

A User Profile Definition in context of recommendation of Open Educational Resources. An approach based on Linked Open Vocabularies

Janneth Chicaiza, Nelson Piedra and Jorge Lopez-Vargas
Departamento de Ciencias de la Computación
Universidad Técnica Particular de Loja
Loja 1101608, Ecuador
Email: [jachicaiza, nopiedra, jalopez2]@utpl.edu.ec

Edmundo Tovar-Caro
Departamento de Lenguajes y Sistemas Informáticos
Universidad Politécnica de Madrid
Madrid, Spain
Email: etovar@fi.upm.es

Abstract—Open Educational Resources include a diverse range of materials making it the most representative icon arisen within the Open Content movement. Users who access and use OERs could be classified into one of these three groups: instructor, student and self-learner. To provide personalized lists of OERs according to the user profile and personal preferences, the user should be characterized by an open and scalable model. In this paper, an open linked vocabulary is proposed to describe user profiles of the open educational resources, which take into account the challenges and opportunities that an open and extensible platform as the Web can provide to learn about the OER users, and from this knowledge, offer the most appropriate resources.

I. INTRODUCTION

Different projects arising under the philosophy of Open Access (OA) have helped to facilitate and to make accessible to all users the online learning. The OA initiatives framed in the educative context have "significant implications, and allows distance educators to play an important role in the fulfillment of the promise of the right to universal education" [1]. Open Educational Resources (OER) include a diverse range of materials making it the most representative icon arisen within this movement.

Considering the personal interests of people that access to the open educational resources, OERs are not only useful for students of educational institution, but also they are used by self-learners or who are part of a kind of informal learning.

The most used tools to find the resources in the Web are the search engines. However, due to the huge amount of available resources, and because the traditional research systems does not consumption the users knowledge, thousand or millions of results are returned. The failure to find relevant material to support a teaching and learning process can be an inhibiting factor in the process of formal or informal learning.

Unlike the search engines, the recommender systems filter the resources taking into account the users interests, and they are able to give a better user support when the users are trying to find the resources they need. The issue is that these systems to be effective require big amounts of data; in closed or specific e-learning environments there is the possibility of modeling a wide range of users variables such as; preferences, styles, goals, learning aims between others. Considering that one of

the OER users profiles is the self-learner or people who do not belong to a formal education system, thus, the collection of interaction information that helps to build a user profile is limited.

In this paper, an open linked vocabulary is proposed to describe user profiles of the open educational resources. As is known, the platform on which OER users are identified, participate and contribute is the Web; in this regard, there are challenges and opportunities that an open and extensible platform can provide to learn about the OER users, and from this knowledge, offer the most appropriate resources.

The user's representation is based on computable schemes, and it fed on Web data structured and organized through systems of open knowledge. The usage of vocabularies and formal languages of the Semantic Web enables to reuse and interoperability of user's data between several applications. Moreover, the modeling of user interests through knowledge organization systems will allow to implement query and support services so that they can understand the topics and relations around the domain and finally, people will know how to better conduct the search of material.

Continuing with this paper, the main proposal to describe Web users, and their main constraints when modeling heterogeneous users in a open learning environment are put forward in the section II. The design and preliminar validation of the User Profile for OER recommendation tasks are put forward in the section III, and the section IV. Finally, in the section V, the conclusions and future work are appeared.

II. THEORETICAL BACKGROUND

A. Standards and data models to represent users

In order to represent a learner, a widely know specification is PAPI Learner Standard (Public And Private Information) proposed by the IEEE Learner Model Working Group. PAPI was launched in 2001 and has served as a reference to some proposals as [2], [3].

PAPI defines different elements to record descriptive information about students' knowledge and preferences, as well as personal contact information, security settings and general privacy, among others. The draft 8 from the specification lets

us to specify different views of learners' information, this way, those using this model can choose the parts that are appropriate for their application.

Moreover, the consortium IMS GLC (IMS Global Learning Consortium) created the specification, IMS LIP (IMS Learner Information Package)[4]. An initial draft specification appeared in 2001 and, in 2005 the final specification was formalized.

