A Scrutable User Modelling Infrastructure for enabling life-long User Modelling

Demetris Kyriacou

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

Abstract. User Modelling is the core component for the majority of personalisation systems. By keeping a model for every user, a system can successfully personalise its content and utilise available resources accordingly. While researching the literature, one can recognize the importance of achieving interoperability across various platforms and systems while attempting to personalise a large diversity of web resources. Furthermore, scrutable solutions allow users to control any modelling process that uses their information. Finally, privacy of user data while exchanging user models from one source to another must be taken in mind. With this paper, a Scrutable User Modelling Infrastructure is presented which blends together these user modelling 'ingredients' and, by adopting Semantic Web technologies, attempts to model a range of life-long user interactions with a variety of web-based systems from the educational, business and social networking domains.

1 Introduction

Consider the following imaginary scenario: Maria is a computer engineering student and for her assignment is required to develop a system written in Prolog. Maria knows only the basics around the Prolog programming language since she has read a book, bought from Amazon, and has gone through a couple of tutorials found on the course's website. In order to cope with the requirements of the assignment, Maria needs to register with personalisation system XYZ and follow a short course on advanced concepts of Prolog. Maria logs on this new service, which it is proposed in this paper, and exports her Amazon model to XYZ, having previously set her Amazon model's privacy status to public. Furthermore, Maria filters her browsing history and selects to send to XYZ only the part that shows that she has gone through the Prolog tutorials. Now, XYZ knows what Maria is familiar with according to her various models and adapts its content accordingly to teach Maria Prolog concepts and features that she is unfamiliar with but are essential for the completion of the assignment.

What if each one of us had a user model for life from the moment we were born and that model was updated constantly with our every day interactions with various online services from the educational, business and social networking domains? What if we had absolute control over it and we decided which system gets access to which part of our model? What if we could set the privacy status of our information? Then, we would be able to scrutinise it the way we wanted and benefit from any interactions

we chose to make with any service out there, by providing our model before the interaction, and receiving it back at the end, updated with the new resulting data based on how we interacted with that service.

2 User Modelling

While trying to move a step further and model a dynamic user in a variety of contexts, life-long User Modelling (UM), the ability to model a dynamic and changing user throughout lifetime interactions with a diversity of resource providers, appears to be an attractive solution [6]. Focusing on some key UM areas is essential for coping with existing and arising challenges:

2.1 Interoperability

Interoperability 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" [1].

With the recent evolvement of the World Wide Web to the Semantic Web [2], the issue of interoperability has become a burning issue in the area of UM. Exchanging user profiles across various sources in a distributed eLearning (and not only) environment can not be achieved if explicit and widely accepted protocols are not being developed and adopted that will allow description, discovery and exchange of user models, stored in various systems - written in different languages and for different platforms [4, 5].

2.2 Scrutability

The term scrutability means that the model of every user can be controlled by the user him/her self to determine what has been modelled about him/her and how the modelling process was conducted [7].

By adopting srutability in UM, users gain control of their models and therefore they can set their preferences on how the modelling process is applied on them. In addition, users can select in which stereotypes they should be included and which ones they should not. Furthermore, the users can alter the value of any single inference that is used for drawing conclusions about them [7].

2.3 User Privacy

Privacy-Enhanced Personalisation is an area that aims at merging together the techniques and goals of UM with privacy considerations and apply the best possible personalisation inside the boundaries set by privacy rules [8].

As the research in this area shows, there is no ideal solution while attempting to combine these two important elements. Instead, numerous small enhancements must

be implemented, depending on the user and application domains in each case, in order to achieve the best possible solution.

The most important considerations while attempting to model a large diversity of users appear to be the issues of:

- Informing the users about the process of gathering their information.
- Allowing users to know how their data is stored and processed in order to draw conclusions about them.
- Acquiring users' approval when their data is being exchanged from one system to another in order to achieve effective and efficient personalisation.

2.4 User Modelling Standards

It is obvious that in such environment, agreement on common structures and scope of user information modelled is needed. The need for standards was naturally raised and was addressed by two significant organisations, IEEE and IMS and resulted in two widely accepted UM standards PAPI (IEEE) and LIP (IMS) [3].

2.5 Semantic Web Technologies in User Modelling

New directions and guidelines for UM have arisen with the introduction of the Semantic Web [2]. New technologies, such as XML, XML Schema, RDF, RDF Schema, OWL and Web Services, have emerged that allow the content of user models to be expressed in a format that can be read and processed by software agents, thus permitting them to find, share and integrate information more easily and efficiently.

3 Research Questions

Listed below are the research questions that will attempt to answer by the completion of my PhD.

What are the requirements for adopting a scrutable user modeling architecture and a communication protocol, for users AND providers of user models, for enabling exchange of user models amongst them?

