Identifying Relevant Fragments of Learner Profile on the Semantic Web*

Peter Dolog
L3S Research Center
University of Hannover
Expo Plaza 1, 30539 Hannover, Germany,
dolog@l3s.de

Abstract

The aim of this paper is to discuss how to identify distributed learner profile fragments on the semantic web. The learner profile fragments are modelled employing vocabulary suggested by several standards for learner profile. The learner profile fragments are maintained as standalone semantic networks of objects in RDF. The objects are instances of concepts labeled by terms from the standards. The identification of the profile fragments needed for example by adaptation services is performed as unification of identification records maintained on different sites. Queries sent to the edutella P2P network provide virtual views which connect those stand alone object networks. The queries can be constructed according to specific needs of personalization techniques, which can be provided as personalization services in a P2P learning network.

Keywords: Distributed User Modelling, RDF/RDFS, Learner Profile Fragment, Learner Profile Standards

1 Introduction

Recent advances in technologies for web-based education provide learners with a broad variety of learning content available. Learner may choose between different lecture providers and learning management systems to access the learning content. On the other hand, the increasing variety of the learning material influences effort needed to select a course or training package which will effectively build skills required for changed business situation. Adaptive support based on learner needs, background and other characteristics can help in selecting appropriate learning and during learning.

Information about a learner is crucial for enabling such adaptation. As the learner may take courses and training in different learning management systems, fragments of his profile are maintained on different sites. The systems should be able to collect those fragments to enable adaptation. This situation raises a question how to identify the relevant fragments of a learner profile distributed over the systems.

*This work is partially supported by EU/IST ELENA project IST-2001-37264 (http://www.elena-project.org).

In this paper we discuss an approach how to identify the distributed learner profile fragments in P2P environment. The fragments are maintained in RDF according to a vocabulary prescribed by standards for learner profiles.

The rest of the paper is structured as follows. Section 2 provides a sample scenario which drives the descriptions in the paper. Section 3 provides a discussion on simplified user conceptual model typically used in adaptive systems based on terminology taken from several learner profile standards. Section 4 discusses our approach to identification of learner profile fragments based on local identification schemes. Related work is discussed in the Sec. 5. Paper conludes with sumary and remarks on ongoing research (Sec. 6).

2 Sample Scenario

To motivate our approach we refer to a sample scenario. Alice is trying to improve her skills in programming of accounting software. She has a degree in computer science and experience in programming of a text editor. She is looking for a training package where she will experience common problems and approaches in programming the accounting software. Alice has an application to access and search a network of learning providers. Her profile about her learning performance at the university is available from the university provider. Her portfolio is available directly from her application.

As the situation shows, the Alice profile fragments have to be retrieved from several places. Those places usually use different identification mechanisms. For example, university identifies a learner by his matriculation number. The company has its own identification scheme for identifying its employees. Alice uses application which employs different identification scheme as well.

Figure 1 depicts the architectural outline for the Alice scenario. Alice accesses the provided courses through her personal learning assistant (PLA). The PLA uses the Edutella consumer to query connected systems. The PLA maintains the identification entries used at the previously accessed systems (the University and CompuTraining provider in our case). The university provider maintains Alice performance during her university studies. The training provider followed Alice performance in the course on programming accounting software and stores it in its metadata stores.

