Discovering User Profiles

Riddhiman Ghosh, Mohamed Dekhil
Hewlett-Packard Laboratories, 1501 Page Mill Rd., Palo Alto, CA
{riddhiman.ghosh, mohamed.dekhil}@hp.com

ABSTRACT

In this paper we describe techniques for the discovery and construction of user profiles. Leveraging from the emergent data web, our system addresses the problem of sparseness of user profile information currently faced by both asserted and inferred profile systems. A profile mediator, that dynamically builds the most suitable user profile for a particular service or interaction in real-time, is employed in our prototype implementation.

Categories and Subject Descriptors: H.3.4, H.3.5 [**Information Storage and Retrieval**]: Systems and Software – *User profiles and alert services;* Online Information Services – *Data Sharing.*

General Terms: Design, Management.

Keywords: User profiles, information management, semantic web, personalization

1. INTRODUCTION

A user profile, used in service personalization, is a structured construct containing information both directly and indirectly pertaining to a user's preferences, behavior and context. Effective personalization requires services to build and maintain accurate models of a customer's preferences, interests and background through a user profile. While building effective profiles can benefit from research contributions in different areas, including security, usability, statistical prediction and mining, in this paper we focus on the information management challenges in the creation and management of user profiles. We address technical challenges relating to the cold-start problem in user profiles and in discovering suitable profiles based on service requirements and user context. Our paper specifically addresses the following challenges:

- *Profile Sparseness:* Services a user is interacting with for the first time, or newly launched services face the problems of having extremely sparse information models about a user, known as the cold-start problem. Asserted (questionnaire-based) profile capture systems have sparse information since typical users don't enter complete information for every service they encounter—they might have already entered the information elsewhere and may not like to repeat themselves. Inferred profile systems augment questionnaire systems through mining techniques such as mining user clickstreams or interaction patterns. However these services are also prone to make erroneous or very limited inferences without significant user history, and do not make use of profile-relevant data available with other services.
- *User Profile Personae:* A user's context and environment while he is interacting with a service impacts both his behavior and preferences. Users may "wear different hats" and have different modes when interacting with a service, e.g. when they are at home vs. at work, on a weekend or a weekday, shopping in "computer scientist mode", vs. in "baseball fan mode", etc. Currently it is not easy to have different user profile personae—instances of a user profile specific to a particular context—accessible to a service, or to choose which persona to reveal to a service.
- *Using out-of-band data:* In our earlier work [2], we have argued that to enrich user profiles services need to use *out-of-band* profile data, i.e. fragments of the user's interactions or preferences with other services. Point-to-point data sharing schemes and agreements are currently required, and do not always exist due to business reasons. Moreover it is hard to reuse user profile data from different services due to a proliferation of terms used to describe the same details about a user (referred to as "babelization").

WWW 2009, April 20-24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

2. MEDIATED PROFILE DISCOVERY

We have built a user profile management framework in order to address the technical challenges mentioned in the previous section. This framework is a part of HP Labs InfoPal, which is a services and analytics platform for personal projects. The key features of the InfoPal Profile Manager (IPM) are its support for *integrating and using out-of-band profile information*; incorporating and exposing *machine-accessible semantics* for all profile data; providing facilities for *dynamically discovering and building* a suitable user profile from fragments of existing profiles; and support for *user profile personae*. The IPM exposes its functionality through REST-style web-services, which can then be used by different services for create/read/update/delete operations on user profile data—a *profiles-as-a-service* approach.

2.1 Profile Aggregation: User profiles are much more than just a list of "interest" keywords—they hold information regarding user behavior, context and other preferences. We build upon our work in [2], where we showed just as Web 2.0 mashups have allowed for the creation of hybrid applications, so too can user profiles be enriched by building them up from the profile fragments of a user distributed across multiple services. This is possible because of the increasing emergence of services that export their data through standard APIs. The two problems encountered in this approach are, first, the semantic heterogeneity between data exposed by different sources, and second, the issue of lack of central ownership of the data—multi-way data sharing agreements between different services do not exist. Our approach to the latter problem is to take the view that the user is the owner and the control point for his data, and we have developed user-driven aggregation tools for profile data, that build and maintain a user-owned profile corpus. In order for IPM to aggregate and expose profile data to and from other applications/services, all profile information has associated with it explicit semantics using the standards from the Semantic Web (RDF/OWL). We have defined an ontology for our domain, the InfoPal User Profile Ontology (IUPO), reusing concepts from existing well-known ontologies such as FOAF. A version of the IUPO OWL ontology, in N3 syntax, can be inspected at [3]. This ontology, with associated data adapters, acts as the unification point for the information from the different profile sources.

