Tag Interoperability in Cultural Web-based Applications

Federica Cena, Francesca Carmagnola, Omar Cortassa, Cristina Gena
Dipartimento di Informatica, Università di Torino, corso Svizzera 185, Torino, Italy
carmagnola,cena,cortassa,cgena@di.unito.it

Yiwen Wang, Natalia Stash, Lora Aroyo
Eindhoven University of Technology, Den Dolech 2, 5600 MD Eindhoven, The Netherlands
y.wang, n.v.stash@tue.nl; aroyo@cs.vu.nl

ABSTRACT

This paper presents an approach that shows how user interaction data (tags) generated by one application can be exploited by another one in similar domain for integrating user models in distributed and interactive environments. In particular we discuss the tags interoperability among two adaptive systems into the cultural heritage domain.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

tagging, interoperability, user modeling

1. INTRODUCTION

In the last years users are more and more involved in multiple *Web 2.0* [4] environments, such as Facebook.com, Flickr.com, YouTube.com, Del.icio.us, Digg.com. In Web 2.0 systems, users can label resources by means of keywords (tags), insert new contents, share objects, provide comments and so on. Social tagging is of utmost relevance also to the *Cultural Heritage* domain because it offers the opportunity for relationships between cultural heritage institutions, collections and users. In particular, social tagging may be of help: To bridge the gap between the professional language defined by domain experts and the popular "un-trained" language; to encourage individuals to find personal meanings/perspectives in public collections by labeling the artworks; and finally to create public engagement with cultural heritage collections. Projects that explore this challenge, such as the Steve Museum [5], demonstrated the effectiveness of social tagging in engaging visitors with collections, and for the museum to understand what users consider as relevant. The Powerhouse Museum [3] proved that user tagging and resulting folksonomies can be used to improve navigation and discoverability through museum collections.

Personalization has already been identified as a success factor for Web 2.0 applications [4]. User modeling that supports personalization for a single application are reasonably well understood.

This paper presents an approach to show how existing fragments of user data in the form of tags, together with the help of explicit semantics, can be shared across systems to realize personalization in distributed and interactive environments. This poses a considerable number of issues. In this paper we mainly focus on the data-integration issues. Concretely, we provide a method for extracting, conceptualizing and linking user tags contained in public RSS files generated during the interaction of users with a cultural recommender system, iCITY[1]. The tags are mapped to the art-related concepts used in the personalized museum application CHIP[2] and are then used to initialize the user model in a "non-intrusive" manner. Section 2 describes the tag interoperabilty between iCITY and CHIP, while Section 3 presents some discussions and future directions of work.

2. ICITY - CHIP TAGS INTEROPERABILITY ARCHITECTURE

iCITY is a social web-based, multi-device recommender system. It provides suggestions on cultural events (exhibitions, concerts, festival, etc.) taking place in the city of Turin. Moreover, it allows users to insert new events, to add information about events, to put comments and tags. Recommendations are based on the user model enriched with tags, exploited to infer user features (see [1]). The iCITY user model is an overlay of the iCITY Event ontology, created as an RDFS transformation of the event classification in *TorinoCultura*[3], a web portal managed by the municipality of Torino for informing citizens about cultural ongoing events in the city. This ontology contains links both to WordNet[4] synsets and domains. iCITY has a modular architecture for extracting, maintaining, reasoning and exporting user tags which can be shared with other applications via RSS feed. The main components for such an interoperability are the **Importer** and **Exporter Modules**.

In the iCITY registration form the user can provide her social web community account. The *Importer Module* has

[1]Digital Semantic Assistant iCITY. Dept. of Computer Science of Turin, City of Turin and CSP - ICT innovation are project partners; http://iCITY.di.unito.it/dsa-dev/
[2]http://www.chip-project.org/demo
[3]http://www.torinocultura.it/
[4]http://wordnet.princeton.edu/

the aim of retrieving the available RSS files containing the set of tags provided by the user in those web communities [2]. Once all the user tags have been extracted, they are used to infer information about user's interest and knowledge.

The *Exporter Module* generates, for every user, a RSS file with the list of the events tagged by the user. For every event, the file stores: the title, the URL, the description, the reference to the event class and subclass within the public RDFS event ontology, the reference to the Wordnet synsets and domains linked to the subcategory (which provides the semantics), and finally, the list of the tags associated by the user to the event.

