ORES-2010
Ontology Repositories and Editors for the Semantic Web

Proceedings of the 1st Workshop on Ontology Repositories and Editors for the Semantic Web

Hersonissos, Crete, Greece, May 31st, 2010.

Edited by

Mathieu d'Aquin, The Open University, UK
Alexander García Castro, Universität Bremen, Germany
Christoph Lange, Jacobs University Bremen, Germany
Kim Viljanen, Aalto University, Helsinki, Finland

Collaborative Metadata Editor Integrated with Ontology Services and Faceted Portals

Jussi Kurki and Eero Hyvönen

Semantic Computing Research Group (SeCo)
Aalto University, School of Science and Technology, and University of Helsinki
http://www.seco.tkk.fi/, firstname.lastname@tkk.fi

Abstract. This paper presents a generic RDF metadata editor SAHA 3 for collaborative content creation and instant semantic content publishing on the Semantic Web. SAHA 3 is a combination of a user-friendly interface and rich editing tools, that are able to utilize external ONKI ontology repositories as services. The system is integrated with a faceted portal engine HAKO, by which the metadata can be published instantly as a semantic faceted portal. Using the SAHA-ONKI-HAKO integrated system, a semantic portal can be created very easily by end-users by defining metadata schemas and related vocabularies, by annotating content, and by interactively configuring the user interfaces of the editor and the search engine. The system is in use in several semantic web applications and scales up to hundreds of thousands content objects.

1 Introduction

The basic process of implementing a faceted (semantic) portal [1–3] includes the following major steps: 1) Formulate vocabularies/ontologies/facets for representing domain concepts. 2) Design metadata schemas for representing content using (1). 3) Annotate content using (1) and (2), typically by a group of distributed peers in a Web 2.0 fashion. 4) Select the facets (1) and create the portal with semantic search and browsing facilities. This paper presents the tool SAHA 3[1] for the latter two phases. The idea is that given a set of domain vocabularies/ontologies/facets and a metadata schema for annotation (phases 1 and 2), a web-based annotation facility for distributed semantic content creation can be created instantly without programming skills. In a similar way, a faceted portal is automatically and instantly created online after annotating the content by just selecting facets for searching. Again, programming is not needed but only configuring the system using a web-based interactive interface.

We first describe the main features of the SAHA 3 metadata editor for the content creation phase, and then the integrated faceted portal engine HAKO[2] for content publishing, followed by notes about implementation and scalability. Finally, contributions of the work w.r.t related systems are discussed, and some application use cases are listed.

---

[1] http://www.seco.tkk.fi/services/saha/
[2] http://www.seco.tkk.fi/tools/hako/

2 SAHA 3 Metadata Editor Features

The original requirements for the web-based annotation editor SAHA [4] are simplicity (hiding technical concepts related to markup languages and ontologies from its user), adaptivity (to different metadata models), quality (helping and guiding the annotator to good and correct annotations), collaboration (supporting distributed simultaneous annotation at different locations), and portability (using the system on the web without installing any special software). SAHA 3 is a completely re-written version of SAHA with the following main new features: First, there is more support for general RDF editing, e.g. for inline editing of nested metadata, and for extending internal vocabularies. Second, the system is scalable to large datasets up to hundreds of thousands of objects. Third, SAHA 3 incorporates a simple publishing platform for building end-user search portals with full-text and multi-faceted search. The whole pipeline[3] from metadata editing to the end-user portal application is accessible and configurable through a web-based interface.

During annotation, references to external ontologies are handled using the ONKI web service interface [5, 6]. SAHA 3 utilizes autocompletion [7, 8] as a key component to find references. When the user tries to find a concept, SAHA 3 uses at the same time web services to fetch concepts from connected external ONKI ontology repositories, and a local index to find locally defined concepts. Results are shown in one autocompletion result list regardless of origin. The same query can cover several ontology repositories at the same time. References to resources within the project at hand and external to it (in an external ontology repository) are transparent to the user.
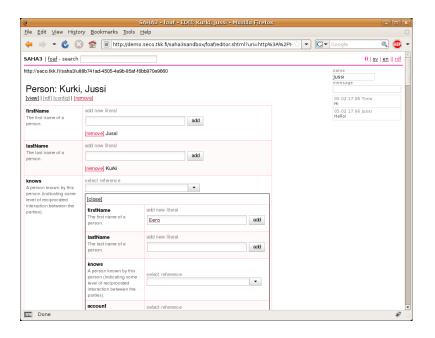
The inline editor is new feature in SAHA 3 that has been found very handy by end-users. The idea is simple: a resource referenced through an object property can be edited inline at the right location in a small version of the editor inside the existing editor. In this way, several levels of editors can be opened recursively within each other, and the RDF network can be edited without moving from one resource window to another. For example, in Figure 1 the FOAF profile of the first author of this paper is edited, and the value of the property "knows" is opened inline as a similar profile editor for the second author of this paper. Although the user interface easily becomes cluttered after a few levels, the inline editor is a handy way to add and edit nested metadata, as in this example shows.

SAHA 3 supports collaborative simultaneous editing. Resources that are being edited by one user are locked from other users. A chat facility has been implemented in the editor to facilitate instant discussions between peer editors (cf. the upper right corner in Fig. 1).