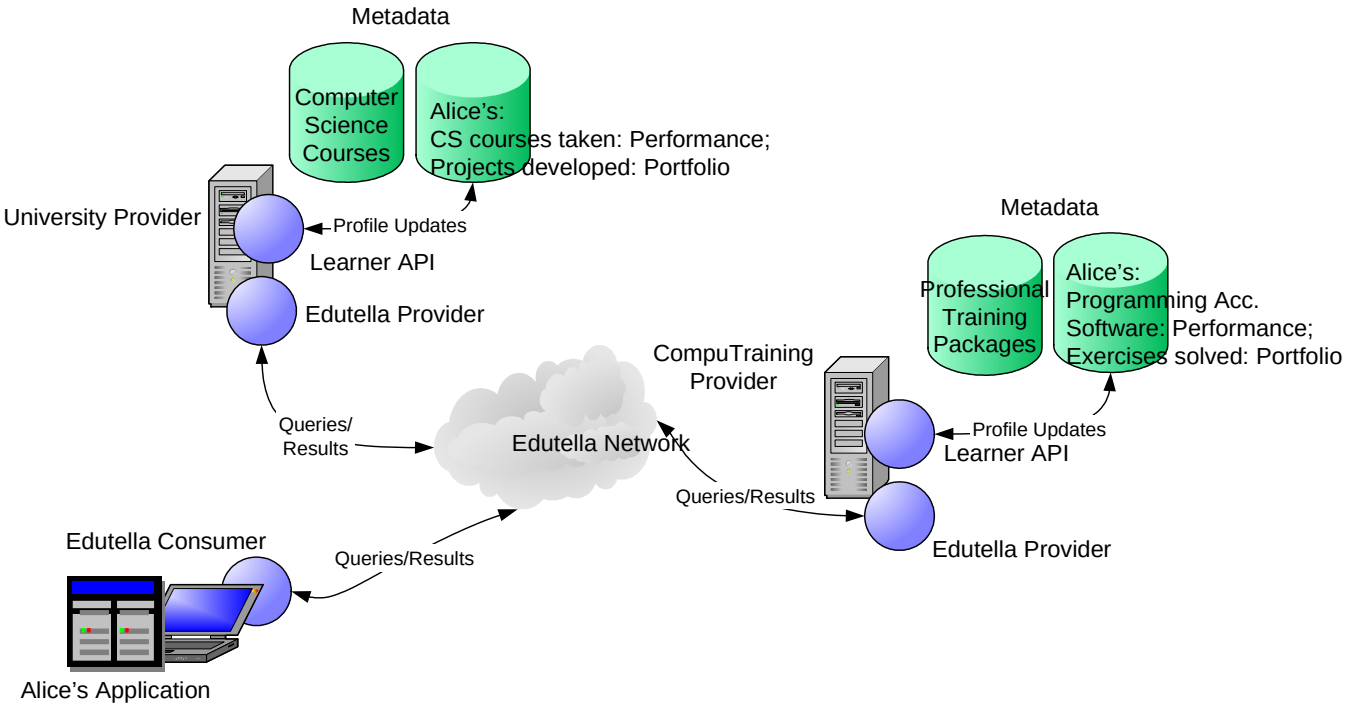

Figure 1: An architecture showing Alice scenario

Both learning providers could use additional external services which followed Alice performance. There are two possibilities to handle this situation. In the first case, the services maintain Alice performance records identified by their identification schemes. The learning provider provides a routing and mapping between its scheme and the service identification schemes. In the second case, the Alice performance from the services is stored at the learning provider. Both situations are possible, thus an algorithm for collecting learner profile fragments has to support both situations.

In addition, the providers and services could use different data models. Data integration problem has to be studied in this context, to be able to exchange learner profile fragments between learning services.

In the following, we will address issues related to learner identification on different distributed systems while the data model for learner profile fragments stays uniform.

3 Modelling Learner Features

Semantic web description formats allow us to express information as a network of associated objects described by a certain type. Therefore, each system, which Alice used to access her training or course, maintains a small network of objects describing Alice in each relevant node of the learning provision network.

The main concepts identified in scenario are performance, portfolio and learner as such. Current versions of learner profile standards provide vocabularies to describe such concepts as discussed for example in [Dolog and Nejdl, 2003]. The use of standards allows us to reduce variability in data models used to maintain learner profile records. For example, the IEEE PAPI describes learner performance as a learning experience measured by achieved competency value and portfolio as anything created during the learning experience or anything which supported the learning experience. Both concepts are described by its properties.

The performance and portfolio objects have to be associated to an object which represents Alice (instance of the Learner class). Such objects have several performance and portfolio records and possibly their real name. To enable multiple identifications (pseudonyms), the Learner class points to several identification records which belong to different systems (providers). This allows us to route requests to particular providers and to use object identifiers used at those providers. As the identification might be time limited, "valid to" and "valid from" dates can be associated to the identification records.

Figure 2 depicts a conceptual model needed for the Alice scenario discussed above. The conceptual model is an excerpt of the conceptual model used in Elena project. Further concepts have been considered, such as learner goal, preferences, competencies, and certificates.

