A (multi'*domain*'sional) Scrutable User Modelling Infrastructure for Enriching Lifelong User Modelling

Demetris Kyriacou, Hugh C Davis, and Thanassis Tiropanis

Learning Societies Lab
School of Electronics and Computer Science, University of Southampton
Highfield, Southampton, SO17 1BJ, United Kingdom
{dke06r, hcd, tt2}@ecs.soton.ac.uk

Abstract. A Scrutable User Modelling Infrastructure (SUMI) is proposed in this paper, which focuses on three key Lifelong User Modelling (LUM) requirements: interoperability, scrutability and privacy. SUMI attempts to enrich the LUM vision by examining two unexplored "pieces of the puzzle": the social e-networking and e-commerce domains. SUMI allows users to gather their various (lifelong) models which they hold with several providers from the social e-networking and e-commerce domains, passes the control to the owners of the user models by providing scrutability and user-controlled privacy privileges, and enables exporting of these user models, or parts of them, towards educational personalization systems, after direct consent of the users-owners. Up-to-date results highlight users' approval of six proposed scrutability and user-controlled privacy privileges which were presented during SUMI's first evaluation, while data portability initiatives from providers of the two examined domains, combined with Semantic Web technologies, constitute the backbone of the proposed solution.

1 Introduction

Lifelong User Modelling (LUM) could be a critical element in enabling lifelong personalized pervasive learning. Although the LUM vision is technically still not possible, it is the way forward. It is worth finding solutions to overcome the various challenges in this area as there are potential benefits, especially in terms of effective sharing of user models with educational systems and between people, while allowing the owners of these models to scrutinize the modelling process [1]. Two unexplored, missing, pieces of the LUM big picture are the social e-networking and e-commerce domains. A rich set of user information can be found in social networking sites like Facebook, LinkedIn, and MySpace, and in e-commerce megastores like Amazon and eBay. Surprising, little of these current rich sources of information are being harvested and integrated into the LUM vision. The potential benefits such information can bring to the table seem to deserve a closer look. What if we could enrich the picture in LUM, by modelling our every day (life-time) interactions with various services from the social e- networking and the e-commerce domains? What if we could add these pieces to the LUM puzzle, resulting in richer user models for -potentially- improved personalization?

My research focuses on identifying the requirements for a Scrutable User Modelling Infrastructure (SUMI) which enables gathering of user models constructed from our daily (lifelong) interactions with services from the social e-networking and e-commerce domains and exporting these models towards educational personalization systems, after direct consent of the users-owners. As the name reveals, special emphasis is given in making such an infrastructure as scrutable as possible which also consists implementation of -user-controlled- privacy enhancements. Following the LUM "rules", SUMI's goal is to pass full control to the user-owner to determine how his/her information will be used, from the moment it is imported in SUMI until it has been exported towards educational personalization (and not only) systems.

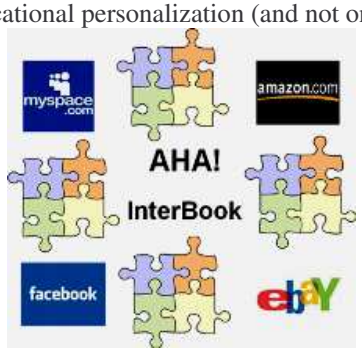

Fig. 1. The missing pieces of the LUM big picture

2 Key Requirements for Lifelong User Modelling

Lifelong User Modelling: LUM was introduced in an attempt to model users' daily - lifelong interactions with several services on the Word Wide Web (WWW) while offering to the users the ability to scrutinize and control the whole personalization process [2]. This huge fast-growing collection of user information can offer potential benefits when suitably used by adaptive educational systems, such as AHA [3], for potentially pervasive lifelong personalisation services. Research in this area has revealed that LUM can play a significant role in setting the foundations for a personalised lifelong learning vision. Offering scrutability privileges to the owners of these models enables user control of how the modelling and following personalisation procedure is conducted, while special consideration for user-controlled privacy empowers users to express their preferences regarding how their data will be retrieved, stored and processed [4]. While the requirements and standards for enabling LUM go beyond the three issues of interoperability, scrutability and privacy, SUMI has focused on these important "ingredients" while investigating the possibility of enriching LUM with sets of user data from the two examined domains.