The IMS LIP standard has been used in the construction of student models and some applications like: Elena¹, L4All², EPET³, Europass CV Application⁴, among others. The specification divides the learners information into eleven categories or data structures, these categories include: biographical and demographical students data, objectives and competencies, accessibility, degrees, activities and relationships.

One of the cons of these two specifications is that they are not extensible, in other words, they have not evolved to consider features and user preferences that arise in an open online learning environment. Another problem identified in [5] is that IEEE PAPI and IMS LIP do not facilitate the sharing and reuse of profiles stored on different servers.

Nowadays, in order to improve interoperability and scalability of user profiles between different systems, most proposals are based on representations based on ontological models. There are implementations of models using OWL or RDF(S) languages. Among the related works about semantic models proposed to represent the user profile, we can highlight the following:

Among the high-level specifications to describe Web users, there are: i)FOAF (Friend Of A Friend)⁵, and ii) vCard⁶ standard proposed by the IETF and currently published in RDF/OWL format by the Semantic web Interest Group. Both FOAF and vCard propose identification metadata and contact as: names, age, nickname and email. FOAF also allows the description of some professional information and user interests. Finally, we can name the collection Person⁷ of the Schema.org specification, driven by major Web search engines.

Another proposal that defines the profile of a Web user shows up in [6], although in this case, the profile is projected in order to incorporate it in recommender systems based on content, for it models the profile based on words of the pages visited by the user.

For purposes of personalization of content and specifically for adapting Web content, the ontology GUMO (General User Model Ontology) [7] defines individual characteristics of a person, as: physical and emotional states, characteristics and personality; in addition, in [7] the Markup Language (UserML) is proposed to manage the communication between different applications to share the user model. Another proposal to facilitate the development of e-learning systems is presented in

[3] custom, where a model student profile based on the IEEE PAPI specification is proposed.

To provide personalized access to documents, [8] defines a domain ontology that organizes the Web documents so that each document is classified as a particular concept in the domain of ontology, then, according to the history of user navigation is the interests of users are collected and recorded in the user model.[9]

Another group of user profiles, focus on ensuring portability and reuse of profiles between systems that adopt the same metadata schema: [2], [5], the LLO ontology (Learner Ontology for Planning of Lifelong Learning), the ontology U2MIO [10].

Although, the proposals mentioned try to model the users profile in a wide spectrum of dimensions, in a Web context, by the dynamics of users data, it seems to be useful to focus in such aspects that in a mid-term will capture the nature of the users interest. To sum up, in order to model the users interest and preferences, the authors have found the following proposals.

- A model to describe the users preferences (local and global) expressed in terms of resources and domain of interests; the learning of users interest, getting trough data mining methods is based on the characteristics or attributes of an object which is graded by a user. [11]
- The experimental vocabulary WI (Weighted Interest Vocabulary) which was developed in 2009 in order to cover the requirements of the Project NoTube; WI models the users preferences inside an environment o determined context; the relevance of significance for the user can be established through weight which value can change depending of the context.
- The STOUPE ontology (Spatio-Temporal Ontology of User Preference) defines three types of preferences (positives or negatives) and the dimensions that can affect them such as activities, devices, localization, time, and event.
- CC/PP (Composite Capability/Preference Profiles) allows describing a devices capacities and the users preferences; this information can be used to guide the content adaptation that is presented in a determined device.
- The FRAP ontology (Framework for Ratings and Preferences)⁸ allows the open interchange of the users preferences. The framework is independent from the domain. [12]
- The same organization that developed FRAP, which is called CTIC, proposes the ontology RECO⁹ that differs in the objective and in the proposal of the ontology FRAP. RECO represents a ordered list of elements given by a recommendation system, whereas FRAP captures the users preferences. The RECO ontology defines a vocabulary to represent preferences and ratings as restrictions through a RDF graph.