2.2 Dealing with Sparseness: IPM addresses the issue of profile sparseness mentioned in Section 1 by allowing a service to request the IPM mediator to build a profile for the user which is suitable for that particular service, from the profiles that exist in a user-owned profile corpus. This is a collection of the profiles of the user as maintained by different services. Even though data sharing agreements between these different sources may not be in place, the user owns his profile information, and user-driven aggregators/agents are used to build this corpus. Each of the user profiles from these different sources stored in the corpus may be described with a specific ontology.

As shown in Figure 1, the service sends the mediator a desired user profile specification (a profile ontology, or a sample profile). This can be considered a "discovery-by-example" approach for locating and constructing user profiles. If after querying the profile corpus, a user profile conforming to the desired ontology description is not found, and a well-known mapping does not exist, the IPM mediator attempts

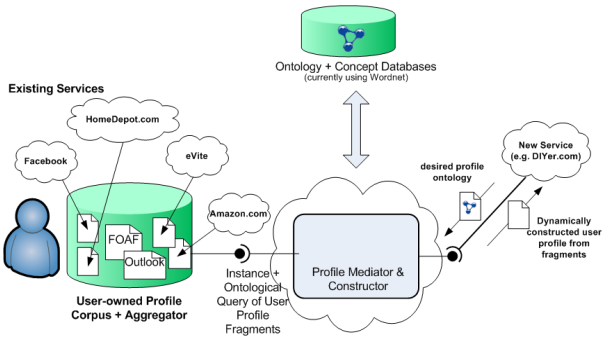

Figure 1: Profile Discovery-by-Example

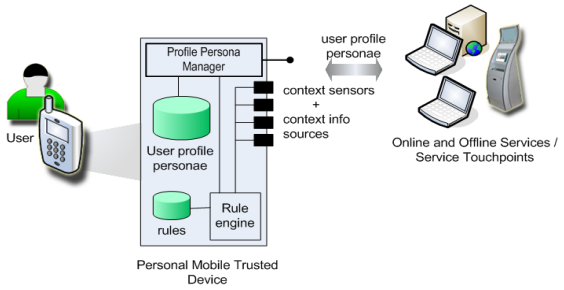

Figure 2: Mobile Profile Personae

to calculate the semantic distance between the constituents (resources and properties) of the ontologies (i.e. terms in the profile *vocabularies*) in the corpus with respect to the specification provided by the requesting service. Profile facts in the corpus within a thresholded semantic distance will be used to construct a user profile for the requesting service. Our initial experiments at calculating the semantic distance have centered on using Princeton's WordNet lexical database, leveraging similarity measures described in [4]. The full version of this paper will describe our experiments, heuristics developed to deal with compound vocabulary terms, and the limitations of this approach. As an example, a user's profile corpus may contain a Facebook.com profile, a HomeDepot.com profile and an Amazon.com profile. If a new do-it-yourself web site, say DIYer.com, needs to build a user profile, the mediator will be able to determine that profile information in his HomeDepot.com profile is the most relevant, and perhaps a subset of his profile pertaining to DIY books or purchases on his Amazon.com profile are relevant, but not information in his Facebook.com profile. The IPM will then dynamically construct a user profile for use by DIYer.com for customizing the user's experience. This is possible even though the user *has not had a significant interaction history with DIYer*, and in spite of there *not being data-sharing agreements among the different services*.

