Creating an Ontology for the User Profile: Method and Applications

Golemati Maria, Katifori Akrivi, Vassilakis Costas, Lepouras Georgios, Halatsis Constantin

Abstract—User profiling is commonly employed nowadays to enhance usability as well as to support personalization, adaptivity and other user-centric features. Insofar, application designers model user profiles mainly in an ad-hoc manner, hindering thus application interoperability at the user profile level, increasing the amount of work to be done and the possibility of errors or omissions in the profile model. This work aims at creating a user profile ontology that incorporates concepts and properties used to model the user profile. Existing literature, applications and ontologies related to the domain of user context and profiling have been taken into account in order to create a general, comprehensive and extensible user model. This ontology can be used as a reference model, in order to alleviate the aforementioned issues. The model, available for download, is exemplified through its application in two different areas, personal information management and adaptive visualization.

Index Terms—user profile, ontology, user modeling, context

I. INTRODUCTION

The continuing progress in network technologies and data storage has enabled the digitization and dissemination of huge amounts of documents. The need for more effective information retrieval has lead to the creation of the notions of the semantic web and personalized information management, areas of study that take advantage of the semantic context of the presented information and the user to facilitate the information storage and retrieval process. The notion of user profiling has been introduced in order to record the user context and personalize applications.

Ontologies have been proven an effective means for modeling digital collections and user context. They are effective, because they may present an overview of the domain related to a specific area of interest and be used for browsing and query refinement. Ontologies model concepts and relationships in a high level of abstraction, providing rich semantics for humans to work with and the required formalism for computers to perform processing and reasoning.

Using an ontology to model the user profile has already been proposed in various applications like web search [9], [12] and personal information management [7]. However, up to this point, ontologies modeling user profiles are application-specific and created specifically for a particular domain. Taking into account the continuing incorporation of ontologies in new applications, there is an emerging need for a standard ontology that will model user profiles; this standard ontology will facilitate communication between applications and serve as reference point when profiling functionalities need to be developed.

This work presents such an ontology for modeling user profiles. The purpose was to create a general yet extendable ontology that will be able to adapt to the needs of every application, maintaining at the same time a general common structure so as to satisfy portability and communication between different applications. After a brief overview of existing work in the area of profiling in relation with ontologies, the methodology for creating the user profile ontology is presented, followed by a presentation of the ontology itself. Examples of the application of the model in two domains are provided in the following section. The last section presents the conclusions and briefly outlines future work.

II. RELATED WORK

Recently the need for software systems to automatically adapt to their users has been recognized and the research on user profiling and context has spread into many disciplines [8]. Context in [13] is categorized in human user context and surroundings context and may also be categorized according to persistence (permanent and temporary) and evolution (static and dynamic). Elaine Rich [14] identifies a three dimensional space of user models: 1) canonical vs. individual user model, 2) explicit vs. implicit user model and 3) long-term vs. short-term user model.

Another important issue is that a user might be found in various contexts. Thus, a context-aware system has to infer which context the user is in a given moment in time, and consequently adapt the system to that context [14].

According to [23], a user model contains all information that the system knows about the user. It is generally initialized either with default values or by querying the user. Users in some cases are grouped in “stereotypes”, like “woman” or “computer scientist”, according to particular characteristics that are application-specific.

Overviews of methods for building user profiles are presented in [14] and [15]. User modeling issues and guidelines are presented in [8], concentrating on modelling of
user knowledge, plans, and preferences. This work focuses on stereotype (as opposed to individual) profiles. The need for a profile that supports reasoning is also stressed out in [14].

The goals listed above can be achieved through the use of ontologies. Ontologies in the form of hierarchies of user interests have been proposed in [9]. Gauch et al. [10] also proposed a system that adapts information navigation based on a user profile structured as a weighted concept hierarchy. The user may create his/her own concept hierarchy and use it for browsing web sites. Razmerita et al. [11] presented a generic ontology-based user modelling architecture applied in the context of a Knowledge Management System.