¹elena-project.org

²lkl.ac.uk/research/l4all/

³eportfolios.ac.uk/ePET/

⁴europass.cedefop.europa.eu/

⁵http://xmlns.com/foaf/spec/

⁶http://www.w3.org/TR/vcard-rdf/

⁷http://schema.org/Person

⁸http://purl.org/frag#

⁹http://vocab.ctic.es/reco/reco.owl

TABLE I. REQUIRED CHARACTERISTICS OF THE USER PROFILE

Feature of the OER User	Current proposals	Requirements of the User Profile
The user can have different levels of knowledge. On the other hand, one of the forms that a user uses to express their learning information needs is through interest expressed as concepts.	Works such as [12], [10] define a model of web user independent from the domain, i.e. they are re-usable in any field; but, they are too general to determine the knowledge of the user in a determined subject; this can affect the performance of the personalization task.	The model must capture the current state of knowledge of a user in a specific field and it must be able to be used to interpret with open knowledge repositories; so that, it can enrich the user profile from these sources. Therefore, recommender systems can take advantage of the potential of the semantic relations, which are expressed, in open systems of organization of knowledge.
An OER user does not center his or her participation in a specific learning system or in a unique material repository; actually, he or she will not be register in any platform.	A considerable number of approaches for users modeling as [14], [11] have focused on the identification of profiles of teachers and students based on their preferences and needs as users of a determined system or a controlled environments.	The model must be flexible to characterize users who little is known, i.e., the profile should model the user preferences depending on the educational material distributed in different repositories, in order to take advantage of the OERs metadata.
OER user can participate in learning communities that emerged spontaneously. Since it has become massive the use of social networks, people use these services to share material, data and preferences.	Approaches as [15] y [16] are experimenting with public data from social sources since they provide an acceptable way to identify user interests. To infer user behavior, techniques based in learning machine are applied; they can be intensive in data processing.	The model should distinguish user profiles from heterogeneous systems (including social services); therefore, it must capture the user data dynamics and model the shared communities interests of users using techniques applicable to the Web. [9]

Another dimension in which some founded Jobs are centered is the characterization of the user as a part of a network. In [13] is proposed an approach for a multi-layer building of communities of interest (COI), which is based on the analysis of the users individual preferences that are described in the profiles. The method builds users profile interests based in specific concepts with the purpose of find similarities between users; the profiles are divided in groups of interest, based on this, some layers of COI are found. Another work that deserves to be named is one presented in [9] in which a model based interest ontology is proposed, it can be used to determine the close neighbors by calculating the semantic similarity of interest models of two users.

B. Characterization of the User of the Open Educational Resources

In this section are analyzed the main conclusions of the present state of users profile modeling against the nature and characteristics of an open learning environment user, supported by oERs; as well as how this relation affects the recommender systems which should provide customized listings according to the users interests and preferences.

From the information in Table I, the authors can conclude that a OER users profile should be flexible; so, it should allow represent users under contrary scenarios: anonymous users or reserve ones with little known information vs. active users in

the web, generally with presence in networks or communities. Also, the model should be extensible and interoperable, so that, it can connect with other representations in order to take advantage of the large amount of open data that are published in different knowledge systems and in OERs open repositories of metadata.

III. DESIGN OF THE OER USER PROFILE

A. User profile requirements

In order to ensure the successful our proposal and according the findings that were found in the literature review, the model of user of OER should keep the following design principles:

- 1) Based on semantic structures, i.e., the model must be able to be combined with controlled vocabularies and open systems of knowledge organization. This feature enables the enrichment and inference of new user profile data and above all enables the organization and classification of users according to common interests. Taking advantage the potential of different knowledge domains they are encoded on the web as linked data. The knowledge about the domain of interest to the user, allow to do better recommendations according to the level of knowledge the user has on an issue.
- 2) Support distributed information: the model should facilitate the representation of user profiles from heterogeneous and distributed systems, which are enabled to share information about learner or teacher.
- 3) Privacy and data protection: In order to ensure the integrity and privacy of user data, the attributes that the user has decided to share in your public profile have to be considered. You must also take into account other information, both personal and critical, which should not be shared with others external agents.
- 4) Personalization. The data model of OER's user must be flexible to support adjustments in two ways: i) extension through the addition of new user features and its environment, and ii) personalization according to specific purposes or environments.

B. User Profile Dimensions

The organization of user characteristics and the environment in which he/she interacts, encapsulating the attributes in different dimensions allows to build a interoperable model of the user, so different applications can share the model with different purposes. The Figure 1, shows the different categories of data that could characterize an OER's user.

