TheHiddenU*
A Social Nexus for Privacy-Assured Personalisation Brokerage

G. Kappel, J. Schönböck, M. Wimmer
Vienna University of Technology, Favoritenstrasse 9-11, A-1040 Vienna
{gerti | johannes | manuel}@big.tuwien.ac.at

G. Kotsis, A. Kusel, B. Pröll, W. Retschitzegger, W. Schwinger, R. R. Wagner
Johannes Kepler University Linz, Altenbergerstrasse 69, A-4040 Linz
{gabriele.kotsis | angelika.kusel | birgit.proell | werner.retschitzegger | wieland.schwinger | roland.wagner}@jku.at

S. Lechner
Netural Communication, Europaplatz 4, A-4020 Linz,
s.lechner@netural.com

Keywords: Social Networks, Personalization, Profiles, Integration, Service Brokerage

Abstract: **Abstract**. Social networks have seen enormous growth over the past few years, providing also a powerful new channel for distributing personalized services. Personalization, however, is exacerbated because social content is scattered across different social networks serving specific human needs and social networkers are particularly reluctant to share social content with service providers, if not under their full control. This paper sketches TheHiddenU, a social nexus exploiting semantic techniques for integrating, profiling and privatising social content, thereby providing the technical prerequisites for personalized brokerage, a new, sustainable business model in the Social Web.

1 INTRODUCTION

Social networks have seen enormous growth being already used by more than two-third of the world's Internet population (Nielsen, 2009). Thus, *service providers* like shopping sites, travel portals, or news sites are urgently forced to exploit social networks as additional distribution channel for providing highly personalized services to *social networkers*. Currently, however, both social networkers and service providers are in an undesirable situation (Halpin, 2009). Firstly, *social content is scattered* over several social networks each serving specific human needs like sharing friendships, business contacts or music interests. Secondly, these isolated views bear the pitfall of *untargeted services*. Thirdly, social networkers have a great sense of ownership being reluctant to share social content with service providers for personalization, at least if it is *obscure* and *beyond their control*, how their social content is being exploited (Kobsa, 2007b).

Let's consider an online media retailer intending a recommendation service for birthday gifts on time, matching the interests of friends while fitting the donor's economic situation – *if only they could know*. At the same time, social networkers might use facebook[1] to connect to friends, Last.fm[2] to serve music interests, and XING[3] for business issues – *if only they could oversee*.

TheHiddenU system proposed herein is intended to act as a social nexus, *integrating* scattered social content into a common social profile and, on basis of that, explicating formerly hidden knowledge by means of *profiling*. At the same time, TheHiddenU can act as a personalization broker. This means that social networkers are enabled to share, in a *privacy-assured* way, necessary profile information with the service provider to receive, e.g., gift recommendations.

The paper is structured as follows. The next section describes TheHiddenU's architecture and approach. Afterwards the envisioned prototypical implementation is discussed, proposing concrete technologies, which seem to be suitable for our purposes. Finally, a comparison of related approaches and first ideas concerning the evaluation of TheHiddenU's feasibility conclude the paper.

* funded by Austrian Federal Ministry of Transport, Innovation and Technology under FIT-IT Semantic Systems (819577, 819584).

[1] www.facebook.com
[2] www.lastfm.com
[3] www.xing.com

2 ARCHITECTURE AND APPROACH

The architecture of TheHiddenU, as depicted in Figure 1, is divided into two major components, a *Sandbox* which is under control of the social networker for integrating, profiling and privatizing social content, and a *Brokerage* area for connecting service providers to gain personalized services. In the following, these two components are described in more detail in a process-oriented manner.

