

A Browser-based Semantic Annotation Tool for Distributed Content Creation

Onni Valkeapää and Eero Hyvönen

Semantic Computing Research Group (SeCo)

Helsinki University of Technology (TKK) and University of Helsinki

Laboratory of Media Technology

P.O. Box 5500, Otaniementie 17, FI-02015 TKK

<http://www.seco.tkk.fi>, firstname.lastname@tkk.fi

Abstract

We present a prototype of a semantic browser-based annotation tool SAHA, that eases the process of creating ontological descriptions of documents e.g. for semantic portals. SAHA supports collaborative creation of metadata by centrally storing the annotations, which can be viewed and edited by different annotators. Concepts defined in external ontologies can be used in annotations by connecting SAHA to ontology servers. The tool is being tested in practical semantic portal projects.

1 Introduction

Provision of rich ontology-based metadata is one of the major challenges in developing the Semantic Web. Various annotation systems have been developed to face this challenge [Uren *et al.*, 2006]. There is, however, a lack of systems that 1) can be easily used by an annotator unfamiliar with technical side of the Semantic Web, and that 2) are able to support the distributed creation of semantic metadata based on complex metadata annotation schemas (ontologies). This paper presents a tool, SAHA, aiming to satisfy these needs. SAHA is browser-based in order to support wide and distributed usage. It has simple user interface that hides complexity of ontologies from the annotator, and adapts automatically to different metadata schemas, and can utilize external ontology services for sharing URIs.

2 Annotations in SAHA

Annotations created with SAHA are based on a user-defined metadata annotation schema, which is defined in OWL. The purpose of a schema, such as Dublin Core, is to define the structure of annotations and to help the annotator to describe resources in a consistent way. Annotations are instantiated classes and properties of annotation schema, which are linked to the document being described. The linking is essential, because 1) an annotation is separate from the document it describes and 2) the way linking is done affects the meaning of relations between the annotation and the document. There are two ways to express a connection between a document and an annotation in SAHA. The first one makes the assertion that the document is an instance of one specific

class defined in the schema. It can be thus used effectively to classify documents. The second method binds an annotation to the document by using a named property. This idea is similar, e.g., to the usage the property “annotates” in Annotea¹.

3 Meta-schema

In addition to describing annotation’s structure in an annotation schema, we also need a way to define *how* the schema is actually used during the annotation process. For example, we may want to define an order for the properties of a class, in which they are shown to an annotator, or state that some property-values should appear as values of the `rdfs:label` property of an instance. Although such rules could be expressed in the annotation schema itself, it is often useful to separate the schema-design from its use [Handschuh and Staab, 2002]. In SAHA, this is done using the *meta-schema*, which is an ontology that describes how a certain annotation schema is used in a particular annotation-scenario. The meta-schema enables the use of same annotation schema in various scenarios. It enhances the interoperability of annotations as it decreases the need to develop distinct annotation schemas for different annotation-cases.

4 Architecture

SAHA is a web application implemented using the Apache Cocoon² and Jena³ frameworks. It is designed as a web application in order to impose as little requirements as possible to the end user’s computational environment. To use SAHA, all an annotator needs is an appropriate web browser with an Internet connection. SAHA uses extensively techniques such as *Javascript* and *Ajax* in order to provide annotator with simple and versatile user interface.

SAHA’s architecture is depicted in figure 1. It consists of 1) the annotator’s web browser, which is used to interact with the system, 2) a Jena ontology model (SAHA Model), which manages the RDF/OWL-data, 3) a PostgreSQL-database (SAHA Store), which is used to store the annota-

¹ <http://www.w3.org/2001/Annotea/>

² <http://cocoon.apache.org/>

³ <http://jena.sourceforge.net/>

tions, and 4) Cocoon's Control Flow, which binds together all components.

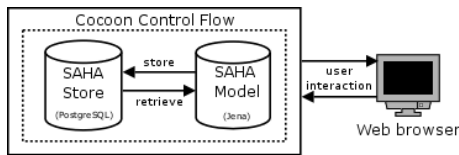


Figure 1: SAHA architecture

5 Utilizing ontology services

An important feature of SAHA is its ability to connect to the ONKI ontology service framework developed in the FinnONTO-project [Komulainen *et al.*, 2005]. It allows annotators not only to use concepts defined in large shared ontologies, but also to share instances they are creating with other annotators that work on different annotation projects. This kind of instance-sharing and use of shared ontologies is vital when pursuing the semantic interoperability between different Semantic Web systems.