Immediately, further questions arise which will contribute to answering the main question set above:

- Interoperability
    - How can we map all these different data models while enabling communication amongst them?
    - How can we allow providers of user models to define their data models for importing and exporting user data in order to enable effective and efficient exchange of user models?
    - What is the optimal solution for storing user data while adopting such architecture, i.e. where is it best to keep all this user

information that it is been used for user modelling? What are the advantages and disadvantages of each choice?

- How can we enable this scrutable user modeling architecture to reflect the educational, business and social networking domain of every user in a way that will impose no barriers on merging and exporting data from any of these three domains, in order to explore the potential for interoperability across these three domains?
- Can Semantic Web technologies assist on the development of such architecture? If yes, which technologies are fitter for such implementation?
- What are the requirements for providers of user models, especially commercial providers, to employ this scrutable architecture and take advantage of the proposed communication protocol, in order to enable exchanging of user models to take place?
    - What format should be used when providing user information to personalisation systems for adaptation purposes?
- To what extent it is possible for such architecture to allow users to scrutinise their models and express their data privacy preferences?
    - How should we define a user model part and how can we allow SUMI users to export parts of their models to subscribed services?

4 Proposed Solution

An initial architecture, a Scrutable User Modelling Infrastructure (SUMI), is proposed, which will be used for demonstration (prototype) and evaluation purposes while attempting to answer the research questions set above

A prototype has already been designed and implemented to meet the majority of the initial core requirements mentioned above. More specifically:

- Every model in SUMI is a representation of an integrated variety of user models every user holds, by interacting with various services on the World Wide Web. The SUMI models' architecture will be later divided into three categories: educational, business and social networking data.
- A SUMI ontology will be developed, while taking in mind the structure of representative services from the educational, business and social networking domains, in order to enable mapping of the various providers' data models for successful communication between them via SUMI.
- Every user has absolute control over his/her SUMI model. The user can decide which models to integrate in SUMI and also who gets to see which part of his/her SUMI model.
- Importing models from services will be achieved by adopting the SUMI import protocol, based on Semantic Web technologies.
- Every user can define the status of the data retrieved by his/her various models. The three categories of data are public, private and hidden.
- The most important feature of SUMI is the users' ability to export a part of their SUMI model to any registered service they prefer. This will be

achieved by adopting the SUMI export protocol, again based on Semantic Web technologies.

5 Evaluation Plan

For the evaluation of SUMI, simulation testing will be conducted which will consist of various generated queries, in order to demonstrate and evaluate the architecture of SUMI, the communication protocol and any potential features. This include the designing and developing of sample services, based on real educational, business and social networking providers of user models, in order to test the SUMI architecture and communication protocol against the pre-defined requirements. The queries will represent hypothetical requests made from (and to) the sample services that will be implemented, for enquiring various user information located in several user models that these services will hold. The SUMI architecture will be used to bring these sample services together, while offering a level of scrutability to the owners of the user models, whereas the introduced communication protocol will enable the exchange of user models to take place in an effective and efficient way.

References

1. Aroyo, L., Dolog, P., Houben, G.J., Kravcik, M., Naeve, A., Nilsson, M., Wild, F., Interoperability in Personalised Adaptve Learning. Educational Technology & Society 9, 2 (2006) 4-18.
2. Berners-Lee, T., Hendler, J., Lassila, O., The Semantic Web. Scientific American, 284, 5 (2001) 34-43.
3. Dolog, P., Nejdl, W., Challenges and Benefits of the Semantic Web for User Modelling. In Proceedings of AH2003 workshop at 12th World Wide Web Conference, Budapest, Hungary, (2003).
4. Dolog, P., Schaefer, M., A Framework for Browsing, Manipulating and Maintaining Interoperable Learner Profiles. In Proceedings of the 10th International Conference on User Modelling, Edinburgh, UK, Springer-Verlag, (2005a).
5. Dolog, P., Schaefer, M., Learner modelling on the semantic web. In Proceedings of PerSWeb-2005 Workshop: Personalisation on the Semantic Web at User Modelling 2005: 10th International Conference, Edinburgh, UK, (2005b).
6. Kapoor, A., Horvitz, E., Principles of Lifelong Learning for Predictive User Modeling. In Proceedings of the 11th International Conference on User Modeling, Corfu, Greece, (2007).
7. Kay, J., Stereotypes, Student Models and Scrutability. In Proceedings of the 5th International Conference on intelligent Tutoring Systems, G. Gauthier, C. Frasson, and K. VanLehn, Eds. Lecture Notes in Computer Science, vol. 1839, London, Springer-Verlag, (2000) 19-30.
8. Kobsa, A., Privacy-Enhanced Personalization. Communications of the ACM 50, 8 (2007) 24-33.