Building Blocks for User Modeling with data from the Social Web

Fabian Abel[1], Nicola Henze[1], Eelco Herder[1], Geert-Jan Houben[2], Daniel Krause[1], Erwin Leonardi[2]

[1] L3S Research Center, Leibniz University Hannover, Germany
{abel,henze,herder,krause}@L3S.de
[2] Web Information Systems, TU Delft, The Netherlands
{g.j.p.m.houben,e.leonardi}@tudelft.nl

Abstract. In this paper we present the so-called Grapple User Modeling Framework (GUMF) which provides user modeling functionality via the Web to applications that would like to offer personalization and adaptation features to their customers. GUMF introduces the notion *dataspaces* which create a logical view on (possibly distributed) user data and provide advanced contextualization and reasoning mechanisms. We showcase the Mypes service that exploits GUMF dataspaces to connect, aggregate, align and enrich user profile information form the Social Web.

1 Introduction

Adaptive systems require information about their users to adapt their functionality [1]. Today, users leave a plethora of traces on the Web that could possibly serve as input for these systems: people provide profile information in social networking services such as Facebook or LinkedIn, annotate their pictures and bookmarks at Flickr or Delicious, and write about their interests in their (micro)blogs. Thereby a lot of useful profile information becomes publicly available [2] . Further, the aggregation of such distributed user data is supported by initiatives such as the Linked Data initiative [3], standardization of APIs (e.g. OpenSocial[1]) and authentication and authorization protocols (e.g. OpenID, OAuth), as well as by (Semantic) Web standards such as RDF, RSS and specific vocabularies such as FOAF[2], SIOC[3], or GUMO [4]. Generic user modeling servers such as CUMULATE [5] or PersonIs [6] facilitate handling of aggregated user data. The Grapple User Modeling Framework (GUMF) [7] follows these approaches and offers means to deliver customized user modeling functionality to its clients.

In this paper we discuss so-called GUMF *dataspaces* that embody and enforce the actual user modeling intelligence of GUMF. We present the *Mypes* service[4] which exploits dataspaces to connect, aggregate, align and enrich user

[1] http://code.google.com/apis/opensocial/
[2] http://xmlns.com/foaf/0.1/
[3] http://rdfs.org/sioc/spec/
[4] http://mypes.groupme.org/

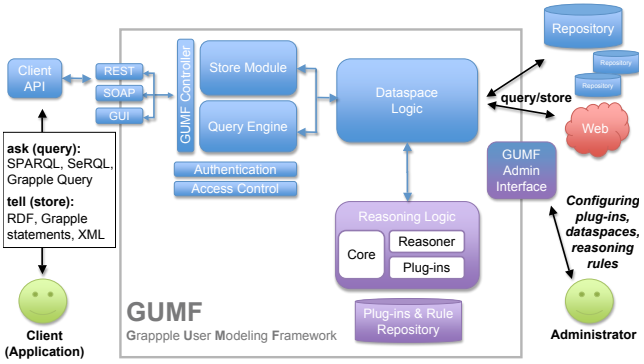


Fig. 1. Architecture of the Grapple User Modeling Framework (GUMF).

profile information form social networking services (Facebook, LinkedIn), social media services (Flickr, Delicious, StumbleUpon, Twitter, Blogspot) and others (Google). Mypes enables users to overview their distributed profiles and via GUMF it allows application developers to integrate aggregated, semantically enriched profile data in RDF or vCard format via a lightweight REST interface into their applications to provide personalization and adaptation functionalities.