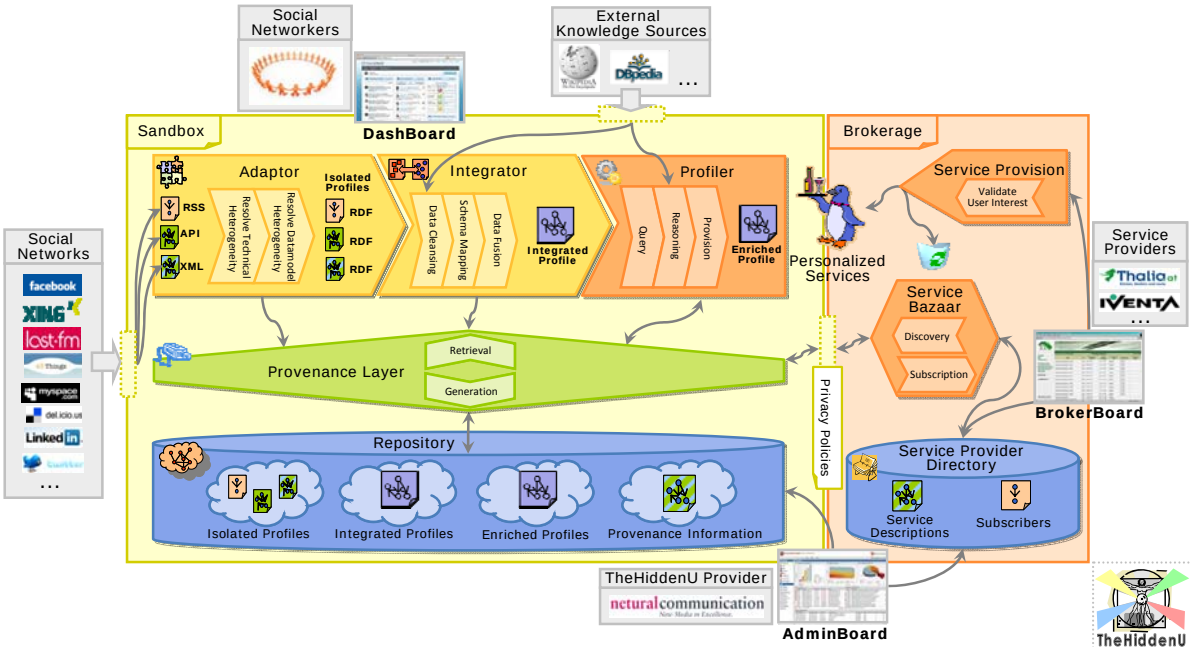
2.1 Hybrid Approach for Social Content Integration

First of all, TheHiddenU provider connects social networks via the *AdminBoard* by *adaptors* resolving technical and data model heterogeneities, resulting in *Isolated Profiles* for each source stored in a *Repository*. Further on, the *Integrator* component allows for a hybrid integration involving both, *schema level* and *instance level*. For this, TheHiddenU provider can apply data *cleansing operators* to resolve lexical heterogeneities, *mapping operators* to resolve structural ones and data *fusion operators* for dealing with duplicates and conflicts to achieve an *Integrated Profile*. Social networkers, in turn, can view, verify and change their integrated profile via the *DashBoard*.

2.2 Semantic-based Profiling & Brokerage

Service providers and social networkers will be provided with an extensible library of reusable *profiler components* and a *language* for their composition. First, *input* components allow for querying recorded profile and provenance data and external sources. Second, *reasoning* components are foreseen for *inferring* new knowledge like a “best-friend” relationship and for *refining* and *augmenting* existing knowledge like computing the strength of an "interested-in” relationship or associating favorite music tracks to external music categories. Third, *provision* components are provided for materializing the newly derived knowledge, resulting in a *Semantically Enriched Profile*.

For brokerage, service providers register their services via the *BrokerBoard* in the *Service Provider Directory*, indicating those parts of the social content necessary for a service to operate (e.g., the gift's occasion) and optional ones (e.g., a price range), together with possible provision channels (e.g., eMail or widget). Social networkers can discover and subscribe services on basis of interest-driven and collaborative filtering using the *DashBoard*. The *ServiceBazaar* realizes a negotiation process, allowing social networkers to balance between privacy concerns and personalization needs. Succeeding service provision takes place in a pseudonymous way.



.
Figure 1: System Architecture of TheHiddenU

2.3 User-empowered Provenance & Privacy-Control

The *ProvenanceLayer* automatically records all information about integration and profiling activities within the Repository. By means of a *ProvenanceViewer* which is part of the *DashBoard*, social networkers are supported to understand which social content is stored, how it is computed, and how it is used. Privacy concerns of social networkers are made explicit in terms of adaptable privacy policies with respect to social content (i.e. social content policies) and with respect to service providers (i.e. service provider policies).