4 Identification of Relevant Distributed Fragments of Learner Profile

According to above proposed conceptual model, any system can choose its own identification mechanism. The system can assign locally unique identifiers to distinguish learners. However, it is required to provide the identifiers according to the conceptual model described above. Learner accesses train-

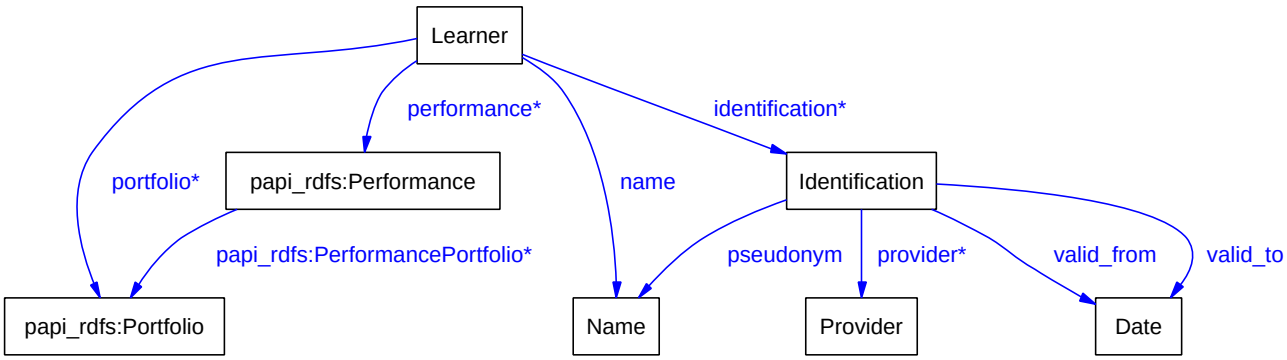

Figure 2: Conceptual model for learner profile from Alice scenario

ing and courses through his personal learning assistant (PLA). However, the PLA uses learning provider services with own identification mechanism. The providers can expose learner identifier used to identify learner a performance record belongs to. The identifier can be then store together with the provider identifier at the learner's PLA. Similarly, if a learning provider accessed further external services, the learner identifiers at those services have to be provided.

Figure 3 depicts an excerpt of instances from Alice' performance and identifications. Alice is identified as Al at the university provider and as Li at the training provider. The model also contains an instance about her learning experience at the university in a course on programming. The learning experience at the CompuTraining in the course on programming accounting software is also depicted.

The parts of the figure which are overlapping are maintained independently at the providers. The unification of the identifications for particular systems is performed when an adaptive system/service searches for learner profile fragments needed for adaptation. The systems can maintain the learner profile fragments by learner API designed according to the learner profile fragments schemas. The API, schemas, and a system prototype for browsing such learner profiles can be found at `http://www.l3s.de/~dolog/learnerrdfbindings/`.

Algorithm. Following algorithm applies when system searches for relevant fragment of a learner performance:

- Retrieve all instances of the Identification concept for current user;
- Search instances of the Learner concept on systems referenced in each identification entry;
- If there are further systems referenced in the identification records at the remote systems, reapply this algorithm with the records;
- Retrieve all objects as instances of concepts needed for adaptation (e.g. learner performance);

To illustrate the algorithm, let us refer back to the Alice scenario. We use the Edutella [Nejdl *et al.*, 2002] to submit queries to the P2P network. The Edutella P2P infrastructure allows us to connect peers which provides metadata about resources described in RDF. Edutella also provides us with a powerful Datalog based query language, RDF-QEL. The query can be formulated in RDF format as well, and it can reference several schemas.

In the following we will use the QEL selection syntax where three parameters (subject, predicate, object) are used to retrieve instances of RDF classes. The syntax of such selection in QEL is s(subject, predicate, object). The selection will retrieve all the resources which contain assertions with the subject, predicate, and object. Any of those parameters can be used as variables.