2.3 User Profile Personae: IPM supports the notion of user profile personae; the IPM service allows users to store different profile personae and choose which persona to forward to a service. A user can have multiple personae per service, or use the same persona with different services. IPM personae, like profiles, are expressed with machine-accessible semantics to allow for easier use across multiple services. The different personae can be stored in the "cloud", or as we have been exploring, on a user's mobile device as a trusted and portable medium of exchanging profile personae with services. Mobile devices are typically sensor-rich and can determine a user's context and environment (low- and high-level); this can aid in deciding the most suitable persona to reveal to a service. We have extended our prior work on using mobile devices for service personalization [6] to realize in IPM a proof of concept of user profile personae. As can be seen in Figure 2, the Profile Persona Manager client component of IPM chooses the most suitable persona from the profile persona corpus to deliver to a service, based on evaluation of service-to-persona matching rules.

3. IMPLEMENTATION

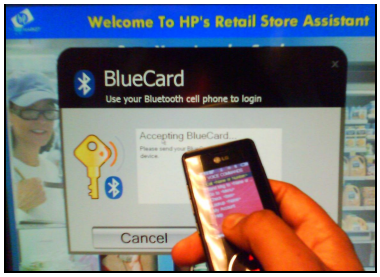

Figure 3: Projecting a User Profile Persona

The IPM prototype has been built and has been used in a retail focused application of InfoPal, customizing a user's interaction based on his aggregate profile. This was obtained using user profile aggregation tools we've built to collect calendar/event information (from Google Calendar using the GData services), from FOAF social network information, and from explicit user intent from our user intent capture framework [5]. Profile manager functionality is exposed via REST Web Services. We've also built the user profile persona capability, and Figure 3 shows a screenshot of a user introducing himself to a service via a chosen profile persona on his mobile device. The current implementation does not provide for automated contextual switching of personae; we are working on a demonstrator, that combines context information (both low- and high-level) in deciding the appropriate profile persona.

4. CONCLUSION AND RELATED WORK

In this paper we presented techniques for the discovery and construction of user profiles, using a profile mediator to build the most suitable user profile for a service. We presented solutions to enhance profile information with contextual data, address the profile sparseness problem faced by personalization services, and allow for user profile personae. In terms of related work, asserted-(questionnaire), e.g. [7], and inferred-profile systems are commonly seen on the Web. However they alone are not enough for effective profile capture. Users may not enter complete information at every service. Also, information needed for personalization is often orthogonal or contextual in nature to the service or interaction at hand. This information either lies elsewhere, or was not anticipated when the questionnaire was designed. Another approach is domain-specific data sharing schemes between two or more parties. An example of this is the Facebook Beacon service [10] which has arrangements with pre-defined retail vendors to share user profile data to contribute to a Facebook user profile. Similarly [1] outlines techniques for explicit data sharing schemes to build user profiles. The disadvantage with this approach is that explicit point-to-point data sharing schemes and agreements are needed by the different services. Google OpenSocial [8] and Yahoo Y!OS [9] are also trying to address some of the challenges in Section 1. However, approaches that use business or incumbency leverage to require all services to use their particular API and format for managing profiles may be lacking, given that linguae francae on the Web can seldom be enforced.

5. REFERENCES

[1] Houben, J., et al., State of the art: semantic interoperability for distributed user profiles, Telematica Institut Report, 2005.
[2] Ghosh, R., Dekhil, M., Mashups for semantic user profiles, *Proc. 17th Intl. World Wide Web Conf.*, Beijing, 2008.
[3] InfoPal User Profile Ontology, www.purl.org/net/iupo
[4] Pedersen, T., et. al., WordNet::Similarity – Measuring the similarity of concepts, AAAI 2004.
[5] Jain, J., Ghosh, R., Dekhil, M., "Multimodal capture of consumer intent", *26th ACM SIGCHI Conf. on Human Factors*, Florence, 2008.
[6] Ghosh, R., Dekhil, M., BlueCard: Mobile device-based authentication and profile exchange, *Proc. 3rd IEEE Intl. Conf. on Systems, Computing Sciences and Software Engineering*, 2007
[7] Windows Live ID, https://accountservices. passport.net /ppnetworkhome.srf
[8] Google OpenSocial, code.google.com/ opensocial.
[9] Yahoo Y!OS, developer.yahoo.com/yos
[10] Facebook Beacon Service, http://www.facebook.com/business/?beacon