3 PROTOTYPICAL IMPLEMENTATION

The rationale behind the selection of techniques intended for the implementation of TheHiddenU is mainly based on our experience in the areas of *ontology-driven situation-awareness* (Baumgartner et al. 2010), *model-driven engineering* (Wimmer et al. 2009) and *aspect-oriented software engineering* (Wimmer et al. 2010). To start with, the central repository will build on existing RDF triple stores like *Jena*[1]. To provide a unified and extensible representation of social content, an *ontology* will be developed, incorporating the core concepts prevalent in user models (Kobsa, 2007a) (e.g., personal data, user knowledge, capabilities, interests, and relationships) and in social ontologies (Sutterer et al. 2008) and standards like *OpenSocial*[2], *FOAF*[3] (Friend of a Friend) allowing to describe persons, their activities, and relationships, *SIOC*[4] (Semantically-Interlinked Online Communities) contributing interactions and resources, and *RELATIONSHIP*[5], specializing relationships (e.g., family, professional). The adaptors acquiring knowledge from social networks, should be based on existing approaches e.g., the *Linking Open Data Project*[6] and the *Data Portability Project*[7]. To realize our user interfaces, *DashBoard*, *AdminBoard* and *BrokerBoard*, the open source Ajax framework *ICEfaces*[8] will be used.

[1] jena.sourceforge.net
[2] www.opensocial.org
[3] www.foaf-project.org
[4] sioc-project.org
[5] vocab.org/relationship.html
[6] tinyurl.com/OpenDataProject
[7] www.dataportability.org
[8] www.icefaces.org

3.1 Social Content Integration

For realizing the Integrator, we will apply *model-driven techniques*, in particular the *Eclipse Modeling Framework (EMF)*[9] and the *Graphical Modeling Framework (GMF)*[10], allowing a graphical representation of the integration operators in a mashup-based manner. For the definition of new integration operators a domain specific language should be offered whereby, e.g., the *Epsilon Merge Language*[11] would be well suited. Finally, the resulting mix of code snippets must be transformed e.g., on basis of OpenArchitectureWare[12] into our runtime environment Jena allowing for a proper mix of SPARQL and Java code.

3.2 Profiling and Brokerage

Profiling will be based on the same technologies as used for the Integrator since we envision again a mashup-based approach, naturally differing in the realized operator logic and the underlying execution platform. Reasoning operators should make use of, both, *declarative* techniques on basis of generic rules (e.g., subsumption) as well as user-defined ones and *imperative* techniques in case declarative reasoning is not flexible enough. In this respect, SWRL[13] seems to be promising, allowing the combination of OWL with RuleML on basis of inference engines like Pellet[14] or AllegroGraph[15] providing declarative reasoning with RDFS++ and logic programming with Prolog. For service brokerage, we will base on existing semantic web standards like UDDI and OWL-S.

3.3 Provenance and Privacy

For introducing provenance recording into the integration and profiler components, Higher-Order Transformations (HOTs) expressed by the *ATLAS Transformation Language*[16] will be used, the Provenance Viewer will be again based on EMF/GMF. Concerning privacy, the approaches of Personis (Kappel et al. 2009) and Kobsa (Kobsa, 2007b) are of foremost interest since Personis puts forward the privacy principle of scrutability and Kobsa focuses on dynamic, adaptable privacy policies.

[9] www.eclipse.org/emf
[10] www.eclipse.org/gmf
[11] www.eclipse.org/gmt/epsilon
[12] owa.itemis.de
[13] ruleml.org/swrl
[14] clarkparsia.com/pellet
[15] www.franz.com/agraph
[16] www.eipse.org/m2m/atl

4 RELATED WORK

There is already plethora of related approaches in the areas of integrating, profiling and privatizing social content, which is briefly reviewed in the following. For a more detailed comparison of TheHiddenU to related work it is referred to (Kappel et al. 2009).

