

The Finnish Law as a Linked Data Service

Matias Frosterus, Jouni Tuominen, Mika Wahlroos, and Eero Hyvönen

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

Abstract. Juridical information is important to organizations and individuals alike and is linked to from all walks of life. The Finnish government has published the Finlex Data Bank¹ for searching and browsing legislation documents. However, the data there is not yet open, is based on a traditional XML schema, and does not conform to new semantic metadata standards. There are many difficulties in maintaining and using the site in, e.g., data harvesting, interoperability, querying, and linking that could be mitigated by the Semantic Web technologies. This paper presents an approach and a project—including first results—for publishing and using the Finnish legislation as a 5-star Linked Open Data service.

Publishing and using juridical information is challenging in many ways. It is produced by different parties, such as governmental bureaus, ministries, different levels of courts, research organizations, and media. The content is heterogeneous and produced using differing tools, data formats, and practices. The links between documents are often informal and/or not made explicit. The law in general is a dynamic, changing entity: for example, it is important to be able to refer to different versions of a law at different points of time. These challenges can be addressed through the use of linked data techniques [1]. Further consideration should be paid to offering the linked juridical data in the form of ready-to-use services and end-user applications.

Our goal² is to create a 5-star linked open data service of the Finnish legislation, a semantic version of Finlex. Based on the service, a next generation of public and commercial legal systems can be built. The work involves 1) juridical ontology development, 2) metadata modeling of legislative documents, 3) the publication of the content as linked data, and 4) the evaluation of the system via application demonstrators.

Ontologies We have begun the work by harmonizing different in-use thesauri and building an ontology of legal concepts. A project by the Finnish Ministry of Justice combined ca. 30 vocabularies and thesauri used by different courts in Finland, and we further added two vocabularies used by commercial legal information services. The collected ca. 9,000 different terms were then mapped with the larger KOKO ontology cloud (ca. 45,000 concepts) that is used in Finland. In addition, the anticipated ontology infrastructure includes two classification schemes of law.

¹ <http://www.finlex.fi/en/>

² The work is a collaboration with the Ministry of Justice, the Ministry of Communications, Edita Publishing Ltd, and Talentum Corp. under the national Linked Data Finland project (<http://www.seco.tkk.fi/projects/ldf/>) funded mainly by Tekes.

Metadata model A first version of a metadata model for representing the structure and linking in the Finnish law has been designed in RDF, based on the Finlex XML schema.

Automatic annotation We aim at annotating the subject matter of legislation, case documents, and legal news using the presented ontologies and metadata models. In this way, the documents could be interlinked also in terms of their content (e.g., materials related to family crimes), not only in terms of the legislative structures now used in Finlex (e.g., that a law X is based on or replaces another law Y). The first dataset, a collection of 35,000 legal news items, is now being annotated.

Publication The data will be published as linked data, including datasets for the laws, court cases, and preparatory works, which are often critical when interpreting the laws. Aside from a standard SPARQL endpoint publication, the laws will be published also in the ONKI ontology service, providing not only a browsing facility but a selector widget with autocompletion for annotation, and APIs for machine use [3]. Each law and its version has a URI that can be found and fetched as an unambiguous reference from ONKI.

Evaluation via Demonstrators New semantic search and recommendation facilities for Finlex will be demonstrated using our SAHA-HAKO tool³ [2]. This tool provides means for data editing and analysis of data quality against the metadata models. The system provides a SPARQL endpoint to its data store, a download facility, and incorporates a faceted search engine on top of SPARQL. In this case, useful facets include the laws and their sections, time, subject matter, as well as the law classification systems in use. The recommendation system would allow, e.g., studying a given case with links to similar cases that reference to the same laws. This can be further refined by noting the cases where the law has changed, so that these can be viewed separately. Finally, users can be offered a possibility of subscribing to the laws and specific sections of laws of interest to them, so that they would receive a notification, if a law in their interest has changed. This is especially pertinent in situations where, e.g., an organization's internal regulation rests upon a specific law whose changes can be critical to the validity of the regulation.

References

1. Hoekstra, R.: The MetaLex Document Server legal documents as versioned linked data. In: Proceedings of the 10th international conference on The semantic web - Volume Part II. pp. 128–143. ISWC'11, Springer-Verlag, Berlin, Heidelberg (2011)
2. Kurki, J., Hyvönen, E.: Collaborative metadata editor integrated with ontology services and faceted portals. In: Workshop on Ontology Repositories and Editors for the Semantic Web (ORES 2010) at ESWC 2010. CEUR Workshop Proceedings, Vol. 596 (2010)
3. Viljanen, K., Tuominen, J., Hyvönen, E.: Ontology libraries for production use: The Finnish ontology library service ONKI. In: Proceedings of ESWC 2009. Springer-Verlag (2009)

³ <http://code.google.com/p/saha/>