Interoperability: Interoperability is a crucial issue in the area of UM and can be described as "a condition that exists when the distinctions between information systems are not a barrier to accomplishing a task that spans multiple systems" [5]. With the introduction of the Semantic Web [6] new technologies have been proposed for enabling interoperability across domains. Exchanging user profiles across various

sources in pervasive environments can not be achieved if explicit and widely accepted protocols are not developed and adopted, which will allow description, discovery and exchange of user models coming from several domains, stored in various systems, written in different languages and for different platforms [5].

Scrutability & User Privacy: Offering scrutable solutions when gathering user models in pervasive and ubiquitous environments is part of the LUM vision which focuses on user-centered rather than application-centered models. It allows user control over the personalisation process, enables self-awareness and acceptability for sharing one's user models for potential benefits [4]. In addition, (user-controlled) privacy of information is a crucial factor which affects modeling decisions and influences users' trust and confidence for allowing educational systems to access and process their information for potentially improved personalisation services. As the research shows, there is an essential trade-off between using personal information and the risk of exceeding the boundaries set by privacy rules, which requires careful implementations of privacy enhancements, based on the applications' domains [7].

3 Identified Problem: Enriching the LUM Vision

While we find UM in a state of transition, between the 'old' personalization approach of autonomous stand-alone and online systems to the 'new' frameworks for achieving interoperable user profiles, although is moving forward, is still been applied single-dimensionally:

- Most adaptive systems developed, especially the ones developed before the introduction of the Semantic Web, are only using their own internal models, when offering personalization services to their users.
- Newly introduced frameworks and architectures, while offering a solution in achieving interoperability across peer systems, do not meet some special requirements for involving systems beyond the educational domain [8].
- User Modelling Servers, a client-server architecture for allowing central information storing and simultaneously data access and retrieval, are mostly designed and developed to meet commercial requirements [9].

We are loosing user information, which is flowing around the WWW, because we are not thinking multi'*domai*n'sionally. Two unexplored rich sources for user information are the social e-networking and e-commerce domains. We can enrich LUM if we find a way to model our every day (life-long) interactions with services on the WWW from the social e-networking and the e-commerce domains, in order to enrich user information sets which are used by educational systems for personalization purposes. But, is this feasible? How can we realistically acknowledge the maturity status of our current knowledge while taking advantage of existing methods and technologies in order to achieve this multi'*domain*'sional approach? For the purpose of this research we have analyzed the two best representatives, based on their liquidity, platforms in the social e-networking domain: Facebook and Google's OpenSocial, and in the e-commerce domain: Amazon and eBay.

Knowledge: Firstly, we need to understand that there is a sea of user information outside the educational domain. Specifically, the social e-networking domain, e.g.

Facebook, MySpace, etc., and the e-commerce domain, e.g. Amazon, eBay, etc., welcome daily millions of users on their respective sites, and a rich set of user information is stored in several systems in these domains.

Methods: Recent data portability announcements from the two key players in the social networking domain [10, 11] have revealed these providers' initiatives to pass user data back to their 'owners'. In addition, various versions of APIs are available from providers of user models in the social networking and the e-commerce domains, which can be used by developers to 'take a dip' inside the providers' databases, retrieve user information held by these providers after direct consent by the user-owner, and enrich their websites with social and/or e-commerce features.

Technologies: Semantic Web Technologies have provided us with effective solutions in viable problems, like description of resources in a machine-understandable way and standards for communication and exchange among independent providers across various platforms [6]. UM is not an exception, thus by taking advantage of these technologies we can develop infrastructures that will allow us to take a step forward towards multi'*domain*'sional visions.

4 Proposed Solution & Up-to-Date Results

4.1 Achieving Interoperability across Domains

Special consideration has been given to collecting the requirements for employing a Scrutable User Modelling Infrastructure (SUMI), in an attempt to enable exchanging of user models from the social networking and e-commerce domains towards educational personalization services [12].