4.1 Social Content Integration

Recently, commercial *meta social networks* like *Power.com*[1], *snag*[2] or *OrSiSo*[3] emerged, providing a single-point-of-access to other social networks but miss an integrated view resolving heterogeneities as intended by TheHiddenU. Research projects already build on an ontology-based infrastructure, like (Gosh et al. 2009), (Golemati et al. 2007), where a social profile is built up by aggregating content spread over multiple services, or (Noor et al. 2009), where tags of a certain social networker are analyzed to build up a social profile for recommendations. In contrast to TheHiddenU, however, brokerage is not an issue. Regarding integration, in the last 30 years, numerous approaches where brought forward in the areas of *data engineering* (Bleiholder et al. 2008), (Elmagarmid et al. 2007), (Rahm et al. 2000), *ontology engineering* (Baumgartner et al. 2010), (Le-Phuoc et al. 2009) and *model engineering* (Bleiholder et al. 2008), (Elmagarmid et al. 2007), which are of high value for our approach. Nevertheless, the majority of them address the whole spectrum of integration tasks necessary in TheHiddenU in an isolated way, not least since data cleansing and fusion deal with instances whereas schema mapping deals with types.

4.2 Profiling and Brokerage

Regarding the profiling aspect of TheHiddenU, first of all, research in the area of personalization and recommender systems (Bonhard et al. 2007), (Kobsa et al. 2007a) is of high relevance. A crucial challenge will be, however, to reason with incomplete and vague information since it can neither be predicted which information social networkers have entered, nor which social networks they have joined. In (Lukasiewicz et al. 2009), formal and practical means for automated reasoning with incomplete and vague information are proposed, which will be adapted to our context. Concerning existing reasoning techniques, in (Huang et al. 2008) an overview on different approaches is given, whereby rule-based reasoning for dynamic and incomplete knowledge best fits our requirements. More widely related to our approach are social network analysis methods which focus on exploring general properties of a whole social network using graph theory (Ereteo et al. 2008), whereas we intend to reason about a single social networker. Concerning the brokerage aspect, relevant work was brought forward in the coalescing areas of Service Oriented Architecture (SOA) and semantic web as well as more distant, also in the area of agent-based and self-organized systems (Dustdar et al. 2009).

4.3 Provenance and Privacy

Concerning provenance, research primarily focused on describing the *origins of data* (Tan, 2007) in terms of *workflow-oriented* approaches providing a coarse-grained focus on data processing, *data-oriented approaches* aiming at a fine-grained view and approaches incorporating *social aspects* (Harth et al. 2007). Only few explicit models, formally representing provenance have been published, being a recent research topic especially in combination with privacy aspects (Kobsa, 2007b). Approaches for *retrieval and visualization* of provenance most often do not go beyond simple provenance queries (Simmhan et al. 2005), although being overwhelming for users (Chapman et al. 2007). Finally, privacy approaches in the realm of personalization have been postulated recently, only (Kobsa, 2007b).

5 EVALUATION METHOD

The feasibility of TheHiddenU will be evaluated in three phases. Firstly, the proposed integration approach will be analyzed in the course of *experiments*, establishing in particular a test-set for content integration by acquiring real world data of volunteer social networkers, supplemented by synthetic data containing diverse heterogeneities. Secondly, expressivity and productivity gained by our profiling language will be evaluated in an *empirical study*, incorporating about 100 students of our semantic web course. Finally, *case studies* will be conducted together with our demonstrators, the *eMedia retailer Thalia*[4] and the *eRecruiting company IVENTA*[5], in order to evaluate the TheHiddenU in a real world setting. In particular, three exemplary social networks will be integrated by the system provider, dedicated profilers will be developed by our demonstrators to personalize their services and volunteers will refine their integrated

[1] ww.power.com
[2] ww.dapper.net/applicaitons/Snag
[3] www.orsiso.com
[4] www.thalia.at
[5] www.inventa.at

profile and subscribe to services. On basis of structured interviews, it will be explored, first, to which extent personalization and service provision met the volunteer's requirements, second which social content is shared with the service provider to analyze privacy attitudes and third, to what extent the provided provenance information procured privacy control over the volunteers social content.

REFERENCES

Baumgartner, N., Gottesheim, W., Mitsch, S, Retschitzegger, W., Schwinger, W.: BeAware! Situation Awareness, the Ontology-Driven Way. accepted for *DKE Journal*, Elsevier, 2010.

Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. *Int. Journal on Semantic Web and Information Systems (IJSWIS)*, 2009

Bleiholder, J., Naumann, F.: Data fusion. *ACM Computing Surveys*, 41(1), 2008

Bonhard P., et al.: The devil you know knows best – How online recommendations can benefit from social networking. *Proc. of the 21st British CHI Group Annual Conf. on HCI*, 2007

Chapman, A., Jagadish, H. V.: Issues in Building Practical Provenance Systems. *IEEE Data Eng. Bull*., 30(4), 2007

Dustdar, S., et al.: Self-Adaptation Techniques for Complex Service-oriented Systems. *Proc. of 5th Int. Conf. on Next Generation Web Services Practices*, IEEE, 2009

Elmagarmid, A. K., Ipeirotis, P. G., Verykios, V. S.: Duplicate Record Detection: *A Survey. IEEE Transactions on Knowledge and Data Engineering*, 19(1), 2007

Ereteo, G., et al.: A State of the Art on Social Network Analysis and its Applications on a Semantic Web. *Proc of the 1st Social Data on the Web Workshop (SDoW'08)*, Karlsruhe, Oct. 2008

Ghosh, R., Dekhil, M.: Mashups for semantic user profiles. *Proc. of the 17th Int. Conf. on World Wide Web (WWW'08)*, ACM, 2008

Golemati, M. et al.: Creating an Ontology for the User Profile: Method and Application. *Proc. of the 1st Int. Conf. on Research Challenges in Information Science (RCIS'07)*, 2007

Halpin, H.: Ten Theses on the Future of Social Networking. *Proc. of W3C Workshop on the Future of Social Networking*, Barcelona, 2009

Harth, A., Polleres, A., Decker, S.: Towards a Social Provenance Model for the Web. *Proc. of Workshop on Principles of Provenance (PrOPr'07)*, 2007

Huang, Z. et al.: A Survey of Web Scale Reasoning. FP7 Deliverable, LarKC/2008/D4.1, 2008

Kappel, G., Kotsis, G., Kusel, A., Lechner, S., Pröll, B., Retschitzegger, W., Schönböck, J., Schwinger, W., Wimmer, M.: TheHiddenU - A Social Nexus for Privacy-Assured Personalisation Brokerage, *Technical Report, University Linz*, 2009, http://www.ifs.jku.at/download/TheHiddenU.pdf

Kay, J.: Scrutable Adaptation: Because We Can and Must. *Proc. of 4th Int. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems*, Springer, LNCS 4018, 2006

Kobsa, A.: Generic User Modeling Systems. The Adaptive Web, Springer, LNCS 4321, 2007a

Kobsa, A.: Privacy-enhanced personalization. *Communication of the. ACM* 50(8): 24-33, 2007b

Lukasiewicz, T., Straccia, U.: Description Logic Programs under Probabilistic Uncertainty and Fuzzy Vagueness. *Journal of Approximate Reasoning*, 50(6), 2009

Nielsen: Global Faces and Networked Places – Report on Social Networking's New Global Fingerprint. http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/03/nielsen_globalfaces_mar09.pdf, 2009

Noor, S., Martinez, K.: Using Social Data as Context for Making Recommendations: An Ontology based Approach. *Proc. of the Workshop on Context, Information And Ontologies (CIAO'09)*, 2009

Le-Phuoc, D. et al.: Rapid prototyping of semantic mash-ups through semantic web pipes. *Proc. of the 18th Int. Conf. on World Wide Web (WWW '09)*, ACM, 2009

Rahm, E., et al.: Data Cleaning: Problems & Current Approaches. *IEEE Data Eng. Bull*., 23(4), 2000

Simmhan, Y. L. et al.: A survey of data provenance in e-science. *SIGMOD Rec.* 34(3), 2005

Sutterer, M. et al.: UPOS: User profile ontology with situation-dependent preferences support. *Proc. of the 1st Int. Conf. on Advances in Computer-Human Interaction*, IEEE, 2008

Tan, W.: Provenance in Databases: Past, Current, and Future. *IEEE Data Eng. Bull*., 30(4), 2007

Wimmer, M., Kusel, A., Schoenboeck, J., Kappel, G., Retschitzegger, W., Schwinger, W.: Reviving QVT Relations: Model-based Debugging using Colored Petri Nets. *Proc. of ACM/IEEE 12th Int. Conf. on MDE Languages and Systems (MODELS'09)*, Denver, 2009

Wimmer, M., Schauerhuber, A., Retschitzegger, W., Schwinger, W., Kappel, G., Kapsammer, E.: A Survey on UML-based Aspect-Oriented Design. accepted for *ACM Computing Surveys*, 2010.