In the field of ontology design, efforts have been made by several research groups to facilitate the ontology engineering process, employing both manual and semi-automatic methods. Semi-automatic methods focus on the acquisition of ontologies from domain texts. In [2], for example, a framework is proposed with this objective, incorporating several information extraction and learning approaches. Comprehensive surveys of existing methodologies can be found in [3] and [4]. Throughout the ontology creation process, the designers may take into account a set of design criteria, such as clarity, coherence and extensibility [5].

III. ONTOLOGY CREATION ISSUES

As seen form the previous section, ontologies, as a notion, have already been introduced in the context of user profiling. The ontologies used however in relation with user profiles are mostly limited to taxonomies of user interests. Bearing in mind that for most applications profiling is not restricted to user interests but also encompasses other user characteristics (such as education, expertise and computer literacy level), our approach is to incorporate them in a user profile ontology. This section, after a brief definition of the ontology concept, presents description of our method for creating the ontology.

A. Ontology Definition

As defined in [1], an ontology is a formal explicit description of a domain, consisting of classes, which are the concepts found in the domain (also called entities). Each class may have one or more parent classes (is-a or inheritance links), formulating thus a specialization/generalization hierarchy; a class has properties or slots (also called roles or attributes) of a specific type describing various features of the modeled class, and restrictions on the slots (also referred to as facets or role descriptions). Classes may have instances, which correspond to individual objects in the domain of discourse; each instance has a concrete value for each slot of the class it belongs to. An ontology together with a set of individual instances of classes constitutes a knowledge base.

B. Ontology Creation Resources

For the creation of the ontology we adopted a top-down approach, firstly selecting important general concepts, which were later enriched and specialized. The focus of the ontology is the static profile of the user i.e. his/her more or less permanent characteristics, and not the dynamic ones (like his/her current position).

Gruber’s design criteria [32] (clarity, coherence, extensibility, minimal encoding bias, minimal ontological commitment) were taken into account during the creation process. In order to create a simple yet adaptable model, profile information models maintained by various applications, like [33] were examined and general ontologies like the ones presented in [29] were considered.

At this point no automatic concept extraction has been used, as the information in the available profile models did not contain high level concepts but rather instances of possible concepts and slot names. Consequently, the ontology designers team proceeded by analyzing the semantics of the profile models and suggesting concepts that would adequately model them. Table I exemplifies this procedure by presenting how certain information from the ICQ [33] user profile were mapped to ontology constructs.

<table>
<thead>
<tr>
<th>Property Modeling in the Profile Ontology</th>
<th>ICQ Profile</th>
<th>ICQ Profile</th>
<th>Modeling in the Profile Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place of Birth</td>
<td>City</td>
<td>Slots of the “Person” class</td>
<td></td>
</tr>
<tr>
<td>Personal Info</td>
<td>Home Page</td>
<td>Slot of the “Person” class</td>
<td></td>
</tr>
<tr>
<td>Zip</td>
<td>State</td>
<td>Not necessary, may be calculated by the date of birth</td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>Country</td>
<td>Slot of the “Person” class</td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>Date of Birth</td>
<td>Not necessary, may be calculated by the date of birth</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Zodiac Sign</td>
<td>Could be added as Instances of the “Education” class</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Languages</td>
<td>Could be added as Instances of the “Education” class</td>
<td></td>
</tr>
</tbody>
</table>

User profile models sourced from bibliography were also considered and concepts from these were appropriately adapted and included in the ontology. Information from bibliographic sources was exploited for selecting the basic set of upper level classes. Tazari et al [17] suggest the following concepts as important for user profiling: User identity, characteristics, capabilities, universal preferences, state of the user, application-specific preferences. Other concepts like current activity, current terminal, location, motion state and orientation are mentioned, but have not been included in this ontology as they refer to a dynamic profile. They also propose a group of parameters concerning personal information (name, birthday, address, bank account, and credit card), general characteristics (physical factors: weight and height, physical disabilities and abilities: reading, speaking and writing), education, occupation, interaction-related information, expertise and user state.

Interests ([9], [10], [19], [23]) and preferences ([8], [23]) are considered of particular importance for most applications that incorporate profiles. Interests are in some cases organized in hierarchies of concepts [9], [10]. Abilities, both physical and mental also seem to be relevant [24]. For
example, the ability of a user to mentally rotate two- or three-dimensional objects affects the interpretation of a picture [16]. The gender factor also has been proven to affect the performance of different users while interacting with the same system [18].

