's expectations, represented by his profile, via the implementation of a user model.

The user profile corresponds to a set of information describing the user. It contains data that represent user preferences. The implementation of a user profile requires the creation of a user model. To identify the most suitable user model representation in a mediation system, we presented a study about the different approaches found in the literature. Many authors have classified user modeling approaches basing on a variety of criteria.

As our goal is to evaluate the user profile, we were based on user model representation and distinguish between overlay, keyword, stereotype, constraint-based, collaborative filtering and Bayesian network models. Each representation is mainly used to model a set of user's dimension and is generally applied in a particular domain of use. Practically, the user dimension considered in a particular system depends on the envisioned field of application. Referring to the particularity of mediation system and as deduced from the comparison table of user models that we constructed, we recommend the use of keyword user model or collaborative filtering approach.

In our future work, we will focus on applying one of the suggested representations to personalize mediation systems. This model will implement a set of dimension that we qualified necessary in a mediation system.

REFERENCES

J. Widom, "Research Problems in Data Warehousing," *4th Int. Conf. Inf. Knowl. Manag.*, pp. 25–30, 1995.

G. Wiederhold, "Mediators in the architecture of future information systems," *Computer (Long. Beach. Calif).*, vol. 25, no. 3, pp. 38–49, 1992.

K. Cheverst, K. Mitchell, and N. Davies, "Adaptive Hypermedia," vol. 45, no. 5, pp. 47–51, 2002.

Y. Elallioui and O. El Beqqali, "User profile Ontology for the Personalization approach," *Int. J. Comput. Appl.*, vol. 41, no. 4, pp. 31–40, 2012.

S. Calegari and G. Pasi, "Ontology-Based Information Behaviour to Improve Web Search," *Futur. Internet*, vol. 2, no. 4, pp. 533–558, 2010.

R. Guha, V. Gupta, V. Raghunathan, and R. Srikant, "User Modeling for a Personal Assistant," *Proc. Eighth ACM Int. Conf. Web Search Data Min. - WSDM '15*, pp. 275–284, 2015.

A. Kobsa, "Generic User Modeling Systems," *User Model. User-adapt. Interact.*, vol. 11, pp. 49–63, 2001.

G. Wiederhold, "Mediation in information systems," *ACM Comput. Surv.*, vol. 27, no. 2, pp. 265–267, 1995.

M. Lenzerini, "Data Integration: A Theoretical Perspective," *Proc. Twenty-first ACM SIGMOD-SIGACT-SIGART Symp. Princ. Database Syst.*, pp. 233–246, 2002.

S. Schiaffino and A. Amandi, "Intelligent User Profiling," *Ifip Int. Fed. Inf. Process.*, vol. LNAI 5460, pp. 193–216, 2009.

S. Ouaftouh, A. Zellou, and A. Idri, "User profile model: A user dimension based classification," *10th Int. Conf. Intell. Syst. Theor. Appl.*, pp. 1–5, 2015.

P. Brusilovsky and C. Tasso, "User modeling for web information retrieval," *Pref. to Spec. issue User Model. User Adapt. Interact.*, vol. 14, no. 2–3, pp. 147–1571, 2004.

E. Rich, "User Modeling via Stereotypes," *Cogn. Sci.*, vol. 3, no. 3597, pp. 329–354, 1979.

A. Mitrovic, "Fifteen years of constraint-based tutors: What we have achieved and where we are going," *User Model. User-adapt. Interact.*, vol. 22, no. 1–2, pp. 39–72, 2012.

N. Tintarev and J. Masthoff, *Recommender Systems Handbook*, vol. 54. 2011.

J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative Filtering Recommender Systems," *Int. J. Electron. Bus.*, vol. 4321, no. 1, pp. 291–324, 2007.

X. Zhou and C. Conati, "Inferring user goals from personality and behavior in a causal model of user affect," *Proc. 8th Int. Conf. Intell. user interfaces - IUI '03*, p. 211, 2003.

F. De Rosis, S. Pizzutilo, A. Russo, D. C. Berry, and F. J. Nicolau Moulina, "Modeling the user knowledge by belief networks," *User Model. User-adapt. Interact.*, vol. 2, no. 4, pp. 367–388, 1992.

S. Sosnovsky, "Ontological Technologies for User Modeling," vol. 5, no. 1, pp. 32–71, 2010.