Mapping between Digital Identity Ontologies through SISM

Matthew Rowe

The OAK Group, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, UK
m.rowe@dcs.shef.ac.uk

Abstract. Various ontologies are available defining the semantics of digital identity information. Due to the rise in use of lowercase semantics, such ontologies are now used to add metadata to digital identity information within web pages. However concepts exist in these ontologies which are related and must be mapped together in order to enhance machine-readability of identity information on the web. This paper presents the Social identity Schema Mapping (SISM) vocabulary which contains a set of mappings between related concepts in distinct digital identity ontologies using OWL and SKOS mapping constructs.

Key words: Semantic Web, Social Web, SKOS, OWL, FOAF, SIOC, PIMO, NCO, Microformats

1 Introduction

The semantic web provides a web of machine-readable data. Ontologies form a vital component of the semantic web by providing conceptualisations of domains of knowledge which can then be used to provide a common understanding of some domain. A basic ontology contains a vocabulary of concepts and definitions of the relationships between those concepts. An agent reading a concept from an ontology can look up the concept and discover its properties and characteristics, therefore interpreting how it fits into that particular domain. Due to the great number of ontologies it is common for related concepts to be defined in separate ontologies, these concepts must be identified and mapped together.

Web technologies such as Microformats, eRDF and RDFa have allowed web developers to encode lowercase semantics within XHTML pages. It is common-place for different vocabularies to be used within such pages, particularly when describing identity information. It must be declaratively specified to the agent how an unknown vocabulary is related to the vocabulary set that it knows and understands. The need to specify mappings between digital identity ontologies is currently being investigated by the W3C Social Web Incubator Group[1]. Our contribution in this paper is the presentation of a mapping vocabulary between ontologies used to define digital identity information called the Social Identity

---

[1] http://www.w3.org/2005/Incubator/socialweb/

Schema Mapping (SISM). Each mapping provides a relation between two concepts in disparate ontologies where the relation defines the semantics of the mapping which can cover equivalence, associative and hierarchical relations.

This paper is structured as follows: Section 2 defines the problems that have motivated the creation of mappings between digital identity ontologies. Section 3 presents the Social Identity Schema Mapping (SISM), detailing the semantics of the mappings, explaining the design choices and providing example mappings. Section 4 discusses how SISM is currently being used in applications. Section 5 presents related work and section 6 discusses the conclusions we have drawn from SISM and plans for future work.

2   Problem Definition

2.1   Monitoring Personal Information Online

Metadata models can now be embedded within web pages using lowercase semantics: The Resource Description Framework in Attributes (RDFa) [8] and Microformats embeds ontology concepts and a set vocabulary of terms within the attributes of XHTML elements respectively. These metadata models can be gleaned from web pages through the use of Gleaning Resource Descriptions from Dialects of Language (GRDDL) [2] where transformations specified within a web page's header generate an RDF model from the page. The transformation specifies the ontologies to be used within the gleaned model, which may be unknown ontologies. This is a common problem when monitoring personal information on the web using automated means: Web pages are parsed in order to glean metadata models containing identity information, if the models use unknown ontologies then automated approaches are unable to interpret such models. Such approaches must be explicitly informed of the relationship between the parsed concepts and a known vocabulary of concepts.

2.2   Data Portability

The majority of web users now have profiles on several distributed Social Web platforms. Such platforms open up their data through an API allowing it to be reused and combined with data from other platforms. In most cases the response returned by the API is XML according to an XML schema used by the platform and due to each platform using a distinct schema, reusing data automatically becomes limited. If we consider that each Social Web platform that a given web user is a member of contains a distinct facet of his/her digital identity, then compiling these facets together would create a complete profile of the web user. This profile could then be reused by the web user when accessing additional Social Web platforms, allowing personal information access to be controlled.

3   Social Identity Schema Mapping (SISM)

SISM solves the problem of information heterogeneity by providing mappings between different identity ontologies. An agent is able to consult the mappings

and derive a relation between a parsed concept and a known concept where the relation describes the semantics of the mapping in a machine-readable format. SISM is available on the web[2] as RDF containing mappings defined as triples containing a source concept as the subject and the target concept as the object, the predicate contains the semantics of the mapping.

3.1 Mapping Identity Ontologies

SISM contains mappings between five ontologies which contain concepts used to define identity information: The Friend of a Friend ontology [3], the ontology for VCards [4], the XFN ontology [5], and the Nepomuk ontologies; Personal Information Model Ontology [6], and the Nepomuk Contact Ontology [7].