CHIP system (Cultural Heritage Information Personalization) allows museum visitors to create their personalized experience in the Rijksmuseum of Amsterdam, both in the virtual collection on the museum Web site and in the physical museum by quickly finding the right path covering her interests [6]. In particular, the CHIP Artwork recommender provides the user with an interactive way to: express her art preferences; quickly find artworks of interest; build a user profile of these interests and preferences. This user profile is further used in the Tour Wizard to assist the user in (semi-)automatically creating museum tours, and in the Mobile Tour Wizard, where it is updated with the artworks seen and rated during the museum visit.

iCITY-CHIP Tags Interoperability. In the interoperability between iCITY and CHIP, we use an open API to request and link user data. Once the user personal information (login) 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: Rijksmuseum specific concepts, or shared domain vocabularies, such as Getty AAT, ULAN, TGN, IconClass, or general purpose lexical data, such as WordNet. The main challenge in achieving the interoperability of user tags between iCITY and CHIP is to provide a dynamic mapping mechanism, which allows for a constant stream of user tags from iCITY to be mapped to concepts from the internal vocabularies of CHIP. This will allow to use iCITY tags to populate the user profiles of new users in CHIP and to be able to instantly generate a tour of recommended artworks in the Rijksmuseum.

To implement the tag mapping from iCity to CHIP, we use the Simple knowledge organization System (SKOS) Mapping Vocabulary Specification[5] created for linking thesauri to each other with relationships *skos:exactMatch*, *skos:broadMatch*, *skos:narrowMatch*, and *skos:relatedMatch*. For this first stage alignment, the mappings are still based on the lexical match of tags. The semantic and syntactic mappings between the iCITY user tags and the CHIP concepts is realized in two steps: (i) to identify the type (e.g. creator, place, material, etc.) of tags as a simple restriction; (ii) to map tags by using SKOS Core mapping relations. Maintaining a certainty level for each mapping allows for tuning of those concept's relevance. Further evidences from the tag cloud and/or the user model allows for a good accuracy as well as the user's direct feedback/confirmation. However, some challenges still remains. *Tags are messy*: the mapping is realized in two steps: first, to identify the type of the tag, e.g. whether *Giovanni* is a person, a kind of material or a place; and then to the map tags using SKOS mapping relations. The disambiguation of mapping can be delegated to the user for a varification, or to collect further evidences from the tag cloud. *Tags grammatical variation*: often tags appear in various grammatical forms, which do not completely match the CHIP concept form, e.g. noun, adjective, etc. Maintaining additional relationships or distances between the different term forms allows for clustering of all possible mappings for a given tag. Using mapping to WordNet can facilitate this process. *Effect of tags for recommendations*: after mapping each tag to the CHIP vocabularies, tags can be exploited for generating recommendations in CHIP. Our idea is to treat the tags differently depending on their relations with the annotated events which are described in the user's RSS file from iCITY. Tags can also be used separately for the event for generating different recommendations. *Ranks of recommendations*: to rank the recommendations based on different tags, we are considering to maintaine a dynamic weight for each tag. It could be defined by factors like the *frequency* of use of this tag in the user profile, the *uniqueness* of use of this tag in the whole system, etc.

[5]http://www.w3.org/2004/02/skos/mapping/spec/

3. CONCLUSION AND FUTURE WORK

In this paper we have presented an approach that shows how user tags generated by one application can be exploited by another one in a similar domain for integrating user models, and thus avoiding the well known "cold start problem" of recommender systems. In the scenario, user tagging about cultural events of iCITY is used to enrich the user profile for generating personalized recommendations in CHIP. Several issues need to be addressed in future research. Among these, we would mention the possibility of mapping user tags to various ontologies to fully realize tagging interoperability. We are planning to propose additional mechanisms, such as the use of SKOS matching operators, to deal with the possible mis-alignment of tags and domain-specific ontologies. We are going to exploit the CHIP user profile into iCITY. The inferences made by the CHIP recommender could be extremely useful to refine the iCITY user model and could also be exploited to support the iCITY reasoning component in solving some semantics disambiguation problems.

4. REFERENCES

[1] F. Carmagnola, F. Cena, O. Cortassa, C. Gena, and I. Torre. Towards a tag-based user model: how can user model benefit from tags? *Proc. of UM2007, LNCS*, 4511:445–449, 2007.

[2] F. Carmagnola, F. Cena, and C. Gena. User modeling in the social web. In B. Apolloni, R. Howlett, and L. Jain, editors, *KES (3)*, volume 4694 of *Lecture Notes in Computer Science*, pages 745–752. Springer, 2007.

[3] S. Chan. Tagging and searching: serendipity and museum collection databases. In *Museums and the Web 2007*, pages 87–99, 2007.

[4] T. O'Reilly. What is web 2.0. design patterns and business models for the next generation of software. 30/09/05.

[5] J. Trant and B. Wyman. Investigating social tagging and folksonomy in art museums with steve.museum. In *World Wide Web (WWW '06)*, 2006.

[6] Y. Wang, R. Sambeek, Y. Schuurmans, L. Aroyo, N. Stash, L. Rutledge, and P. Gorgels. Be your own curator with the chip tour wizard. In *Museum and the web (MW '08)*, 2008.