6 User interface

The user interface of SAHA is built on two main pages. In the first page depicted in figure 2, an annotator sees the class hierarchy of an annotation schema and a list of annotations belonging to each class. A new annotation is created by choosing a class and defining a URL of the document to be annotated. In order to give more specific description of the document, the annotator can also create a new sub-class for an annotation schema class.

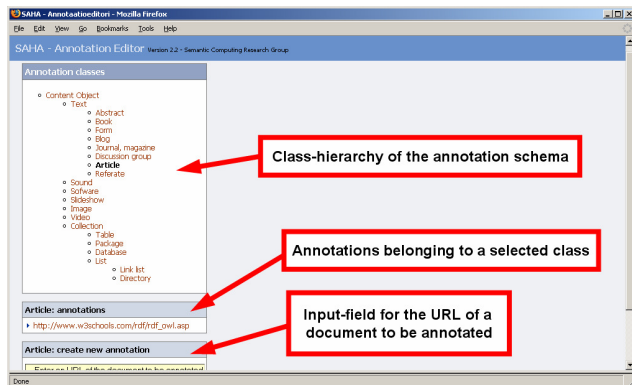


Figure 2: SAHA's user interface: the class-selection page

In the second page depicted in figure 3, annotator can edit the annotation and view the document to be annotated. The page shows the properties of the selected annotation-class in a simple form. In case of an object property, annotator can query existing instances, create a new one, or use the ONKI service to find a suitable concept defined in an external ontology. All queries to the ONKI service or to a local knowledge base containing instances use semantic autocompletion [Hyvönen and Mäkelä, 2006], a semantic extension for the idea of completing input search keywords on-line, proposed

in Google Suggest⁴. In order to make the annotation process as simple as possible, technical concepts, such as URIs, are hidden from the user. Ontology-editing features, such as defining new properties are excluded as well.

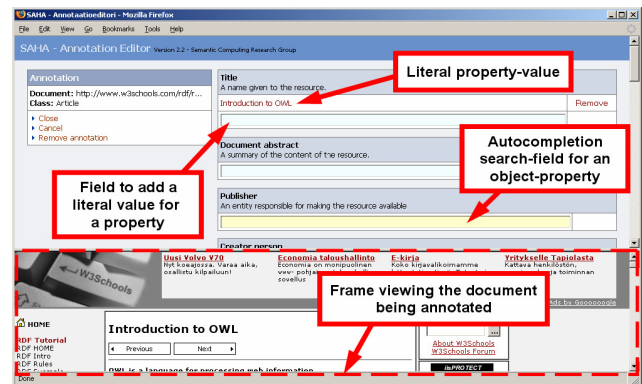


Figure 3: SAHA's user interface: the annotation-page

7 Discussion

SAHA is a working prototype. It is in trial use for the distributed content creation of the semantic eHealth portal TerveSuomi.fi⁵. Future plans include using SAHA to provide metadata for various semantic portals developed in the FinnONTO-project⁶, and to research the integration of the semiautomatic annotation framework Poka⁷ with SAHA.

Our research is a part of the FinnONTO-project funded mainly by the Finnish Funding Agency for Technology and Innovation (Tekes).

References

- [Handschuh and Staab, 2002] Siegfried Handschuh and Steffen Staab. Authoring and annotation of web pages in CREAM. In *Proceedings of the 11th international conference on World Wide Web*, Honolulu, USA, May 2002.
- [Hyvönen and Mäkelä, 2006] Eero Hyvönen and Eetu Makelä. Semantic autocompletion. In *Proceedings of the 1st Asian Semantic Web Conference (ASWC)*, Beijing, China, September 2006.
- [Komulainen *et al.*, 2005] Ville Komulainen, Arttu Valo and Eero Hyvönen. A Tool for Collaborative Ontology Development for the Semantic Web. In *Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005)*, Madrid, Spain, Nov. 2005.
- [Uren *et al.*, 2006] Victoria Uren, Philipp Cimiano, José Iria, Siegfried Handschuh, Maria Vargas-Vera, Enrico Motta and Fabio Ciravegna. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Journal of Web Semantics*, 4(1):14–28, January 2006.

⁴ <http://labs.google.com/suggest/>

⁵ <http://www.seco.tkk.fi/applications/tervesuomi/>

⁶ <http://www.seco.tkk.fi/projects/finnonto/>

⁷ <http://www.seco.tkk.fi/applications/poka/>