User expertise, either computer-related or related to another domain is a concept necessary for many profiling applications [23]. Defining a universal and adequately objective expertise measure with clearly defined categories is not an easy task [25] and is out of the scope of this work. However, by studying the existing literature, properties relevant to user expertise and competence have been identified and included to the user model ontology. The ontology is described in the following section and is available in Protégé and RDF format in [22].

IV. ONTOLOGY DESCRIPTION

This section presents a brief description of the user profile ontology. The ontology may be extended through inheritance and the addition of more classes, as well as concept instantiation according to the needs of a specific application. As a result, it may be used for the representation of both stereotype profiles (i.e. user profiles that represent a specific user category, like “computer expert”) and individual ones.

This ontology presents information that is mostly static and permanent, its structure allows though for more dynamic characteristics (e.g. current position of the user when moving) to be easily incorporated. Moreover, the temporal aspect of some of the ontology classes has been taken into account. The ontology allows the existence of multiple instances of classes that represent characteristics that may change with the passage of time. These classes include a period representing the user’s life as well as a time period which represents the duration of that particular aspect. For example, a user may have had a “Contact” of type “friend” from 1989 to 2004. The domain of the slot “person” of the “Contact” class is the class “Person”. This way, relations between different users may be modeled as well.

“Interest”, “Preference”, “Ability”, “Characteristic” and “Thing” contain only three slots: “type”, “name” and “score” (or “value” in the case of “Thing”). “Thing” has two subclasses, “Living Thing” and “Non Living Thing” as modeled in the WORDNET ontology [30] [31]. In the case of interests, apart from the “type” slot, which is a String, a slot named “interest type” of type “Interest” has been added to allow the creation of interest hierarchies, as the ones suggested in [9] and [10]. Table III shows an example.

User expertise according to [26] may be defined as a combination of three dimensions: breadth, the extent or variety of different tools, skills and knowledge the user may possess, depth, the completeness of the user’s current knowledge of a particular domain, and finesse, which refers to innovativeness and creativity. Breadth and depth are developed over time through a combination of study and hands-on use, whereas finesse is more related to the user’s personality. These properties are included in the user ontology as slots of the “Expertise” class. The notion of “experience atoms” is introduced in [25]. They are defined as elementary units of experience as a result of activity in a particular domain. Experience atoms may be expressed in the user ontology as individual instances of the “Expertise” class.

Experience referring to the use of computers is very often related to duration and frequency of usage [28]. A questionnaire of perceived user expertise in a series of end user computing related sub-domains is used in [27] in order to calculate the expertise level of the user by the combination of the scores supplied by the user in each question. The idea of assigning a score or level to expertise is expressed through the “score” slot in the “Expertise” class.

It should be mentioned here that in the case of the “Expertise” class, the aim was to collect from the existing literature user characteristics that may serve as indications or factors during the assessment of the user expertise level. The definition of the expertise levels themselves and the expertise measures are application-specific and out of the scope of the
current work.

To sum up, the “Expertise” class has been created as a container for both expertise measures and expertise scores in order to accommodate the particular needs of individual applications that make use of profiling. The following section provides two examples of such applications, and how the user model ontology may be used in each case.

V. CASE STUDIES/EXAMPLES

In order to demonstrate the adaptability of the proposed user profile ontology, two case studies will be presented, one related to personalized, adaptive visualization and the other to personal information management.

A. Personalized Visualization

A context-related research is being developed in the framework of digitizing the Historical Archive of the University of Athens, Greece. The corpus of the above-mentioned archive is very large (more than 4,000,000 documents) and consists of documents issued in the University since its foundation (1837). Currently, the Historical Archive can be visited by anyone who is interested in searching for information relative to its contents. Visitors submit requests to the archive staff, which subsequently undertakes the task of locating the relevant documents and presenting them to the visitors.

In the above-mentioned framework, a novel information retrieval system is being developed in order to render the corpus available directly to its users. The user profile is used by the system for deciding which visualization suits the user best for the presentation of information, so as to be employed the next time s/he returns to search for information.