Mapping constructs from the Web Ontology Language (OWL) [7] and the Simple Knowledge Organisation System (SKOS) [6] are used within SISM. OWL is a well established and widely used ontology language, and SKOS is also a widely used language for describing knowledge organisation systems such as thesauri. The combination of both OWL and SKOS constructs cover the range of mappings needed in order to map concepts from the available identity ontologies. OWL semantics employ strong bindings of equivalence and subsumption. If we want to say that two concepts from distinct ontologies are similar or related in some way but are not equivalent then we cannot express this using OWL constructs. Conversely, SKOS semantics employ relaxed bindings of a relation. If we want to denote a mapping of equivalence between concepts then OWL constructs are better suited given that SKOS constructs do not cover such equivalence. We will now discuss the design of SISM and explain which constructs were used and how they were applied.

3.2 Collections of Mappings

Figure 1 shows mappings between the *foaf:Person* and *vcard:VCard* classes where each class is defined as an instance of *skos:Concept*. According to the SKOS primer [6] such concept definitions encompass all OWL class and property types (i.e. *owl:Class*, *owl:ObjectProperty* and *owl:DatatypeProperty*). SISM contains a concept scheme (instance of *skos:ConceptScheme*) defined by the hash URI *#sism*. Each mapped concept is related to this concept scheme through the *skos:inScheme* relation. This allows an agent to query the semantic web asking for all the schemes that a given concept belongs to. If we imagine that thousands of mapping vocabularies are published, each containing a SKOS concept scheme, then an agent is able to see the differing use that a given concept has been applied

---

to and enrich its vocabulary of interpretable concepts. Mapped concepts are grouped into collections to organise them into logical sets. For instance, a user or agent could ask SISM for all the mappings which describe the name or the email address of a person. Collections are expressed as instances of *skos:Collection*. As figure 1 shows, the collection of person concepts is identified as a resource using the hash URI *#person*. Concepts are placed within a collection using the *skos:member* relation to denote membership.
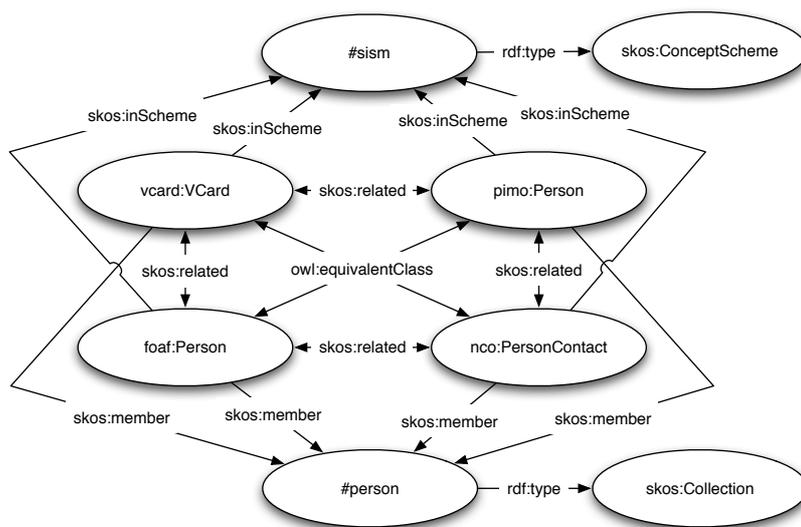

Fig. 1. Collection of mapped Person concepts

Figure 1 shows mappings between classes in different identity ontologies. We define the *foaf:Person* and *vcard:VCard* concepts as being related in some way but not equivalent: *vcard:VCard* defines an information resource containing contact information whereas *foaf:Person* defines an instance of a person. We therefore regard each concept to be related in some way, and express this using the *skos:related* relation. The relation between *foaf:Person* and *pimo:Person* is more straightforward as each ontology defines a resource as a person. Therefore we used the strict expressivity of OWL to relate the two concepts as being equivalent using the *owl:equivalentClass* relation.

Hierarchical relations between concepts in SISM are dealt with through the use of the SKOS relationship definitions *skos:broader* and *skos:narrower*. SISM contains a collection of mappings between URL concepts, the *foaf:homepage* property is defined as having a broader concept defined by the *vcard:url* property. Therefore using *skos:narrower*, the inverse of *skos:broader*, *vcard:url* can also be defined as having a narrower concept defined by the *foaf:homepage* property. We chose SKOS constructs to express generalisation and specialisa-

tion due to their relaxed semantics, OWL constructs also allow the definition of hierarchical relations but impose strict restrictions on subsumption. Equivalent property concepts in distinct schemas are defined as equivalent using the *owl:equivalentProperty* relation.