In addition, as shown in Figure 1, different categories of user data may influence others: i) due to the exercise of academic and professional activity, the user can acquire different skills and abilities; ii) the previous experience and the current user activity can determine their learning objectives and motivations; iii) finally, i) and ii) more the context information can influence the interests and current decisions of the user.

Although the user profile can be used for different purposes, on this paper, the user preferences constitute the fundamental dimension of our design. The following reasons justify this decision: i)) the users background both academic as

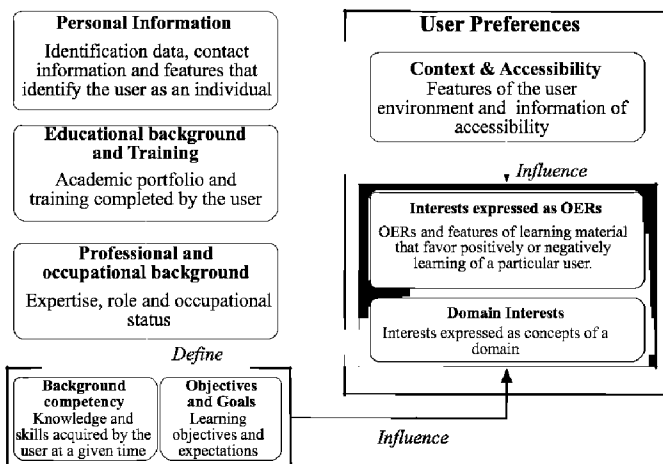


Fig. 1. Dimensions of the User Profile

professional have an important role in customization tasks oriented to content adaptation [17], this differs from our primary purpose, the recommendation of OERs. ii) The collection of information about the objectives and expectations of learning, in an open learning environment can be complex or limited.

C. User's Preferences

A preference is a mental state of an individual respecting from a subset of items from a universe of alternatives.

To provide lists of relevant OERs according to the interests of a particular user, it is proposed to specify a grade or rating for each user preference through a quantitative or qualitative scale; the value that is assigned to represent the importance of this element to the user.

In order to support different methods of filtering information, the user model has been designed to withstand at least three types of user preferences:

1) *User Interests expressed as concepts of a domain of knowledge*: The interest model of the user, being based on thesaurus or other knowledge organization systems, it can be used to exploit hierarchies of concepts and provide the necessary support the user can define in more precise interests. [9]

Example 1: Topic recommendation. An user (:User1) has expressed interest by the subject of Software Engineering (:SoftwareEngineering). Open Sources of Knowledge as DBPedia¹⁰ enable the automatic extraction of the subtopics of this discipline, which can be recommended to the user to refine their interests, if it is required.

```

IF
  [type (:Software_Engineering , :Concept) AND
   interest (:User1 , :Software_Engineering) AND
   narrower (:Software_Engineering , :Software_design) AND
   narrower (:Software_Engineering , :Software_requirements) AND
   narrower (:Software_Engineering , :Software_architecture)]
THEN
  [recommends (:User1 , :Software_design) OR
   recommends (:User1 , :Software_requirements) OR
   recommends (:User1 , :Software_architecture)]

```

¹⁰<http://dbpedia.org/About>

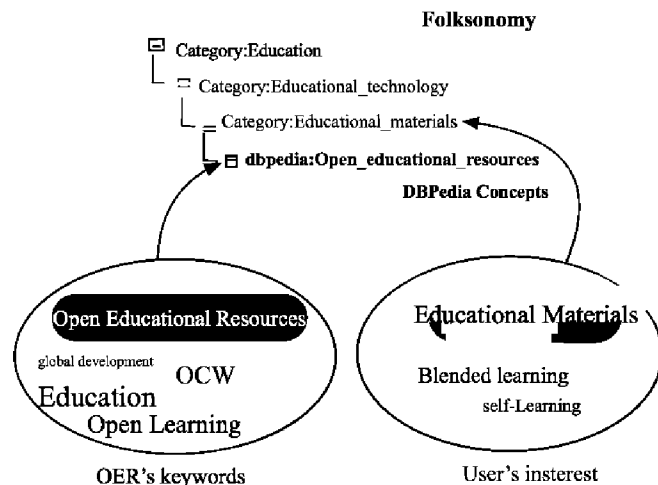


Fig. 2. Mapping between OER concepts and User interests

Another application of interest expressed as linked concepts by semantic relations is the location of the resources related to given subject and their variations or derivations.