As we assume a uniform data model suggested above, the query can be formulated in terms of the data model.

```
s(Alice, learner:identification, Ident),
    s(Alice, rdf:type, learner:Learner),
    s(Alice, learner:learner_id, LID),
    s(Ident, rdf:type,
      learner:Identification),
    s(Ident, learner:provider, PID),
    s(PID, rdf:type, learner:Provider)
    s(Ident, learner:ID, LPID).
s(Alice, learner:performance, Perf),
    s(Perf, rdf:type,
      learner:Performance),
    s(Perf,
      learner:learning_experience, LEX).
```

First, all identification records of Alice are retrieved together with local learning performance. The selection query for learner identifiers is constructed based on the Identification concept (the learner: prefix is an abbreviated namespace of the learner schema). The remote learner identification is maintained as a pair of provider and learner identifiers (PID, LPID) maintained in the provider and ID attributes. It is allowed to have one learner identifier valid for several providers. In that case, multiple pairs are retrieved. According to the Alice scenario, the program finds the identifications of the university provider and the CompuTraining provider.

The join selection for performance is constructed based on the Performance concept. The performance maintains a learning_experience attribute where an identifier of a resource taken during the study is stored.

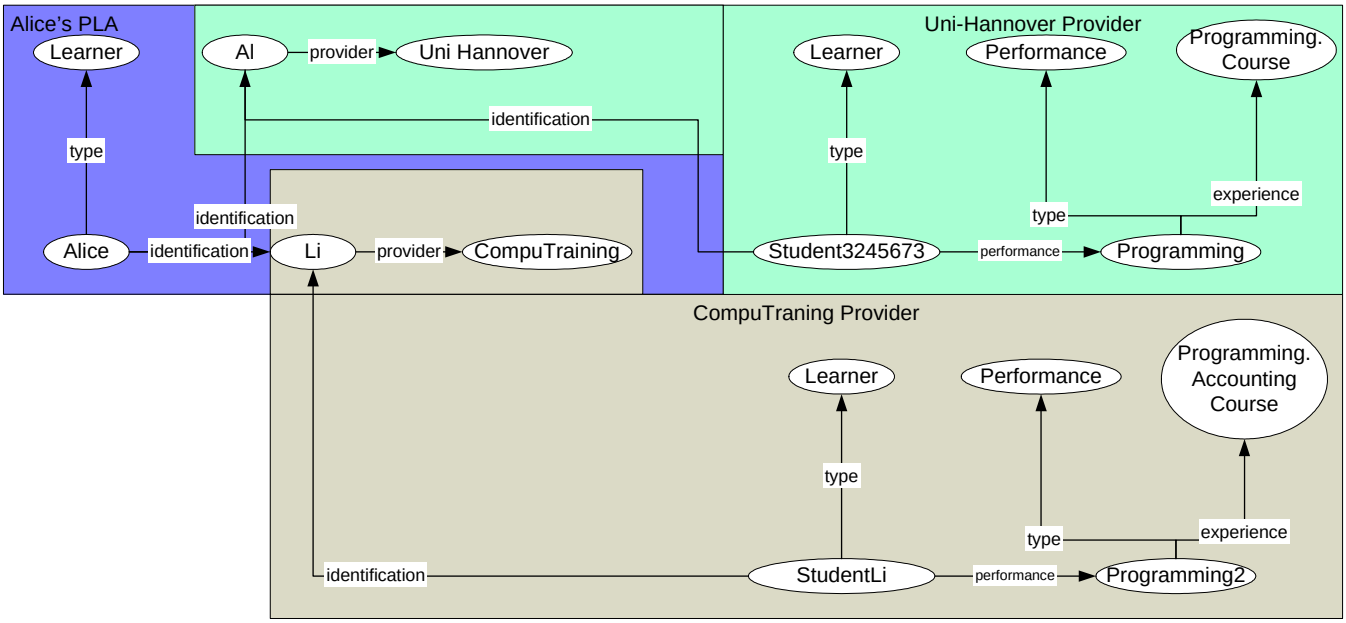

Figure 3: An excerpt of instances of Alice's performance at different systems under different identification

