RSS-based Interoperability for User Adaptive Systems

Yiwen Wang[2], Federica Cena[1], Francesca Carmagnola[1], Omar Cortassa[1], Cristina Gena[1], Natalia Stash[2], and Lora Aroyo[23]

[1] Dipartimento di Informatica, Università di Torino
Corso Svizzera 185, Torino, Italy
{carmagnola, cena, cgena}@di.unito.it
[2] Eindhoven University of Technology, Computer Science
{n.v.stash, y.wang}@tue.nl
[3] VU University Amsterdam, Computer Science
l.m.aroyo@cs.vu.nl

Abstract. This paper presents an approach to exploit widely used tag annotations to address two important issues in user-adaptive systems: the cold-start problem and the integration of distributed user models. The paper provides an example of re-use of user interaction data (tags) generated by one application into another one in similar domains for providing cross-system recommendations.

1 Introduction

The *Web 2.0* phenomenon introduced various social applications enabling online collaboration and encouraging the participation and contribution of spontaneous social networks. Users are increasingly involved in multiple Web 2.0 environments, such as Facebook.com, Flickr.com, Del.icio.us, etc. However these applications are still "digital islands" in terms of personalized experience - not truly interconnected in a way which allows users to capitalize on the full potential of a distributed multi-application environment. Most of those services maintain a different identity, e.g. login information, preferences or profile of users with a limited integration of these data between different applications. However, tags inserted by users could be extremely useful for adaptive web applications [2], e.g. to enrich and extend the user model. User usually tags to highlight and organize the items she is interested in, in order to retrieve them later. Thus the action of tagging can be be analyzed in order to make interesting inferences on the user model [3]. The exploitation of tags for improving the user model, requires that systems could understand the semantics of the tags (e.g., applying suitable strategies borrowed from automatic Word Senses Disambiguation).
The focus of this paper is to illustrate how existing fragments of user data in the form of tags can be brought together with the help of explicit semantics, and in this way allow for an adequate personalized experience across the boundaries of particular applications. This poses a considerable number of technological demands. Working in a distributed setting implies that personalization considers

both *data-integration issues*, i.e. how the information from different applications is related, as well as *context-modeling issues*, i.e. in which space/time/mode the statements about a user are valid. In this paper we look at the data-integration issue.Concretely, we provide a method for extracting, conceptualizing and linking user tags contained in public RSS files generated in the interaction of users with a social recommender system iCITY [3]. The tags are mapped to art-related concepts used in the personalized museum applications CHIP [1].

The paper is structured as follows: Section 2 describes the architectural specifics of iCITY and CHIP. These are further elaborated in Section 3 wherein we present the conceptualization of iCITY tags, and the mapping of such tags into CHIP user model. Finally, in section 4 we draw conclusions and future work trends.

2 iCITY - CHIP User Interoperability Architecture

iCITY is a social web-based, multi-device recommender system. It provides suggestions on cultural events in the city of Turin, and allows users to insert new events, comments and tags. Recommendations are based on the user model enriched with tags, exploited to infer user features (see for details [3]). iCITY has a modular architecture for extracting, maintaining, reasoning and exporting user tags, which can be shared with other applications via a RSS feed. The main components for interoperability are:
The *Importer Module* which is responsible for the extracting the tags, available in form of RSS files, from external sources, e.g. web community like flickr.com and del.icio.us. Once all the user tags have been extracted, they are used to obtain useful information about user's interest and knowledge. In order to understand their meaning, the system looks for correspondences between the tags and the *synsets* and the *domains* of the MultiWordet database.
The *Exporter Module* which generates, for every user, a RSS file containing the list of the events tagged by the user. For every event, the RSS file stores: the title, the URL, the description, the reference to the event category and subcategory in the iCITY event ontology, the reference to the Wordnet synsets and domains linked to the subcategory, and finally the list of the tags associated by the user to the event. In this way, a recipient system can import this RSS file containing the tags and reason about them. The recipient can try to disambiguate the meaning of tags thanks to the information, provided in the RSS file, about the event subclass they belong to and the references to WordNet domain and synset.
The **CHIP** sytem[4] illustrates a personalization infrastructure for semantically enriched museum collections. We use the digital database of the Rijksmuseum ARIA[5] (750 master pieces) and its mappings to external vocabularies [1], namely the three Getty vocabularies[6], as well as the subject classification Iconclass[7]. The