2 Grapple User Modeling Framework

Figure 1 shows the architecture of GUMF. The elements at the top provide the essential, generic functionality of the framework; elements part at the bottom right provide generic as well as domain-specific *reasoning logic*. Client applications can access GUMF either via a RESTful or SOAP-based API. Further, there is a *Java Client API* that facilitates development of GUMF client applications[5]. Client applications mainly approach GUMF to store user information (handled by the *Store Module*) or to query for information (handled by *Query Engine*). User profile information is modeled by Grapple statements [7], which are basically reified RDF statements about a user, enriched with provenance metadata. GUMF currently supports SPARQL and SeRQL queries as well as *Grapple query*, a pattern-based query language that exploits the Grapple statement structure to specify what kind of statements should be returned by GUMF. Authorized client requests are answered by GUMF's *Dataspace Logic*. Dataspaces concentrate the actual user modeling capabilities of GUMF. They are equipped with data storage repositories that either reside at the GUMF server or are distributed across the Web (possibly maintained by the client application itself), and with (reasoning) plug-ins that further enrich the data that is available in the repositories or transform user data into a structure/format that is appropriate for the GUMF client applications that request user data.

The *Administrator* of a GUMF client application can configure dataspaces and plug-ins via the *GUMF Admin Interface*. Activating or deactivating plug-

[5] cf. GUMF help pages: http://pcwin530.win.tue.nl:8080/grapple-umf/help/

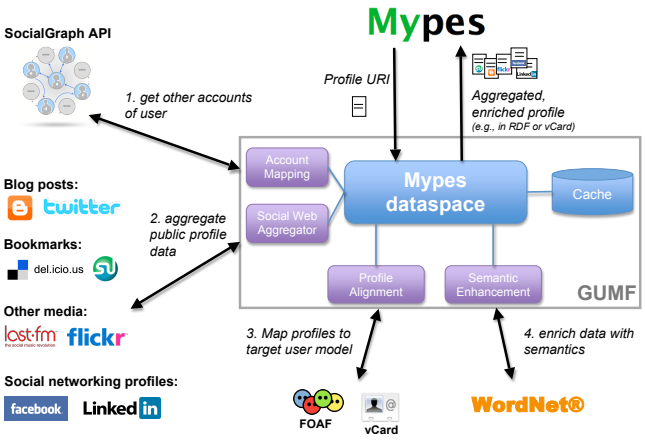


Fig. 2. Aggregation and enrichment of profile data with Mypes.

ins directly influences the behavior of dataspaces. Further, administrators can adjust the plug-ins and reasoning rules to their needs. In the next section we will highlight GUMF dataspaces and describe plug-ins we implemented that support aggregation of user data from the Social Web. These plug-ins and the corresponding dataspace enable Mypes to visualize the distributed user data traces. Moreover, as GUMF dataspaces can be shared across different client applications, we also enable other client applications to benefit from the profile aggregation.

3 Mypes: Enriching User Profiles via GUMF

To illustrate the functionality of GUMF dataspaces we present the Mypes service which is a GUMF client application that particularly highlights GUMF's ability to support the task of gathering information about users for user adaptive systems [1]. In this section we first present the GUMF components we implemented to allow for profile aggregation and enrichment before we describe Mypes features in more detail.

3.1 Linkage, Aggregation, Alignment and Enrichment of User Data

The Grapple User Modeling Framework aims to provide a uniform interface to user data that might be distributed on the Web. To feature access to distributed user data and to align and enhance the data, GUMF and the corresponding GUMF components depicted in Figure 2 respectively perform the following steps.

1. **Account Mapping** Given a user's URI of an online account, the account mapping plug-in gathers other online accounts of the same user by exploiting the Google Social Graph API[6], which provides such mappings for all users who linked their accounts via their Google profile, for example:

[6] http://code.google.com/apis/socialgraph/

traditional profile attributes	Facebook	LinkedIn	Twitter	Blogspot	Flickr	Delicious	Stumble Upon	Last.fm	Google
nickname	x	x	x	x	x	x	x	x	x
first name	x	x							
last name	x	x							
full name	x	x	x		x				x
profile photo	x		x		x				x
about		x							x
email (hash)	x				x				
homepage	x	x	x						x
blog/feed			x	x	x	x	x	x	
location		x	x		x				x
locale settings	x								
interests		x							
education		x							
affiliations	x	x							
industry		x							
tag-based profile					x	x	x	x	
posts			x	x	x	x	x		

Table 1. Profile data for which Mypes provides crawling capabilities: (i) traditional profile attributes, (ii) tag-based profiles (= tagging activities performed by the user), and (iii) blog, photo, and bookmark posts respectively.

```
"http://www.google.com/profiles/fabian.abel": "claimed_nodes": [
    "http://delicious.com/fabianabel",
    "http://fabianabel.stumbleupon.com",
    "http://www.last.fm/user/fabianabel/", ...]
```

For those users where no mappings can be obtained via API, it is possible to provide appropriate mappings by hand. The account mapping module finally provides a list of online accounts that are associated to a particular user.