Users who come to the Historical Archive to retrieve information vary in multiple ways. For example, they have different educational levels, ranging from users who only completed elementary or secondary school to users who possess postgraduate degrees (MScs, PhDs) in various scientific subjects. They also have differences in their experience using the computer, ranging from those who are beginners to the ones very experienced and computer-skilled.

Regardless of their computer expertise, users have been found to have different ways for foraging the information they are interested in, which depend on individual differences, a major factor that influences the user profile. Apart from personal preferences and existing knowledge, cognitive abilities, specific aims and tasks to be solved, the gender, the age, the profession and the living environment of the user constitute properties of individuality, which is a fundamental part of the user profile. Moreover, the steps a user performs while trying to reach the information needed, the -so called- “history” of the user, plays an important role in sketching out his/her profile. Important properties related to the user profile are listed in Table IV.

<table>
<thead>
<tr>
<th>TABLE IV: PROPERTIES OF THE USER CONTEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User Context Property</strong></td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>University relation/role/title</td>
</tr>
<tr>
<td>Information Retrieval Knowledge</td>
</tr>
<tr>
<td>Aim</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Profession</td>
</tr>
<tr>
<td>Living environment</td>
</tr>
<tr>
<td>Abilities</td>
</tr>
</tbody>
</table>

Apart from the user profile, the visualization environment proposed for the information search in the Historical Archive takes into consideration both the system context and the document collection context. Tables V and VI list representative properties of the system and document collection contexts.

In order to select the most prominent visualization method for each case, the values for all properties of the user, system and document collection context are computed and, subsequently, the computed property list is matched against the feature profile of each available visualization method. Matching is performed through a set of rules, with each rule indicating whether a particular feature of a visualization method is considered to be helpful, impeding or neutral for a specific context characteristic. For example, the rule

(user_context, spatial_memory, yes) =>

(metaphor, landscape, 70)

states that if a visualization is to be performed for a user having spatial memory, then methods employing the
landscape metaphor are considered as “strong candidates” (as indicated by the score value 70), since the particular user’s ability allows him/her to exploit the visualized items’ spatial placement so as to perceive the visualization more effectively [20]. Score values are drawn from the range [-100, 100] with positive values being used for “helpful” features and negative values being used for “impediments”. For more information on the visualization method selection algorithm, the interested reader is referred to [21].

<table>
<thead>
<tr>
<th>System Context Property</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input devices</td>
<td>Mouse, Keyboard, Joystick, Specialized input devices (3D mouse, glove, etc.)</td>
</tr>
<tr>
<td>Output devices</td>
<td>2D monitors, 3D monitors, Head mounted displays</td>
</tr>
<tr>
<td>Other hardware equipment</td>
<td>Processor, Memory, Graphics</td>
</tr>
</tbody>
</table>

The user profile is represented in the system via the user profile ontology. An example of such a user profile, also available as an Instance in the profile ontology available in [22], is the following.

A female (gender) 20 years old (age) student (profession), Maria Papadopoulou (name), wants to retrieve information about the Department she is studying in. She wants to write an article to publish in the Department’s newspaper (activity) about the evolution of the Department of Informatics and Telecommunications as far as teaching in it is concerned. She is very experienced in using the computer (computer expertise: high) and in searching for information, mainly though the Internet (web search expertise: high). She uses a PC with traditional I/O devices (system context: mouse, keyboard, 15”, 2D monitor). She has already visited the Historical Archive in the past and according to her previous interaction she likes exploring 3D environments (registered preferences: history).

The system collects the above contextual information and matches it against the features of all available visualization methods. The method found most appropriate for the contexts at hand is selected to perform the visualization.

B. Personal Information Management

As part of the EU DELOS Network of Excellence [6], the TIM project (Task-centered Information Management) is studying the potential for users to store files, email, etc., indexed by personal ontologies. Design and implementation of the prototype tool OntoPIM [7] is still in early stages but several key issues are already apparent.

OntoPIM relies on the use of a Personal Ontology that describes the user’s domain of interest. The ontology is personal in the sense that it reflects the user’s view of the domain(s). It is used to assign semantics to the information contained in the user document repository in order to be able to retrieve this information more easily. With the use of the Semantic Save, it provides the user the possibility to store any object of interest according to its semantics, i.e. to relate it to the concepts of the Personal Ontology, where an object may be an e-mail, a document, a picture, or any other type of data.