SUMI Models' Architecture: A SUMI model can be considered as a model of models, where users can add models to their SUMI collection and import the content of these models, define the privacy status of each one of these models or parts of them to determine how other users will be allowed to access them, and should they decide to do so, export their models, or parts of them, to subscribed educational services.

The first step requires defining the architecture of SUMI models which will "prepare" the infrastructure to accept the various models imported by the SUMI users. A 4-category architecture of SUMI models was resulted after the comparative evaluations of the two representatives in each domain:

Generic User Data: Any input that is being entered by the user manually in any way AND it is common to both domain representatives; e.g. The "Interests" field that users can find in Facebook and OpenSocial websites (Hi5, MySpace, etc.).

Service-Specific User Data: Any input that is being entered by the user manually in any way AND it is NOT common to both domain representatives; e.g. MySpace's "Favourite Heroes" which is not provided in Facebook.

Service-Generic Generated Information: Any user information, represented by an attribute or attribute list, which is generated by the provider based on previous user input AND it is common to both domain representatives; e.g. "Items Recommendations" which users receive from both eBay and Amazon.

Service-Specific Generated Information: Any user information, represented by an attribute or attribute list, which is generated by the provider based on previous user input AND it is NOT common to both domain representatives; e.g. Amazon's "Items New Releases" which you do not find in eBay.

SUMI Ontology: A SUMI ontology has been developed, and can be found at [13], while taking in mind the structure of the two best representatives from the social networking and e-commerce domains in order to enable mapping of the various providers' data models, for successful communication between them and educational personalization systems via SUMI. Any provider of user models can define its internal data model using an online SUMI service, found at *http://www.mysumi.org*, which is connected with the SUMI ontology. The ontology introduces five categories while adopting some other categories from an already existing and well-accepted ontology: the General User Modeling Ontology (GUMO), which was created for uniform interpretation of distributed models in intelligent semantic web enriched environments [14]. Although we avoid re-creating several features and categories by adopting GUMO, we feel it needs to be extended since it lacks some attributes for successfully modelling providers from the social e-networking and e-commerce domains. Below we provide a short description of each introduced category in the SUMI ontology while we explain the mapping, if any, to GUMO categories:

User*:* This category describes SUMI users and maps to GUMO's *Person* category.

Provider*:* The category *Provider* describes all providers of user models. They are divided into two sub-categories: Social Networking and E-Commerce.

AttributeCategory*:* This describes the four categories of attributes which can be retrieved from providers of user models using their APIs. It reflects the SUMI models' architecture as explained above.

Attribute*: Attribute* describes all attributes that can be retrieved from providers of user models using their APIs. Each attribute belongs to a group of attributes and each group of attributes belongs to one *AttributeCategory. Attribute* uses the *GUMO:SituationalElements* category to map each attribute in SUMI with a situational element in GUMO using the *hasGUMORelation* property. For example the attribute *Favourite_Movie* in SUMI is mapped with the element *Film* in GUMO.

DictionaryConcept*:* This category provides meaning to the attributes described in *Attribute* using the property *sameAsDictionaryConcept.* As the property shows, we map each attribute with dictionary concepts in order to provide explicit and well-accepted meaning to the attributes' definitions. Further explanation is provided below.

Defining meaning with Dictionary Concepts: Defining attributes' meaning using a dictionary, which was written expressly for the purpose of explaining terms to people, can always explain any ontological reasoning and relationships by showing the users the relevant dictionary entries [15]. SUMI follows the same approach and provides meaning to the various attributes, gathered by providers from the two examined domains, using the online Oxford English Dictionary [16]. It maps each attribute with a dictionary concept to provide explicit and well-accepted meaning and presents them to users when inspecting their SUMI collections. For example, the attribute "Interests" which is common to providers in the social e-networking domain has been mapped with the dictionary meaning: "A thing in which one has an interest or concern". This approach helps avoiding multiple interpretations by people living in different countries, with different cultures and different cognitive social models.

