From Personal Notes to Linked Social Media *

Laura Drăgan and Alexandre Passant and Tudor Groza and Siegfried Handschuh
Digital Enterprise Research Institute
National University of Ireland, Galway
IDA Business Park, Lower Dangan, Galway, Ireland
firstname.lastname@deri.org

Abstract
Semantic technologies are available, and gain popularity on the Web as well as on the desktop, but both (desktop and Web) act as large data silos, making personal and online data difficult to interlink. We propose a system that enables easy publishing of personal notes as linked social media content, while at the same time semantically enriching the desktop resources with information retrieved from the Linked Data cloud. The transformation, publication and linking process is integrated with the familiar desktop applications and online blogging platforms, providing a better usability experience.

Introduction
During the last years, two different paradigms emerged, bringing semantics to the Web and to the desktop. On one side, the Linked Data (LD) initiative and the Linking Open Data (LOD) project had an enormous success in linking isolated datasets on the Open Web, by means of Semantic Web (SW) technologies. On the other side, the Semantic Desktop (SD) enabled interlinking isolated application data existing on the desktop, also using SW technologies.

One of the new challenges is how to interconnect personal semantic data created on the desktop with data from the SW. The problem has two aspects: (i) publishing and integrating personal information into the LOD cloud, and (ii) enriching desktop data with information from the Web, (iii) while assuring the privacy of the data. The tasks are difficult because the data is not uniformly described, although both environments use SW technologies.

In this paper we show how semantic personal notes from the desktop can easily be published as LD and transformed into social media content and automatically connected with the LOD cloud, without losing meaning or context. The result is not only the transformation and publishing process, but also a personal linked dataset, well integrated within the larger cloud of Linked Data.

Motivating Use Case
Two relevant characteristics of blog posts are: (i) they belong to a context made up of the references existing in their content, such as people, places, projects, events or other blog posts; (ii) their topics are of interest to the author and thus are very likely to have references to elements present on her desktop (e.g. persons mentioned in text are in the local address book application).

The SD maintains all local resources interconnected, while the applications encourage linking and creation of relations between new and existing resources. The semantic note-taking tool SemNotes\(^1\) automatically creates the links between the personal notes taken and the desktop resources mentioned in their content. Personal notes are often the starting point of blog posts, but when published online, all the relations that enrich the notes are lost because they refer to resources available only locally. However, the same resources often exist within LD cloud, therefore the links to them can be replaced in the publishing process. The resulting blog posts retain the meaning and context of the initial notes, without revealing personal data from the desktop.

System Design
We propose a system that enables semantic sharing of personal notes by extending the functionality provided by the semantic note-taking tool SemNotes. The notes are automatically annotated with relations to the desktop resources they mention in the text. The transition of the semantic notes from the desktop to the Web is done in three steps: (i) find equivalent online references for the local resources linked to the note, and enrich the desktop information with those references, (ii) transform the note by replacing the local resources with the Web ones, and (iii) publish the note online according to the LD principles (Berners-Lee 2006).

From the point of view of the architecture as well as of the data, our system is divided into local and remote. The local part handles local (private) data, while the remote module handles online (public) data.

As shown in Fig. 1, the local module collaborates with existing components of the Semantic Desktop: the central storage of semantic data as well as the note-taking application used (which we extend with a plugin). It also interacts with

\(^1\)The work presented in this paper was supported by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2). Copyright © 2010, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

\(^1\)http://smile.deri.ie/projects/semn
a series of proxies that directly access Linked Data repositories. On the remote side, the interaction with the LD cloud is done via services like “Ping the Semantic Web”.

The distinction between the local and remote parts of our system is reflected also in the vocabularies used. The NEPO-MUK Semantic Desktop (Bernardi et al. 2008) defines and uses a set of ontologies2, well-suited to represent desktop information. However these ontologies are not used, so far, in Web applications. On the Web, and especially in the Linked Data realm, popular vocabularies include the following, as described in (Bizer, Cyganiak, and Heath 2007): (1) FOAF, (2) SIOC, (3) DOAP, (4) GeoNames, (5) the Music Ontology, and (6) models such as Dublin Core and SKOS for general metadata. We consequently defined a set of mappings between both sets of ontologies, to enable interoperability.

The local component matches private local resources from the desktop to public Web resources and uses the information identified as referring to the same object to enrich the local data. This step is executed locally, in order to avoid possible privacy issues.

The remote part publishes the transformed personal notes as Linked Data according to the publishing principles. We use popular ontologies like the aforementioned ones for the data we publish online. In addition, we defined a scheme for naming the various objects published from the desktop on the Web:

http://vmgal17.nuig.ie/notes/note/id
http://vmgal17.nuig.ie/notes/resource/id
http://vmgal17.nuig.ie/notes/tag/label

The notes, which are the primary objects, have an unique id generated by the server. For the publishing process to be consistent, we generate server URIs for the resources referenced by the notes. Each resource URI will have internally a list of Web URIs that were found to match it. Because tagging is an important feature of the social media, and is available both on the SD (and therefore for the notes) and on the SW, we consider tags as a special kind of resources. The resources and the tags are shared on the server, increasing the interlinking of the data. Following the LD principles, all server URIs are dereferenceable and return information about themselves, in RDFa.

Related Work

In (Bojars et al. 2008) SIOC was used to enable interoperability between social data on the Web, but it can also be adapted to more specialized communities such as the Health Care and Life Sciences or Enterprise 2.0 environments.

The SemBlog project (Takeda and Ohmukai 2005) allows users to add data from personal ontologies to their blogs. (Karger and Quan 2004) describes semantic blogging in the context of the Semantic Web browser Haystack. The closest to our goals is SemiBlog (Möller and Decker). SemiBlog is 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, although in user-friendly way, while our approach relies on automatic export. The BlogAccord system (Cayzer 2006) enhances the blogging experience by adding information from metadata available on the Web from different services to existing posts.

Conclusion and Future Work

In this paper we presented a method for bridging the gap between local semantic data and Linked Data. The system has two distinct major parts, one on the desktop and another one online. Each has its own specific roles and functionalities. The system enables easy sharing of linked personal information while maintaining the existent relations between the resources. It protects the privacy of the data by using online resources that represent the same entities as the local ones, but which are already available in the Linked Data cloud.

References


---

2http://www.semanticdesktop.org/ontologies/
3http://rdfs.org/sioc/nepomuk