SUBLIMA: a software stack for Inter-operable Topic-Centred Information Portals

Robert H.P. Engels[1] and David Norheim[2]

[1] ESIS, Oslo, Norway
{ robert.engels@esis.no }
[2] Computas AS, Oslo Norway
{ dn@computas.no }

Abstract. SUBLIMA aims at publishing topic-based information in interoperable portal solutions and is available under an open source license.
Keywords: Library, Archive, Semantic Web Portal, SPARQL, query dispatching, CMS, Jena, Apache, Cocoon, Struts2 end

Introduction

Norway's public sector has been keen on adapting open standards in governmental organisations. The Norwegian Archive, Library and Museum Authority (ABM-utvikling)'s program on a Norwegian Digital Library has led to the SUBLIMA project, in which an open standard based, open-source software for creation and drift of topic portals is developed.

Potential users of this software are libraries, archives, musea and other interestees wishing to offer inter-operable topic-based portals to their users.

In order to fulfil the requirements an implementation of a complete open-source based stack based on W3C endorsed Semantic Web standards has been implemented.

Technology: defining and implementing the SUBLIMA Open Source project

Since a major requirement for the SUBLIMA stack is genericness, two organisations serve as independent domain providers. In addition to this, there are a number of potential users evaluating the two implementations. Definition of the stack's components is based on experience from previous projects, in addition to testing of several components in an in-house environment.

The system requirements called for a flexible data model, using vocabularies from e.g. Dublin Core. The domain model needed where to be migrated from the existing taxonomy with associative relations as well as full flexibility in the depth of hierarchies. The domain model also needed to be organic, and were to be used and maintained by the librarian responsible for each portal implementation.

Based on this, the decision was taken to implement the system using W3C Semantic Web recommendations. Building on previous practical experiences with

the use of RDF and OWL in knowledge management and search & navigation scenarios, the project partners applied this experience in the development of the architecture.

The client consists of a search interface allowing users to search using free text and advanced meta-data search. The search string is transfered into a structured SPARQL query to run against the SPARQL dispatcher.

The dispatcher allows federation of the query over various SPARQL endpoints and SPARQL wrappers or transformers to other query languages, e.g. the library sectors CQL language.

The architecture back-end consist of an RDF Store with SPARQL interfaces. Though the existing choice is Postgress database with Jena on top (or rather Joseki) the SPARQL interface allows the replacement of the underlying storage with more scalable store when or if the performance becomes an issue. The SPARQL endpoint also makes use of free text indexing using LARQ (an extension of the SPARQL implementation of Jena using Apache Lucene). The dispatcher allows federation of the query over various SPARQL endpoints and SPARQL wrappers/transformers to other query languages, e.g. the library sectors CQL language. The topic-ontologies are modelled using SKOS vocabulary as they are currently rather informal hierarchies with associative links. The system allows for evolution into a future OWL representation of the ontology. SUBLIMA extensively reuses common RDFS and OWL vocabularies. Current implementations of the SUBLIMA stack use graphical (hyperbolic-view based) visualisations as well as text-based interfaces for browsing and selecting ontology nodes, depending on application domains.

Implementations: the Oslo Public Library (Deichmanske bibliotek) and the Medical Library (University of Oslo)

The Oslo Public Library serves the county of Oslo and is Norways largest public library . The collection includes both fiction, non-fiction, and other media, in addition to a wide selection of childrens literature, currently holding about 1850 topics and 4600 resources, yielding ca. 100.000 triples in Turtle. Another implementation at the University in Oslo is the Scandinavian Medical Information for Layman (SMIL), which offers quality controlled meta-data related to health, illnesses and treatments. The current SMIL base consists of 8500 records creating around 250.000 triples in four languages represented in Turtle. Both portals and the Open Source licensed software stack used to create them are available online and will be demonstrated at the poster sessions at the ISWC08.