4   Applications

4.1   Transforming Metadata Models

To date the main application of SISM has been the transformation of metadata models. As we explained in section 2, one of the main problems that motivated the creation of SISM was the increase of lowercase semantics within XHTML markup and the disparate ontologies used. We have successfully applied SISM to normalise metadata models to our desired identity ontologies by generating inference rules which when applied to a given RDF model transforms the model. Transformed models can then be integrated and the internal data compared more easily given the same ontological concepts. This approach has been successfully deployed in an application that automatically monitors personal information on the web [4].

4.2   Interlinking Distributed Identity Fragments

The second application of SISM has been to map XML schemas used by Social Web sites to concepts within SISM. Through this additional mapping metadata models, defined as XML, returned from querying Social Web platforms are lifted to RDF. By lifting these responses to RDF representations we have been able to integrate information from such platforms [5] thereby investigating how digital identity fragments distributed across the Social Web can be combined into a single complete profile which the user has control over.

5   Related Work

The UMBEL project[8] has investigated and produced a lightweight ontology to provide a mapping layer between RDF models published on the web. This works by linking similar concepts in distinct ontologies to an upper concept, therefore providing mappings in an upper level. SISM differs by using collections for meditation rather then concepts.

Work by [9] presents an approach to data portability across Social Web platforms by defining identity information using FOAF. However, such platforms will not consume RDF using FOAF, instead exported data must be lifted to RDF and then lowered again into the required form. Making social data portable and therefore reusable across Social Web platforms has been discussed in [1] through the use of SIOC[9]. Social data such as shared content and discussions can be

---

[8] http://www.umbel.org/
[9] http://www.sioc-project.org

described using SIOC and when the web user exports his content from one site to another it is machine-readable and therefore reusable. This strategy has been deployed as a fully functional WordPress plugin[10].

6   Conclusions and Future Work

In this paper we have presented the Social Identity Schema Mapping (SISM), a currently deployed and very much in-use vocabulary of mappings between disparate digital identity ontologies. At the time of writing this paper SISM was the only mapping vocabulary for digital identity ontologies available on the web. The need for SISM is highlighted in the W3C Social Web Incubator Group's current work investigating the portability of social data and the overlap in conceptual elements available to define the semantics of such data. SISM is based on both SKOS and OWL constructs which capture the possible semantics of relations between concepts and is currently supporting two areas of work: transforming metadata models and interlinking identity fragments distributed across the Social Web.

Future work will include the addition of more ontologies into SISM to cover the maximum range of digital identity ontologies. We hope that SISM becomes the basis for interpreting relations between digital identity ontologies, therefore we are currently implementing a RESTful service which returns an RDF model from a web page according to required ontologies.

References

1. Bojars, U., Breslin, J.G., Peristeras, V., Tummarello, G,. and Decker, S.: Interlinking the Social Web with Semantics. Intelligent Systems, IEEE. Vol. 23, no. 3, pp. 29-40. (2008)
2. Connolly, D.: Gleaning Resource Descriptions from Dialects of Languages (GRDDL). W3C Recommendation. `http://www.w3.org/TR/grddl/` (2007)
3. Khare, R.: Microformats: The next (small) thing on the Semantic Web?Internet Computing, IEEE. Vol. 10, pp 68-75. (2006)
4. Rowe, M.: Applying Semantic Social Graphs to Disambiguate Identity References. In: Proc. 6th European Semantic Web Conference. Heraklion, Crete. LNCS, vol. 5554, pp 461-475 (2009)
5. Rowe, M.: Interlinking Distributed Social Graphs. In: Proc. Linked Data on the Web Workshop, WWW09. Madrid, Spain. (2009)
6. Isaac, A. and Summers, E.: SKOS Simple Knowledge Organization System Primer. W3C Recommendation. `http://www.w3.org/TR/skos-primer/` (2009)
7. Smith, M., Welty, C., and McGuinness, D.L.: OWL Web Ontology Guide. W3C Recommendation. `http://www.w3.org/TR/owl-guide/` (2004)
8. W3C Working Group.: RDFa Primer: Bridging the Human and Data Webs. (2008) `http://www.w3.org/TR/xhtml-rdfa-primer/`,
9. Yeung, C., Liccardi, I., Lu, K., Seneviratne, O., and Berners-Lee, T.: Decentralization: The Future of Online Social Networking. W3C Workshop on the Future of Social Networking Position Papers. (2009)

---

[10] http://wiki.sioc-project.org/w/SIOC_Import_Plugin