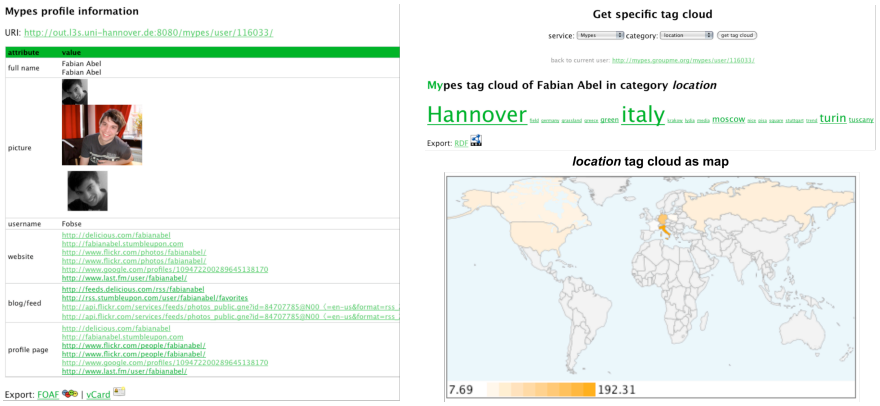
2. **Social Web Aggregator** For the URIs associated to the user, the aggregator module gathers profile data from the corresponding services. In particular, traditional profile information (e.g., name, homepage, location, etc.), tag-based profiles (tagging activities), and posts (e..g, bookmark postings, blog posts, picture uploads) are harvested from nine different services as depicted in Table 1.
3. **Profile Alignment** The profiles gathered from the different services are aligned with GUMF's uniform user model by means of hand-crafted rules, i.e. the user data is modeled by means of Grapple statements [7] using FOAF or vCard as domain-specific vocabulary for the actual user attributes (e.g., name, homepage, etc.). For example, given the *full name* of a user's Google or Flickr profile, GUMF creates a statement as follows.

```
@prefix gc: <http://grapple-project.org/grapple-core/> .
<http://grapple-project.org/gumf/statement-2010-05-19-bob-name>
  rdf:type     gc:Statement;
  gc:user      <http://bob.myopenid.com>;
  gc:predicate <http://xmlns.com/foaf/0.1/name>;
  gc:object    "Bob Mayer";
  gc:created   "2010-05-19T16:23:04.243+01:00" .
```

4. **Semantic Enrichment** Tag-based profiles are further enriched and clustered by means of WordNet[7] categories so that GUMF client applications can, for example, access particular parts of a tag-based profile such as facets related to *locations* or *people*.

The four plug-ins can be plugged into dataspaces. We applied them to the Mypes dataspace which forms the basis for the Mypes service.

[7] http://wordnet.princeton.edu/



(a) traditional profile (b) filtering tag-based profile

Fig. 3. Aggregation of traditional and tag-based profiles.

3.2 Mypes Service

Mypes exploits GUMF to provide an interface that is valuable for casual users, who would like to overview their distributed profile data. Further, it makes the enriched public profile data that is available via the Mypes dataspace (see Figure 2) also available via a lightweight RESTful interface (in FOAF and vCard format). Hence, thanks to GUMF and the plug-ins presented above, Mypes just has to focus on representing the public profile data – either for end-users or for systems that require additional information about their users.

Given the Google profile URI of a user, Mypes queries GUMF and the Mypes dataspace particularly to obtain the corresponding aggregated, enriched profile data. It then enables users to overview their public profiles, for example, users can inspect particular profile values, analyze the completes of their profiles on a bar chart and understand to what kind of information about themselves is publicly available. Mypes thus raises the users' awareness of their profile data distributed across different services on the Web.

