Publishing Semantic Personal Notes as Linked Data

Laura Drăgan
DERI, National University of
Ireland, Galway
laura.dragan@deri.org

Alexandre Passant
DERI, National University of
Ireland, Galway
alexandre.passant@deri.org

Tudor Groza
DERI, National University of
Ireland, Galway
tudor.groza@deri.org

Siegfried Handschuh
DERI, National University of
Ireland, Galway
siegfried.handschuh@deri.org

ABSTRACT

There is an obvious shift of focus towards the Web, as people spend more of their time and store more of their data online. However, the desktop still handles a large amount of the personal data. The Semantic Desktop brings semantics to the desktop, better interlinking and organization of the data, thus allowing better management. However, in spite of common representation formats, personal and online data is still difficult to interlink, notably because of the different vocabularies used to describe it as well as the lack of common identifiers between desktop applications and Web-based services. We describe here a solution for easily publishing and sharing of personal notes as Linked Data. We provide a two-phased publishing and interlinking process. It can be used to publish any kind of information from the desktop to the Web, enabling integration of small chunks of personal knowledge into the Web of Data and focusing on a user-driven approach of knowledge management.

1. INTRODUCTION

The Linked Data initiative and the Semantic Desktop are two relevant subdomains of the Semantic Web vision. The first focuses on globally interlinking resources on the Web, while the second enables information integration and interlinking on the desktop. Although the two share the representation models (RDF(S)/OWL), there is still a gap between data from the Web and the desktop, due to different sets of vocabularies that are used, and generally distinct identifiers.

We describe here an approach for integrating data from these two environments, specifically for publishing personal notes from the desktop (using Semantic Desktop technologies) to the Web (using the Linked Data principles). The main requirement is to publish data online without losing the personal context established on the desktop. Our approach consists of two main steps: (i) preparing the desktop data for sharing, and (ii) publishing it online. In addition, it requires two prerequisite steps, provided by existing applications: (i) the note-taking process and annotation of the note (adding the context), and (ii) the identification of Web URIs which represent the same real-world thing as the desktop resources that belong to the context of a note.

Our contributions include: (i) mappings between the relatively small number of desktop vocabularies and the most popular Web vocabularies. The mappings are used in the transformation of the desktop data, represented with the desktop ontologies, to data represented with the Web vocabularies, ready to be published online; (ii) a process for the publication of desktop information on the Web using the Linked Data principles, while at the same time protect the sensitive private data from being shared unwillingly; (iii) a system implementation that allows sharing of semantic personal notes as semantic blog posts, interlinked with existing information within the Linked Data cloud.

2. USE CASE - SEMANTIC BLOGGING

Two important characteristics of blog posts are: (i) their topics are of interest to the author and thus are very likely to have references to things present on the desktop; (ii) they belong to a context consisting of the references made in their content. Writing a blog post in a desktop application can offer several benefits, if the application is a semantic one, on a Semantic Desktop, where desktop resources are interconnected [1]. Semantic note-taking tools like SemNotes[1] automatically generate relations between the notes and the desktop things mentioned in their content. Such annotations give context to the note and should be preserved when the note is published as a blog post on the Web, since it enables serendipitous browsing and information discovery, as they contain relevant additional links to other entities.

Currently, personal notes, even the ones semantically enriched using Semantic Desktop applications, must be published as blog posts by being manually copied into a blogging tool. In this way, any additional semantic information available on the desktop becomes lost or, if copied, leads to broken references as they point to the local resources which are not accessible outside of the desktop. The note-taking to publishing process is sometimes shortcut by using the drafting functionality that some systems like WordPress or Blogger offer, so that users can directly take the notes in

---

[1] http://smile.deri.ie/projects/semn

the blogging tool, usually online, thus replacing the desktop note-taking application. Using online tools deprives the user from having the personal context automatically added to the blog post, since desktop information cannot be easily integrated in Web-based interfaces.