<table>
<thead>
<tr>
<th>Document Collection Context Property</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories of documents</td>
<td>Criterion of categorization, Number of elements, Relation between categories</td>
</tr>
<tr>
<td>Text documents</td>
<td>Full text, Image, Manuscript only, Meetings’ minutes</td>
</tr>
<tr>
<td>Metadata</td>
<td>Author, Title, Type, Date of issue, Department of issue, Keywords, Categories</td>
</tr>
<tr>
<td>Collection origin</td>
<td>Static, Dynamic</td>
</tr>
</tbody>
</table>

Bearing in mind that every user has his/her own domain of interest, personalization issues are very relevant here. OntoPIM proposes the creation of an initial library of ontologies suitable for various user groups and domains. These ontologies should be created beforehand after an elaborate user study.

There are cases, of course, that some final tuning will be necessary to adjust the selected ontology template in order to accurately reflect user characteristics and interests. This can be done both manually (by the user) and automatically (by the system), employing a user profiling mechanism. The ontology presented in this work may be used to model the personal ontology core.

The profile ontology may, to some extent, be populated automatically with user information that is available within the file system. This may include:

- **Chosen language and time-zone.** These could give information about the user nationality and living conditions. Dialing codes and IP addresses can be also used for determining the users’ location.
- **Current file structure.** If the user has created a more elaborate file structure than that already provided by the operating system to store his/her files, then it could also be a source of new concepts. A dictionary of synonyms could be used here to make the matching more effective. The user could also be prompted to indicate folder structures that contain documents relevant to his/her interests or activities. For example, if the user has a folder named “Articles” with sub-folders like “basketball” or “gardening” used to further categorize the documents, these concepts may be used to populate the “Interest” Class of the profile. The file content could be used as well to support this concept extraction.
- **Address Book.** The system can also scan address books in order to retrieve contact information and populate the “Contact” class with instances.
- **Calendars and to-do lists may be used to identify user activities.**
• The web cache and bookmark/favorites structure could also be a possible source for deriving interests and preferences. The user can also be prompted to provide information concerning his/her personal data, interests, preferences, contacts, etc.

As an example available in [22], Elias Daradimos (name), 29 years old (age/date of birth) is a network administrator (Profession). The OntoPIM user profile extraction mechanism identifies him as a resident of Athens, Greece (Living Conditions) and creates a list of his contacts names and e-mails (Contacts). By examining application preferences and system settings, asking him to fill in a form, and by scanning indicated parts of the file system, the mechanism concludes that the user is interested in electronics, airplanes, scanning indicated parts of the file system, the mechanism and system settings, asking him to fill in a form, and by e-mails (Contacts). As an example available in [22], Elias Daradimos (name), 29 years old (age/date of birth) is a network administrator (Profession). The OntoPIM user profile extraction mechanism identifies him as a resident of Athens, Greece (Living Conditions) and creates a list of his contacts names and e-mails (Contacts). By examining application preferences and system settings, asking him to fill in a form, and by scanning indicated parts of the file system, the mechanism concludes that the user is interested in electronics, airplanes, scanning indicated parts of the file system, the mechanism and system settings, asking him to fill in a form, and by e-mails (Contacts).

VI. CONCLUSIONS AND FUTURE WORK

This work is an attempt to create an ontology that incorporates concepts and properties used to model the user profile. Existing literature, applications and ontologies related to the domain of user context and profiling have been taken into account in order to create a general, comprehensive and extensible user model. The model, available through [22], is also presented through two examples in different areas, personal information management and adaptive visualization.

As this model focuses more on static user characteristics, it is our future aim to study the incorporation of dynamic and temporal characteristics in order to cater for a wider range of applications that include profiling. Furthermore, the acquisition of the profile properties for individual users through questionnaires is investigated, in order to complement the user profile ontology with a means to populate it.

REFERENCES


domak/arbeitschwerpunkte/acm/midas/root.html


Ontology personalization context-aware user profile. This is a preview of subscription content, log in to check access. Preview. Å