3 Faceted Search Engine HAKO

A SAHA 3 project can be published through the search interface of the HAKO portal engine. HAKO supports both free-text and faceted search [1–3]. For exam-

---

[3] SAHA Sandbox: http://demo.seco.tkk.fi/saha3sandbox/saha3/main.shtml

Fig. 1. Editing FOAF profiles in the SAHA 3 editor. New value for "knows" is created inline as a new instance of a person.

ple, Fig. 2 depicts a HAKO application using the RDF store of the Kirjasampo-system[4] containing tens of thousands of instances of literary work of different kinds. The user has entered the keyword "tolstoi". Free-text search is done as a prefix search by default. On the left, one can see the hit list of 70 books and plays related to Tolstoi distributed over the facet categories, from where the user can refine the search either by type or by theme. The facets are configurable—the administrator can select any object property to be a facet top category.

Apart from the facet administration, HAKO interface is actually only a front-end to the SAHA 3 model. The configurable, shared data model (including the indices), and the dataset used by SAHA 3 are the same as those used by HAKO. This means that all modifications made with the SAHA 3 editor are reflected on the HAKO interface instantly.

The system is implemented in Java on top of Spring[5] framework. The data model is based on TDB[6] RDF database. Full-text search is backed by Lucene[7]. The editor interface is built using DWR[8] and Dojo[9] AJAX-components.

---

[4] http://kirjasampo.fi/

[5] http://www.springsource.com/

[6] http://openjena.org/TDB/

[7] http://lucene.apache.org/

[8] http://directwebremoting.org/
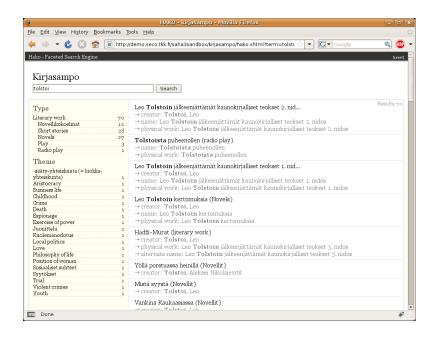
[9] http://www.dojotoolkit.org/

Fig. 2. SAHA3/HAKO multifaceted search interface on "Kirjasampo"-project.

4 Discussion

SAHA 3 makes use of external distributed ONKI ontology repository services, which is very handy when utilizing external third party ontologies and large vocabularies. From the interface design viewpoint, inline editing has been found a very useful feature. A major novelty of the system is the integration of the editor with a portal engine in real time. The idea of instant creation of a faceted search engine bears resemblance with SIMILE Exhibit[10]. However, in our case the system is integrated with a metadata editor, ontology services, and Semantic Web data models, and the search technology is based on Lucene (and not JavaScript) scaling up to very large datasets. SAHA 3 has been used in projects containing over 100,000 instances, and it has not shown signs of slowing down, or requiring large amounts of memory. The Tomcat server[11] used runs well below 500MB with large projects.

SAHA 3 is in production use in the HealthFinland[12] system and the Kirjasampo project, where some 50 librarians have been annotating metadata about tens of thousands of novels, short stories, authors, and other objects related to Finnish literature for a semantic portal. The editor has been used is many ways

---

[10] http://www.simile-widgets.org/exhibit/

[11] http://tomcat.apache.org/

[12] http://www.seco.tkk.fi/applications/tervesuomi/

in CultureSampo[13], e.g. for creating the semantic narrative descriptions of the Finnish History Ontology[14] and the Semantic Kalevala epic[15]. HAKO has been used in several industrial application demonstrations for searching documents.

References

1. Pollitt, A.S.: The key role of classification and indexing in view-based searching. Technical report, University of Huddersfield, UK (1998) http://www.ifla.org/IV/ifla63/63polst.pdf.
2. Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., Lee, K.P.: Finding the flow in web site search. CACM **45**(9) (2002) 42–49
3. Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S.: MuseumFinland—Finnish museums on the semantic web. Journal of Web Semantics **3**(2) (2005)
4. Valkeapää, O., Alm, O., Hyvönen, E.: A framework for ontology-based adaptable content creation on the semantic web. Journal of Universal Computer Science **13**(12) (2007)
5. Viljanen, K., Tuominen, J., Hyvönen, E.: Ontology libraries for production use: The Finnish ontology library service ONKI. In: Proceedings of the ESWC 2009, Heraklion, Greece, Springer–Verlag (2009)
6. Tuominen, J., Frosterus, M., Viljanen, K., Hyvönen, E.: ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. In: Proceedings of the ESWC 2009, Heraklion, Greece, Springer–Verlag (2009)
7. Hyvönen, E., Mäkelä, E.: Semantic autocompletion. In: Proceedings of the First Asia Semantic Web Conference (ASWC 2006), Beijing, Springer–Verlag (2006)
8. Hildebrand, M., van Ossenbruggen, J., Amin, A., Aroyo, L., Wielemaker, J., Hardman, L.: The design space of a configurable autocompletion component. Technical Report INS-E0708, Centrum voor Wiskunde en Informatica, Amsterdam (2007)

---

[13] http://www.seco.tkk.fi/applications/kulttuurisampo/

[14] http://www.seco.tkk.fi/ontologies/histo/

[15] http://www.seco.tkk.fi/applications/kulttuurisampo/kalevala/

[16] http://www.seco.tkk.fi/projects/finnonto/