

Using SKOS to Integrate Social Networking Sites with Scholarly Information Portals

Arnim Bleier, Benjamin Zopilko, Mark Thamm and Peter Mutschke

GESIS-Leibniz Institute for the Social Sciences
{arnim.bleier, benjamin.zopilko, mark.thamm,
peter.mutschke}@gesis.org

Abstract: Web 2.0 platforms have become a ubiquitous way of information exchange, but are seldom integrated with the Web of Data. To overcome this situation we propose the usage of SKOS thesauri acting as back-of-the-book index providing domain-specific axes transcending applications. We illustrate this concept with a use-case in the social sciences domain but applications in other domains are possible.

Keywords: User-Generated Content, Digital Libraries, Semantic Web, SKOS, Social Sciences

1 Introduction

The Social Web represents the vision of user-friendly online platforms fostering the generation of content in a collaborative manner. In the case of emerging scholarly online platforms it is the hope that innovative patterns of knowledge creation and dissemination enhance collective intelligence by overcoming the role-asymmetry between producer and consumer known from the Web 1.0. Members of such communities can form virtual networks around topics of their academic interest propelling the exchange of ideas along the social graph. However promising this might be, the concept breaks with the borders of the application and leaves a gap left to fill for enabling linked and open Web 3.0 information systems.

In the remainder we suggest domain specific LOD traversal axes crosscutting application boundaries. We begin with the problem statement. Next we sketch out the process of linking heterogeneous academic data sets via SKOS thesauri and later extend this idea to unstructured user-generated content found on Social Networking Sites (SNS). We conclude with having a look into challenges still faced as well the current status of a prototype.

2 Problem Statement

Leaving the distinction between content provider and user with the advent of the so called Web 2.0 was the empowerment leading to a flood of content with only light-

weight explicit semantics such as mentions of other users or tagging. But as Schmidt rightfully argues [1] this new “producer” role suffers another distinction, this time towards the ontology manager and in the case of scholarly Web 3.0 platforms prohibiting a rich semantic encoding of new ideas in the first place.

The W3C Incubator report on a “Standards-based, Open and Privacy-aware Social Web”[2] provided valuable visions and bottom up projects (e.g. GNU social or Diaspora) have targeted cross-application interoperability, but few of these initiatives focused on the empowerment of users when it comes to the semantic dimension of their content to provide a linkage to rest of the Web of Data. A related situation is faced on the side of heterogeneous LOD services; this time however highly specialized application specific vocabularies prevent a coherent picture to emerge and make it difficult to traverse (scientific) data along domain concepts.

We argue that both cases are related in requiring a domain-specific representation that is precise enough to capture existing concepts, but also leaves flexibility to express new ideas.

3 Integrating Heterogeneous Data on the Web of Data

In this section we will have a closer look at the process of building up LOD services at GESIS combining fine-grained vocabularies for individual data sets and more coarse-grained representations for the thematic traversal of the social sciences domain. As the leading German social sciences infrastructure facility GESIS publishes large amounts of scientific information in form of library references, survey studies and corresponding statistical data sets on several sites (i.e. Sowiport¹, SOFIS² or ZACAT³) addressing different use cases and user categories. The development of such targeted applications scenarios yielded services enjoying high usage. However, leaving many information sources unconnected proved disadvantageous for growing beyond the originally foreseen use cases. Addressing these shortcomings with a complete rebuild of applications was not a choice, but an integrative approach was needed. The Simple Knowledge Organization System (SKOS) proved as a useful choice, allowing classic knowledge representations to be encoded, in terms of a high-level thesaurus, for the Web of Data. With the RDF representation of the Thesaurus for the Social Sciences [3] (TheSoz) a formal multilingual representation of the social sciences domain has been developed. This TheSoz⁴ acts as a back-of-the-book index for the social sciences and glues together data items belonging to various application domains; now we are looking into ways extending this concept to third-party applications such as academic Social Networking Sites.

¹ <http://www.gesis.org/sowiport/>

² <http://www.gesis.org/en/services/research/sofis-social-science-research-information-system/>

³ <http://zacat.gesis.org/webview/>

⁴ <http://lod.gesis.org>

4 Connecting Social Networking Sites

As we have addressed GESIS is providing different kinds of data sets and the Thesaurus for the Social Sciences provides the glue making it possible to traverse them along domain-specific axes. A similar situation is faced in case of integrating Social Networking Sites. If one agrees that applications centered on user-generated content should be aligned with the Web of Data a two-way mechanism would be needed to support ingoing as well as outgoing links to and from further LOD resources. While the subject of supporting application/rdf+xml request types on a partnering SNS is an open issue, progress on supporting outgoing links and requests to further LOD resources has been made. Users of our prototype can either manually select TheSoz concepts tags they think are suitable to their contribution(s) or use an automatic suggestion service recommending appropriate thesaurus concepts. The usage of an automatic suggestion service proves in particular useful since it requires only little user knowledge of the vocabulary itself and makes adoption of the service more likely. While these “TheSoz tags” can act just as traditional tags with a human readable label integrating seamless into the expected user experience, they are in fact a smart resource. Since the thesaurus is multilingual, literal forms of labels provide translations and semantic relations with other thesaurus concepts provide refinement and inclusion [4] in the tag-space. Moreover the semantic machine- and human-readable meaning of these tags does not end at artificial application boundaries but provides connections to other applications for the traversal of information along axes of user interests.

5 Current Status and Challenges

The discussed concept has still a long way to go. A particular challenge to get initial user involvement is the optimization of the multi-label classifier used in the TheSoz concept recommendation and consequently we are considering ways to integrate multi-modal data (e.g. mentions or the structure of discussion threads) into the feature space of the classifier. Most importantly to us, however, will be the user feedback to our prototype on the iversity⁵ platform launching this fall.

References

1. Schmidt, J. and Pellegrini, T.: Das Social Semantic Web aus kommunikationssoziologischer Perspektive. In: Social Semantic Web, pp. 453—468. Springer (2009)
2. Harry, H., Tuffield, M.: A Standards-based, Open and Privacy-aware Social Web: W3C Incubator Group Report. W3C (2010)
3. Zapilko, B., Sure-Vetter, Y.: Converting the TheSoz to SKOS. GESIS Report (2009)
4. Miles, A. and Bechhofer, S.: SKOS simple knowledge organization system reference. W3C (2008)

⁵ <http://www.iversity.org>