Example 2: OER recommendation. In Figure 2, it can see that it is possible to map user interests expressed as concepts and keywords of OERs, thus, systems can provide recommendations for resources without them having the same words of the user. The same approach can be used to find similar users and provide recommendations based on collaboration.

2) *Documents and/or preferred attributes of OERs*: The user's preferences for a particular group of OERs, or their features that describe it, they can be used to apply approaches based on content filtering, in order to take advantage of the large amount of metadata exposed by open repositories of learning material.

3) *Demographic information and accessibility*: This information is useful to filter the most appropriate items of information according to user context. This dimension includes: location, language, age, and information accessibility and user device used to access the educational content. The PAPI [18] specification also adds metadata such as learning style and physical limitations.

By including this information in the user profile, systems sensitive to the context could be created according to the particular tastes of each user; for example, for a student who manages one language, it may result a priority to find the resources that are in her/his mother tongue.

D. Conceptual Model of the User Profile

The design of the OER user profile begins with the construction of the conceptual model. Therefore, at this point, our goal is to identify and describe the terms of the user model to improve the discovery and recommendation of open educational resources

One of the fundamental tasks of the conceptualization of the model is building the terms of the glossary, which specifies the metadata and user variables and concepts related. Table II

TABLE II. EXCERPT OF THE TERMS GLOSSARY

User Concept		
Metadata	Description	Range of values
names	User names	Text
description/summary	Resume or information descriptive of the user	Text
e-mail	Email preferred for the user	email
Localization	User location, i.e. a city, region or country	URL
Language	Source language of the user	URL
Web page	URL of the user's home page	URL
Profile Concept		
Metadata	Description	Range of values
Identifier	User profile ID	URL/string
User	Social account user	URL
Source	Source address of the user profile information	URL
Creation date	Date of the profile creation	Date
Published date	Date of the profile published	Date

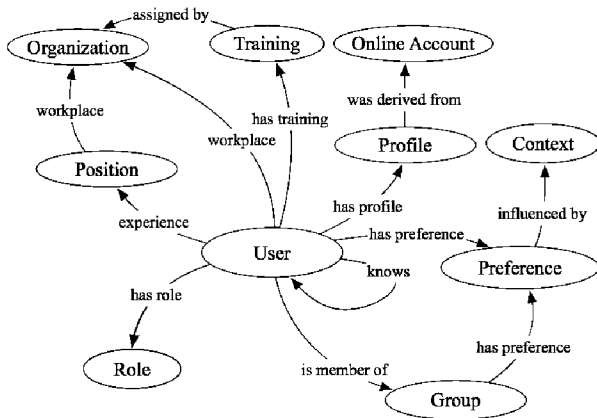


Fig. 3. Conceptual Model. A top level view

shows an extract of the terms identified for concepts User and Profile shown

As shown in Table II, to build the glossary terms may appear new terms, such as Concept Profile. This concept includes the information necessary to identify the web profile of an user, and it is used to identify the source and provenance of the user's information.

From the glossary of terms, the next step was to build the diagram of binary relations between concepts. In the Figure 3 it is presented the high level conceptual model resulting from this process.

E. Reuse of ontological and non-ontological resources

The reuse of ontological and non-ontological resources reduces the time for creating an ontology and associated costs; also it contributes to the quality of ontology. citeVillazon2011

From the analysis of the proposals referred to in paragraph 2.1, it was selected those models that offer greater coverage of the terms identified in the Glossary of Terms and ensured a high probability of interoperability and semantic consistency with the conceptual definition of the model chosen. Main vocabularies that are selected are listed:

- Simple Knowledge Organization System (SKOS)¹¹ vocabulary used to organize the users interest defined by concepts.
- FOAF¹² y vCard¹³ are used to describe the basic information of people and organizations.
- Dublin Core (DC)¹⁴ y Dublin Core Metadata Initiative (DCMI) Metadata Terms¹⁵ provide needed metadata to describe the users preferences depending on of OERs.
- VIVO¹⁶, ontology for describing different concepts in academic and scientific domain.
- Open Provenance Model Vocabulary¹⁷ to describe the origin of the data from Web users.
- Schema.org¹⁸, set of metadata schemes that improve the discoverability of an appeal by Web agents.