As the external providers can have similar identification records for third systems, a query has to be constructed for each tuple (PID, LPID) to find the identifications at the third systems. A query to retrieve also performance values from the external systems is constructed similarly to the previous example.

```
s(Lremote, learner:identification,
  RemoteLID),
    s(Lremote, rdf:type, learner:Learner),
    s(Lremote, learner:learner_id,
      LIDremote),
    s(RemoteLID, rdf:type,
      learner:Identification),
    s(RemoteLID, learner:provider, PID),
    s(RemoteLID, learner:ID,
      LPID),
s(Lremote, learner:identification,
 RemoteLIDI),
    s(RemoteLIDI, rdf:type,
      learner:Identification),
    s(RemoteLIDI, learner:provider,
      PIDExternal),
    s(RemoteLID, learner:IDExternal,
      LPID).
```

If there is a non empty result set of identification entries for the third systems, the query construction is reapplied until there are no more systems to contact.

Discussion. The approach to distributed learner modelling presented in this paper is currently under development in EU/IST Elena project. The exchange model for learner profiles described in this paper has been implemented for example in the PLA [Dolog *et al.*, 2004a] and is currently under development in the Personal Reader [Dolog *et al.*, 2004b]. The advantage of this approach is that it relies on standards for learner profiling which allows to construct uniform queries. The identification mechanism suggested here allows to use local learner identifiers and the mapping between them is performed according to the records which maintain learner identifiers at the neighboring providers and/or services. The records also provide us with routing information for queries, i.e. which providers to contact to retrieve additional information about learner.

There are some open issues which still have to be resolved. There is a very likely situation that the internal data model for learner profiles is different from the one based on standards. The providers have to support query rewriting functions to rewrite received queries into their internal data model. Another solution would be to provide mapping services between schemas employed as discussed for example in [Dolog *et al.*, 2004a; Simon *et al.*, 2004]. Another important problem is how to address different attribute value ontologies for example for concepts learned or competencies acquired. The ontology mapping has to be employed also in that case.

Another problem which is currently discussed is where to put the reasoning about the query construction. The queries for the algorithm proposed in this paper can be constructed at a mediator (e.g. the PLA). Another approach would be that each provider will be able to construct additional queries if there are external systems to be contacted according to the identification records. This would mean that each node in the network will construct and submit queries just to its neighbors.

In the case of inter-organization network, privacy and security issues has to be addressed to protect sensitive data. The identification mechanism has to be combined with distributed policies and credentials evaluation. Both, the identification records and learning related learner features, has to be

protected and disclosed just to trusted parties. An important question in this context is how to protect information which was already disclosed to a system which is asked by third external system to provide the information.

5 Related Work

Work on integration of distributed user model fragments which are needed for specific task was presented in [Niu *et al.*, 2003; Vassileva *et al.*, 2002]. Their work similarly as our is based on an idea that just particular fragments in specific combination are needed for different computation purposes. In our work we applied the standard vocabulary to reduce negotiation overhead needed when the heterogeneous fragments and schemas are employed.

The identification through pseudonymity was applied also in [Kobsa and Schreck, 2003]. The pseudonymity was treated as kind of protection mechanism. Here we apply different pseudonyms in learner identification on different systems to compute relevant fragments of learner profile.

As the learner data are sensitive, the trust and security issues have to be further investigated. We have already proposed extensions towards standard based vocabulary for privacy purposes in learner modelling. Those extrensions and work reported in several other projects like [Kobsa and Schreck, 2003; Bohrer and Holland, 2000; Heckmann, 2003; Maler *et al.*, 2003] should be further investigated in the context of our approach.

Trust negotiation is another interesting related work in the context of open environment. First steps towards trust negotiation in open p2p network was presented in [Nejdl *et al.*, 2003]. Guarded distributed logic programs are used to encode policies and enforce them during resource attribute exchange and negotiation. We are currently investigating whether this approach is suitable also for learner attribute exchange. Our metamodel for policies to protect learner information is flexible and allow us to use any language when there is appropriate interpreter available. Appropriate subclass of the policy class is then used to identify which iterpreter to run.

6 Conclusions and Further Work

This paper reports on recent work for the development of learner profile for the ELENA project. We showed that we can definitely benefit from learner profiles standards because they provide a vocabulary which was agreed in a broader context.