A RESTful Approach: By researching the literature and analyzing the API versions that are available from the representatives of the two examined domains, one can identify that although Web Services fully satisfy the needs for such an infrastructure, in terms of achieving communication between providers across different domains, it is an advance solution to an 'easier' problem. A more 'relaxed' approach can fit perfectly to the requirements for enabling communication among providers and exchange of user information between them and educational personalization services. Thus, SUMI adopts the REST protocol, an easy, effective and efficient approach [17], which is also offered in all four representatives of both examined domains. All attributes inside SUMI are represented with a URI which can be accessed by the subscribed systems via HTTP methods, in order to retrieve the attributes' values. Of course, direct consent of the user-owner is required for enabling SUMI to return the requested attribute's value when the method is called.

4.2 Evaluating Three Scrutability & Three User-Controlled Privacy Privileges

As already stated, our goal is to make SUMI as scrutable as possible while implementing user-controlled privacy enhancements, following the LUM vision which is based on passing the full control to the user-owner of the information. For the purposes of the 1st SUMI user evaluation, which was conducted over a period of 38-days with 107 users, we proposed three scrutability and three (user-controlled) privacy user privileges. Users were exposed to a prototype SUMI service, in order to determine whether the proposed privileges were acceptable to the users, whether the users were able to complete assigned tasks while using these privileges, and whether they understood the consequences of their interactions with the system while completing the tasks [18].

Scrutability Privileges: The examined scrutability privileges which we offered to our users during the evaluation were:

- Adding at least one social e-networking and one e-commerce model to their SUMI collection. Users were exposed to the 4-category models' architecture.
- Importing in SUMI the content of any category of their previously added models. SUMI provided the option between dynamic information, meaning real-time HTTP GET request and retrieval of real-time data from the provider of the user model, and static information, meaning the cache copy that was taken, when the last dynamic import request was generated by the user, kept inside the SUMI database and will be retrieved using SQL queries. Network failures or busy network traffics, are some reasons that users could take advantage of the static information option.
- Exporting their models, or parts of them, to a subscribed educational service – a group formation system. Users were allowed to inspect and approve the transaction details before enabling SUMI to export their information towards the subscribed system.

User-Controlled Privacy Privileges: The three proposed privacy privileges were designed for controlling how others would access a user's SUMI collection and they were all user-controlled, i.e users could customise them as they wished. An example of a user-controlled privacy privilege is the latest privacy enhancement from

Facebook, which announced that users can determine which parts of their model will be accessed by other non-friends users. Previously, users did not have any choice as to how outsiders would access their models. Although a privacy setting existed, which allowed other users to view only the visited user's name and profile picture, it did not give the chance to users to customise this setting as preferred. The three offered user-controlled privacy privileges were:

- Setting the privacy status for all 4 categories of all previously added models. Users had the choice of three privacy statuses: *public*-others can see that the model exists and anyone can view its content, *private*-others can see that the model exists but they have to place a request to the model's owner for viewing the model's content, and *hidden*-others can not see that the model exists, therefore the model's content is accessed only by the model's owner.
- The second privilege consisted of responding to viewing requests from other SUMI users to some 'private' categories of their previously added models,
- The third privilege was about visiting other users' SUMI models and placing viewing requests to some 'private' categories of those users' previously added models.

The conclusion of the evaluation has shown that the users expressed their general approval of the proposed privileges while proposing useful suggestions regarding improvements to the presentation and interface to the system. In addition, participants recommended other scrutability and privacy privileges which they thought SUMI should offer to its users. Elaborate description and detailed results can be found at [18].

5 Future Work Agenda

Further work is required to confidently support our contribution which could enrich the LUM vision as stated in this paper. Consequently a future work agenda has been developed and it includes: further analysis of the announced data portability initiatives [10, 11] and the available APIs in order to identify any potential challenges for implementing such a proposed SUMI service, working closely with educational personalisation systems, such as AHA![3], for developing real-life adaptive lessons which will demonstrate how SUMI can enrich user models with sets of information from the two examined domains, and examining how adopting REST as SUMI's communication protocol can influence any design decisions. Finally, a second user evaluation is already underway, which builds on the feedback from the first evaluation and provides more advanced scrutability and user-controlled privacy privileges to SUMI users. As the results from the first evaluation revealed, users wanted to customize what information SUMI is allowed to keep internally, and how the imported data is treated by the SUMI service, which show a satisfactory degree of user engagement to a potential SUMI service. This raises further questions: Until what point can we pass the control to the users, and by adopting which scrutability and privacy privileges can we best achieve this? How do users react on having full control? Is there a point where they need guidance on how to proceed or can they cope with absolute freedom to inspect and alter the way they are being modeled?