Attributes, properties and entities that are not considered in these vocabularies, were designed and created as part of our proposal.

IV. REPRESENTATION AND VALIDATION OF THE USER PROFILE

Once built the ontological model, various user profiles were defined in order to determine the capacity of representation of the model designed. Namely, three user profiles are considered: i) a person who regularly uses digital educational resources to support their work and acquire the skills required in their occupational tasks; ii) a student of an educational program, who want to supplement the study material provided in class; and iii) an instructor who want to find the material related to a particular subject.

The first type of user could well correspond to an self-learner profile. In e-learning, self-learner is an active user on the web that is not necessarily enrolled in an educational institution and supports their learning processes using resources found on the Web.

Then, certain characteristics that could distinguish to a self-learner when interacting and search online material are described: i) is professional and entrepreneur, ii) generally set their own learning objectives; iii) it subscribed to some social networks such as Facebook, twitter and LinkedIn; iv) regularly, publish and share relevant material that he/she has been used in his/her training process; v) his/her interests are related to different knowledge areas and topics; vi) dominates

¹¹<http://www.w3.org/2004/02/skos/>

¹²<http://xmlns.com/foaf/spec/>

¹³<http://www.w3.org/TR/vcard-rdf/>

¹⁴<http://dublincore.org/documents/dces>

¹⁵<http://dublincore.org/documents/dcmi-terms/>

¹⁶<http://vivoweb.org/ontology/core>

¹⁷<http://open-biomed.sourceforge.net/opmv/ns.html>

¹⁸<https://schema.org/>

- [11] T. Horvth, "A model of user preference learning for content-based recommender systems." *Computing and Informatics*, vol. 28, no. 4, pp. 453–481, 2009.
- [12] L. Polo, I. Mínguez, D. Berrueta, C. Ruiz, and J. M. Gómez-Pérez, "User preferences in the web of data," *Semantic Web*, vol. 5, no. 1, pp. 67–75, 2014.
- [13] I. Cantador, P. Castells, and A. Bellogn, "Modelling ontology-based multilayered communities of interest for hybrid recommendations," in *IST Workshop on Adaptation and Personalisation in Social Systems: Groups, Teams, Communities (CORFU)*, 2007.
- [14] S. B. Dias and J. A. Diniz, "From blended to inclusive learning: Accessibility, profiles, openness, and higher education," *j-jucs*, vol. 19, no. 18, pp. 2722–2742, 2013.
- [15] P. Lops, M. de Gemmis, G. Semeraro, F. Narducci, and C. Musto, "Leveraging the linkedin social network data for extracting content-based user profiles," in *Proceedings of the Fifth ACM Conference on Recommender Systems*, ser. RecSys '11. New York, NY, USA: ACM, 2011, pp. 293–296. [Online]. Available: <http://doi.acm.org/10.1145/2043932.2043986>
- [16] Q. Wang and H. Jin, "Exploring online social activities for adaptive search personalization," in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, ser. CIKM '10. New York, NY, USA: ACM, 2010, pp. 999–1008. [Online]. Available: <http://doi.acm.org/10.1145/1871437.1871564>
- [17] P. Brusilovsky and E. Milln, "The adaptive web," P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Heidelberg: Springer-Verlag, 2007, ch. User Models for Adaptive Hypermedia and Adaptive Educational Systems, pp. 3–53. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1768197.1768200>
- [18] "Draft standard for learning technology - public and private information (papi) for learners (papi learner)," CEN WS-LT Learning Technology Standards Observatory, Tech. Rep., 2001. [Online]. Available: <http://www.cen-ltso.net/Users/Pages/Content.aspx?content=136>