The RDF and RDFS allow us to use different schemas and query languages such as QEL. The QEL allows us to integrate reasoning capabilities over personal profile in P2P network. This is step towards P2P RDF based environment where personalization techniques can be implemented as services.

Many issues still have to be resolved. The technical infrastructure for this approach to personalization has to be investigated in more detail and mechanisms for provision, searching, and using such personalization services have to be introduced. Mapping or mediating between different schemas should be investigated as well when we want to provide communication between different peers. Different identification schemes have to be investigated more deeply to support better exchange of learner profile fragments between distributed nodes. Experiments with analyzed privacy technologies and dynamically switching between them have to be investigated to support flexibility in open environment also in the context of security.

Acknowledgements. We would like to thank Wolfgang Nejdl and Tomaž Klobučar for extensive discussions which helped to improve this work. We would also like to thank anonymous reviewers for comments which helped to improve this paper.

References

[Bohrer and Holland, 2000] Kathy Bohrer and Bobby Holland. Customer profile exchange (cpexchange) specification. http://www.cpexchange.org/, October 2000.

[Dolog and Nejdl, 2003] Peter Dolog and Wolfgang Nejdl. Challenges and benefits of the semantic web for user modelling. In *Proc. of AH2003 — Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems, WWW2003 Conference, Hypertext'03 Conference, User Modelling Conference 2003*, Budapest, Hungary, Nottinngham, UK, Johnstown, PA, USA, 2003.

[Dolog *et al.*, 2004a] Peter Dolog, Nicola Henze, Wolfgang Nejdl, and Michael Sintek. Personalization in distributed e-learning environments. In *Proc. of WWW2004 — The Thirteen International World Wide Web Conference*, New Yourk, May 2004. ACM Press.

[Dolog *et al.*, 2004b] Peter Dolog, Nicola Henze, Wolfgang Nejdl, and Michael Sintek. The personal reader: Personalizing and enriching learning resource using semantic web technologies. In Wolfgang Nejdl and Paul De Bra, editors, *Proc. of AH2004 — International Conference on Adaptive Hypermedia*, volume 3137 of *LNCS*, Einghoven, The Netherlands, August 2004. Springer.

[Heckmann, 2003] Dominik Heckmann. Integrating privacy aspects into ubiquitous computing: A basic user interface for personalization. In *Proceedings of the AIMS 2003, Artificial Intelligence in Mobile System*, 2003.

[Kobsa and Schreck, 2003] Alfred Kobsa and Jörg Schreck. Privacy enhancing technologies — approaches and development. *ACM Transactions on Internet Technology*, 3(2):149–183, May 2003.

[Maler *et al.*, 2003] Eve Maler, Prateek Mishra, and Robert Philpott. Bindings and profiles for the oasis security assertion markup language (saml) v1.1. OASIS Standard, September 2003.

[Nejdl *et al.*, 2002] W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmér, and T. Risch. EDUTELLA: a P2P Networking Infrastructure based on RDF. In *In Proc. of 11th World Wide Web Conference*, Hawaii, USA, May 2002.

[Nejdl *et al.*, 2003] Wolfgang Nejdl, Daniel Olmedilla, and Marianne Winslett. Peertrust: Automated trust negotiation

for peers on the semantic web. Technical Report, Learning Lab Lower Saxony, University of Hannover, November 2003.

[Niu *et al.*, 2003] Xiaolin Niu, Gordon McCalla, and Julita Vassileva. Purpose-based user modelling in a multi-agent portfolio management system. In Peter Brusilovsky, Albert T. Corbett, and Fiorella de Rosis, editors, *In Proc. of User Modeling 2003, 9th International Conference, UM 2003*, Johnstown, PA, June 2003. Springer, LNAI 2702.

[Simon *et al.*, 2004] Bernd Simon, Peter Dolog, Zoltán Miklós, Daniel Olmedilla, and Michael Sintek. Conceptualising smart spaces for learning. *Journal of Interactive Media in Education. Special Issue on Educational Semantic Web*, (9):221–239, May 2004.

[Vassileva *et al.*, 2002] Julita Vassileva, Gordon McCalla, and Jim Greer. Multi-agent multi-user modelling in I-Help. *User Modeling and User-Adapted Interaction*, 2002.