[4] http://www.chip-project.org/demo
[5] http://rijksmuseum.nl/aria/
[6] http://www.getty.edu/research/conducting research/vocabularies
[7] http://www.iconclass.nl/libertas/ic?style=index.xsl

use of common vocabularies provides the new data repository a relational and hierarchical structure for reasoning and making recommendations. Based on this semantics-enriched data model, we have implemented a set of Web-based tools [1], e.g. Artwork Recommender and Online Tour Wizard, and a PDA-based Mobile Tour Guide to collect user input both in the virtual museum space on the Web and in the corresponding physical museum. All user interactions in each of the tools are stored in a user profile, categorized in four clusters: *personal* characteristics, e.g. name, age, gender, which could be initialized by either importing an existing FOAF RDF profile or via an OpenID channel linking the CHIP login data to an existing login information of third party Web application; *social* information initialized by FOAF properties, e.g knows, openid, organization, OnlineAccount, user *ratings* of artworks and topics in terms of VRA Core properties[8], e.g. work, creator, title, creationDate, creationSite, subject; and user *iteraction* with, e.g. virtualTours.

3 iCITY-CHIP User Tag Interoperability

In this interoperability use case, an open API is adopted to request and link user data. Once the user personal (login) information is aligned between CHIP and iCITY, based on the RSS feed we maintain a dynamic mapping of iCITY user tags to the CHIP vocabulaty set (ARIA shared with Getty and IconClass) and general purpose lexical data such as WordNet. This will populates users' profiles (especially first-time users) in CHIP and enable instantly generate recommendations in the Rijksmuseum collection.

The mapping is realized in two steps: (i) to identify the type (e.g. creator, place, material, etc.) of iCity tags as a simple restriction; (ii) to map tags to CHIP art concepts by using the Simple Knowledge Organization System (SKOS) Core Mapping Vocabulary Specifications[9]. For the first stage alignment, the mappings are still based on the lexical match of tags. With a few additional simple restrictions by applying the type of tags, a lexical match gives more confidence to generate a strong semantic match [4]. For example, the semantic equivalence between iCITY tag "Amsterdam" and Getty Thesaurus of Geographic Names (TGN) creationSite "Amsterdam" is expressed with *skos:equivalentConcept* for the type of place. And *skos:narrower* for the type material points from tag "photo" to concepts "Photo collotype", "Photo Gelatin silver print" and "Photo Bromide print" in the Rijksmuseum ARIA hierarchical specialization.

4 Conclusion and future work

In this paper we have presented an approach to exploit widely used tag annotations to address two important issues in user-adaptive systems in the cultural heritage domain: the cold-start problem and the integration of distributed user

[8] http://www.vraweb.org/projects/vracore4/index.html
[9] http://www.w3.org/2004/02/skos/

profiles. We have sketched a scenario, in which user tagging about cultural events gathered by iCITY is used to enrich the user profile for generating personalized recommendations of artworks and topics in CHIP. To realize the tagging interoperability, first we have investigated the problems that arise in mapping user tags to shared vocabularies (ontologies). With the help of SKOS matching operators we propose an approach to deal with the possible misalignment of tags and domain-specific ontologies. Further, we need to address the mapping of user tags to event ontologies (iCITY) and possibly to multiple concepts in the domain-specific ontologies. Further, we need to close the loop by allowing import of the CHIP user profile into iCITY and in this way to refine the iCITY user model.

References

1. Lora Aroyo, Natalia Stash, Yiwen Wang, Peter Gorgels, and Lloyd Rutledge. Chip demonstrator: Semantics-driven recommendations and museum tour generation. In Karl Aberer, Key-Sun Choi, Natasha Fridman Noy, Dean Allemang, Kyung-Il Lee, Lyndon J. B. Nixon, Jennifer Golbeck, Peter Mika, Diana Maynard, Riichiro Mizoguchi, Guus Schreiber, and Philippe Cudré-Mauroux, editors, *ISWC/ASWC*, volume 4825 of *Lecture Notes in Computer Science*, pages 879–886. Springer, 2007.
2. Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, editors. *The Adaptive Web, Methods and Strategies of Web Personalization*, volume 4321 of *Lecture Notes in Computer Science*. Springer, 2007.
3. Francesca Carmagnola, Federica Cena, Omar Cortassa, Cristina Gena, and Ilaria Torre. Towards a tag-based user model: How can user model benefit from tags? In Cristina Conati, Kathleen F. McCoy, and Georgios Paliouras, editors, *User Modeling*, volume 4511 of *Lecture Notes in Computer Science*, pages 445–449. Springer, 2007.
4. Omelayenko B. Tordai, A. and G. Schreiber. Thesaurus and metadata alignment for a semantic e-culture application. Semantic Authoring, Annotation and Knowledge Markup Workshop, 2007.