References

1. Kay, J., Lifelong learner modeling for lifelong personalized pervasive learning. IEEE Trans on Learning Technologie*s*, 1(4), 215--228 (2008)
2. Kay, J., Lifelong user models, memory and learning. ResearchChannel.org, Last retrieved on 29/03/2009 at http://www.researchchannel.org/prog/displayevent.aspx?rID=21195&fID=4752, (2007)
3. De Bra, P., Calvi, L., AHA: a Generic Adaptive Hypermedia System, In Proc. of the 2nd Workshop on Adaptive Hypertext and Hypermedia, Pittsburgh, 5--12 (1998)
4. Kay, J., Scrutable adaptation: because we can and must. In V. Wade, H. Ashman, and B.Smyth, editors, Proceedings of Adaptive Hypermedia and Adaptive Web-Based Systems, 4th International Conference, AH2006, pages 11-19. Springer, (2006).
5. Dolog, P., Schaefer, M., A Framework for Browsing, Manipulating and Maintaining Interoperable Learner Profiles. In Proc. of the 10th International Conference on User Modelling, Edinburgh, UK, (2005)
6. Berners-Lee, T., Hendler, J., Lassila, O., The Semantic Web. Scientific American, 284, 5, 34--43 (2001)
7. Kobsa, A., Privacy-Enhanced Personalization. In Communications of the ACM, 50(8), 24-33 (2007).
8. Bieliková, M., Kuruc, J. Sharing User Models for Adaptive Hypermedia Applications. In H. Kwasnicka, M. Paprzycki (Eds.), Proc. of ISDA 2005, Sept. 2005, Wroclav, Poland, IEEE Computer Society Press, Loa Alamitos, pp. 506-511 (2005).
9. Fink J., Kobsa, A., A Review and Analysis of Commercial User Modeling Servers for Personalization on the World Wide Web. In User Modeling and User-Adapted Interaction - Special Issue on Deployed User Modeling, 10, 209--249 (2000)
10. Facebook Developers News, Announcing Facebook Connect, Last Retrieved on 29/03/2008 at http://developers.facebook.com/news.php?blog=1&story=108
11. Google Inc, Google Friend Connect, Last Retrieved on 29/03/2009 http://www.google.com/friendconnect
12. Kyriacou D., A Scrutable User Modelling Infrastructure for enabling life-long User Modelling. In Adaptive Hypermedia and Adaptive Web-Based Systems - Doctoral Consortium, Hannover, Germany, (2008)
13. SUMI Ontology, Last Uploaded on 29/03/2009 at http://mysumi.org/sumiOntology.owl
14. GUMO Ontology and UbisWorld, Last Accessed on 29/03/2009 at http://www.ubisworld.org/
15. Apted, T., Lum, A., and Kay, J. Supporting metadata creation with an ontology built from an extensible dictionary. In: P.D. Bra and W. Nejdl, Editors. Adaptive Hypermedia and Adaptive Web-based Systems. Springer, (2004)
16. The Oxford English Dictionary Online, Last Accessed on 29/03/2009 at http://www.oed.com/
17. IT Professional Blog – Representational State Transfer (REST), Last Retrieved on 29/03/2009 at http://mauriziostorani.wordpress.com/2008/07/27/rest-representational-state-transfer-and-restful-web-services-concepts-and-examples/, (2008)
18. Kyriacou, D., Davis, H., and Tiropanis, T., Evaluating Three Scrutability and Three Privacy User Privileges for a Scrutable User Modelling Infrastructure. In: User Modeling, Adaptation and Personalization (*In Press*), 22-26 June 2009, Trento, Italy, (2009).