Figure 3(a) shows an example of an aggregated Mypes profile, namely the traditional profile attributes gathered from the diverse services (see Table 1). When accessing `http://mypes.groupme.org/mypes/user/116033/rdf` the FOAF profile in RDF/XML syntax is returned. Mypes exports all available values for a profile attribute, e.g., if a user specifies her name differently at the different services then all these different values are provided.

GUMF also connects the tagging activities users perform in the different tagging systems. As the *semantic enrichment* plug-in (see Fig. 2) extends tag assignments with meta-information that states to which WordNet category the corresponding tag belongs to, it is possible to filter the aggregated tag cloud of a user according to WordNet categories. For example, Figure 3(b) shows the aggregated tag cloud filtered so that only tags related to locations are displayed. For this kind of tag cloud, Mypes provides an alternative visualization: tags related to locations are mapped to country codes (using the *GeoNames* Web

service[8]), which are sent to Google's visualization API to draw a geographical intensity map that highlights those countries that are frequently (possibly indirectly) referenced by tags in the profile (see bottom in Figure 3(b)). Mypes also features RDF export for these (specific facets of) tag-based profiles using the Tag Ontology[9] and SCOT[10] vocabulary.

4 Conclusion

In this paper we described how the Grapple User Modeling Framework (GUMF) enables to enrich user profiles with user data gathered from the Social Web. We presented the Mypes service that exploits this functionality to enable casual users to overview their distributed profiles and allows adaptive applications to re-use the aggregated and enhanced profile information for their own purposes. In our future work we plan to evaluate the actual benefit for end-users of adaptive applications which make their adaptation decisions based on the enriched profiles produced by GUMF and Mypes respectively.

Acknowledgments This work is partially sponsored by the EU FP7 project GRAPPLE (http://www.grapple-project.org/).

References

1. Jameson, A.: Adaptive interfaces and agents. The HCI handbook: fundamentals, evolving technologies and emerging applications (2003) 305–330
2. Abel, F., Henze, N., Herder, E., Krause, D.: Interweaving public user profiles on the web. In Kobsa, A., Bra, P.D., eds.: International Conference on User Modeling, Adaptation and Personalization (UMAP), Hawaii, USA, Springer (June 2010)
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. International Journal on Semantic Web and Information Systems (IJSWIS) **5**(3) (2009) 1–22
4. Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., von Wilamowitz-Moellendorff, M.: Gumo - the general user model ontology. In Ardissono, L., Brna, P., Mitrovic, A., eds.: User Modeling. Volume 3538 of Lecture Notes in Computer Science., Springer (2005) 428–432
5. Yudelson, M., Brusilovsky, P., Zadorozhny, V.: A user modeling server for contemporary adaptive hypermedia: An evaluation of the push approach to evidence propagation. In Conati, C., McCoy, K.F., Paliouras, G., eds.: User Modeling. Volume 4511 of Lecture Notes in Computer Science., Springer (2007) 27–36
6. Assad, M., Carmichael, D., Kay, J., Kummerfeld, B.: PersonisAD: Distributed, active, scrutable model framework for context-aware services. In LaMarca, A., Langheinrich, M., Truong, K.N., eds.: Pervasive Computing. Volume 4480 of Lecture Notes in Computer Science., Springer (2007) 55–72
7. Abel, F., Heckmann, D., Herder, E., Hidders, J., Houben, G.J., Krause, D., Leonardi, E., van der Slujis, K.: A framework for flexible user profile mashups. In Dattolo, A., Tasso, C., Farzan, R., Kleanthous, S., Vallejo, D.B., Vassileva, J., eds.: Int. Workshop on Adaptation and Personalization for Web 2.0 at UMAP '09, CEUR Workshop Proceedings (2009) 1–10

[8] http://www.geonames.org/
[9] http://www.holygoat.co.uk/projects/tags/
[10] http://scot-project.org/scot/