3. SYSTEM OVERVIEW

We propose an approach that enables the publishing and sharing of personal notes by extending the functionality provided by SemNotes. The process has two steps: (i) transformation and (ii) publication. In the first step, the note is transformed locally for publication, and private local data is replaced with public server references. In the second step, the transformed note is published online on a dedicated "SemNotes server", where the resources referenced and the tags assigned, are shared between the notes of all users. As we mentioned above, there are also two prerequisite steps: (i) the note-taking process and annotation of the note, which is the usual note-taking approach, and (ii) the identification of Web aliases for the desktop resources related to a note, where URIs are mined from the Web for locally defined resources, such as people, events or projects. The annotation is done semi-automatically and is an existing feature in SemNotes. The second prerequisite step consists in finding Web resource for each of the desktop entity linked to the note that is about to be published. This step is executed by a desktop service that relies on the Semantic Web index Sindice to retrieve results.

The publishing system is divided between its local part and its remote part. The local part handles local *private* data, while the remote one handles online *public* data. The separation between them extends over three layers: ontology, data and application.

On the *ontology* level, the NEPOMUK desktop ontologies are used locally while popular Web vocabularies like SIOC are used on the server-side. These ontologies are used to describe the *data* exchanged between the applications. Personal desktop data is stored in the local repository, which exists on any NEPOMUK Semantic Desktop, while Web data is distributed in the Linked Data cloud. Finally, on the *application* level, the local component is an extension to SemNotes that provides publishing functionality for notes, and the remote component is a server that hosts and publishes online the notes.

The first step of the process is executed on the local side, by an extension of the SemNotes application. It consists of replacing the links to the local resources mentioned in the note with their Web aliases, and enriching the content of the note with RDFa. Then, the publication step is done by the server, which receives information from the desktop and publishes the note according to the Linked Data publishing principles. On the server, the notes, linked resources and tags have dereferenceable URIs. The resources and tags are shared among notes and users, thus providing object centered sociality. The dataset is also linking to external resources that are found to be `sameAs` the local ones.

4. RELATED WORK

Semantic blogging was introduced by Cayzer and Shabajee in [3]. Karger and Quan describe semantic blogging in the context of the Semantic Web browser Haystack [4]. The existing systems for semantic blogging fall into two categories: (i) desktop applications that involve publishing the actual local resource information together with the blog post, or (ii) online applications that do not have access to desktop data relevant to the user. The first category, represented by tools like SemiBlog [5] or SemBlog [6], gives the user better access to the relevant data from the desktop. However, both tools require that the resources that contain private information are published together with the blog post, which might lead to privacy issues. They are used for exchange of personal information in the blog posts, which differs from our approach of using already published web data as to protect the privacy of the personal information. The process described implies manually adding the metadata, while our approach relies on automatic export. Online services like BlogAccord [2] for music information or Zemanta² blogging assistant, belong to the second category. They enhance the blogging experience by providing access to various online resources to create the context of a blog post, but not to the personal context of the user.

5. CONCLUSION

We presented an approach for publishing personal notes as Linked Data on the Web. The aim of our work was to provide a way for publishing and sharing complete information by preserving the personal context of the notes without compromising privacy. Our solution makes a step towards bridging the gap between local semantic data and Linked Data. We defined a publishing process that comprises two steps: (i) preparation – the note is transformed into a SIOC-based Web representation; and (ii) publication / sharing – the note is published online following the Linked Data principles.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] A. Bernardi, S. Decker, L. van Elst, G. Grimnes, T. Groza, S. H. M. Jazayeri, C. Mesnage, K. Moeller, G. Reif, and M. Sintek. *The Social Semantic Desktop: A New Paradigm Towards Deploying the Semantic Web on the Desktop*. 2008.

[2] S. Cayzer. What next for semantic blogging? In *Proc. of the SEMANTICS 2006 conference*, pages 71–81, Vienna, Austria, November 2006.

[3] S. Cayzer and P. Shabajee. Semantic blogging and bibliography management. In *BlogTalk Proc.*, 2003.

[4] D. R. Karger and D. Quan. What would it mean to blog on the semantic web? In *Proc. of the 3rd ISWC, Hiroshima, Japan*, pages 214–228. Springer, 2004.

[5] K. Möller and S. Decker. Harvesting desktop data for semantic blogging. In *Proc. of Semantic Desktop Workshop, ISWC, Galway, Ireland*, 2005.

[6] H. Takeda and I. Ohmukai. Semblog project. In *Activities on Semantic Web Technologies in Japan, A WWW2005 Workshop*, 2005.

---

²`http://www.zemanta.com`