Publishing and Using Legislation and Case Law as Linked Open Data on the Semantic Web

Eero Hyvönen[1,2][0000−0003−1695−5840], Minna Tamper[1,2][0000−0003−1695−5840],
Esko Ikkala[1][0000−0002−9571−7260], Sami Sarsa[1][0000−0002−7277−9282],
Arttu Oksanen[1,3][0000−0003−2327−6942], Jouni Tuominen[1,2][0000−0003−4789−5676], and
Aki Hietanen[4]

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland
https://seco.cs.aalto.fi, firstname.lastname@aalto.fi
[2] HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
https://heldig.fi
[3] Edita Publishing Ltd.
https://www.editapublishing.fi
[4] Ministry of Justice, Finland
https://oikeusministerio.fi, firstname.lastname@om.fi

Abstract. Legislation and case law are widely published on the Web as documents for humans to read. In contrast, this poster paper argues for publishing legal documents as Linked Open Data (LOD) on top of which intelligent legal services for end users can be created in addition to just providing the documents for close reading. To test and demonstrate this idea, we present work on creating the Linked Open Data service SEMANTIC FINLEX for Finnish legislation and case law and the semantic portal prototype LAWSAMPO for serving end users with legal data. SEMANTIC FINLEX is a harmonized knowledge graph that is created automatically from legal textual documents and published in a SPARQL endpoint on top of which the various applications of LAWSAMPO are implemented. First applications include faceted semantic search and browsing engines for 1) statutes and 2) court decisions, as well as 3) a service for finding court decisions similar to a given one. A novelty of LAWSAMPO is the provision of ready-to-use tooling for exploring and analyzing legal documents, based on the "Sampo" model.

Keywords: linked data · case law · legislation · semantic portal

1 Semantic Finlex Linked Open Data Service

In Finland, legislation and case law have been published as web documents since 1997 as web documents in the Finlex Data Bank[5]. Although the Finlex service is widely used by the public, it does not provide machine-readable legal information as open data, on top of which services and analyses could be built by the ministry or third party vendors. The first version of SEMANTIC FINLEX based on Linked Data was published in 2014 [4]. The data included 2413 consolidated laws, 11 904 judgments of the Supreme

---

[5] http://www.finlex.fi

Court, and 1490 judgments of the Supreme Administrative Court. In addition, some 30 000 terms used in 26 different thesauri were harvested for a first draft of a consolidated vocabulary. During this work, some shortcomings of the initial RDF data model became evident as well as the need for using the then emerging new standards for EU level interoperability: ELI European Legislation Identifier [3] and ECLI European Case Law Identifier [2]. The dataset also consisted of only one version (2012) of the statutory law and was not updated, as new legislation and case law was published in Finlex. These issues were resolved in the new version of the SEMANTIC FINLEX [10] that currently hosts a dataset comprising approximately 28 million triples. The data was enriched by automatic annotation to named entities and references to legal texts, to corresponding vocabularies, and to data sources by combining statistics- and rule-based named entity recognition with named entity linking [13].

The Semantic Finlex service adopts the 5-star Linked Data model[6], extended with two more stars, as suggested in the Linked Data Finland model and platform [7]. The 6th star is obtained by providing the dataset schemas and documenting them. Semantic Finlex schemas can be downloaded from the service and the data models are documented under the `data.finlex.fi` domain. The 7th star is achieved by validating the data against the documented schemas to prevent errors in the published data. Semantic Finlex attempts to obtain the 7th star by applying different means of combing out errors in the data within the data conversion process. The service is powered by the Linked Data Finland[7] publishing platform that along with a variety of different datasets provides tools and services to facilitate publishing and re-using Linked Data. All URIs are dereferenceable and support content negotiation by using HTTP 303 redirects. In accordance with the ELI specification, RDF is embedded in the HTML presentations of the legislative documents as RDFa[8] markup. In addition to the converted RDF data, the original XML files are also provided. To support easier use by programmers without knowledge of SPARQL or RDF, a simplified REST API is provided, too.

2   LAWSAMPO Semantic Portal

To demonstrate and test using the SEMANTIC FINLEX in applications, the semantic portal LAWSAMPO is under development. LAWSAMPO is a new member of the Sampo[9] series of semantic portals, based on the "Sampo model" [6], where the data is enriched through a shared ontology and Linked Data infrastructure, multiple application perspectives are provided through a single SPARQL endpoint, and faceted search and browsing is integrated with data-analytic tooling in the user interface. The faceted search and tooling are implemented using the Sampo-UI framework[10] [8]. The Sampo portals[11]

---

[6] `https://www.w3.org/DesignIssues/LinkedData.html`

[7] `http://ldf.fi`

[8] `http://www.w3.org/standards/techs/rdfa`

[9] In Finnish mythology and the epic Kalevala, "Sampo" is a mythical artefact of indeterminate type that gives its owner richness and good fortune, an ancient metaphor of technology.

[10] Cf. homepage for more info: `https://seco.cs.aalto.fi/tools/sampo-ui/`.

[11] Including, e.g., CultureSampo (2008) for cultural heritage, TravelSampo (2011) for tourism, BookSampo (2011) for fiction literature, WarSampo (2015) for military history, Biogra-

have had millions of end users on the Web suggesting that it is a promising model to create useful semantic portals.

The landing page of the LawSampo portal offers three application perspectives: **1. Statutes.** By clicking on Statutes, a faceted search interface [14] for searching and browsing statutes is opened. The facets on the left include document type (with seven subtypes), statute type, year, and related EU regulation. After filtering out a set of documents (or a particular document) of interest, the user is provided with two options. First, the user can select a document from the result list and a "homepage" of the document opens, showing not only the document but also linked contextual information related to it such as referred EU regulations linked to EU Cellar[12] or other documents from Semantic Finlex referring to it. For example, court decisions in which the statute has been applied can be shown. Second, it is possible to do data analysis based on the filtered documents. For example, a histogram can be created showing the dates of the filtered documents. **2. Case Law.** In the Case Law perspective, a similar faceted search interface opens for searching and browsing court decisions. In this case, the facets include court, judge, and keywords characterizing the subject matter of the judgement. **3. Case Law Search.** The third perspective is an application, where a law case judgement, or more generally any document or text, can be used for finding similar other case judgements. For example, if one gets a judgement from a court, this application can be used to find out what kind of similar judgements have been made before. Several methods for finding similar cases were tested when implementing this application including TF-IDF, Latent Dirichlet Allocation (LDA), Word2Vec, and Doc2vec [11,12].

Figure 1 depicts the end-user interface of the Case Law Search perspective application [11]. The user is able to input a text document as a query to Finlex case law either by uploading a file or by writing text directly into the form. The query document is seen on the subwindow "Document content". Supported file formats for uploading documents are plain text, XML, PDF, and with Tesseract OCR, image formats, such as JPEG and PNG. The text extraction mode to be used can be selected by the drop-down menu on the right bottom corner. The search form also allows the user to choose the algorithm that ranks the documents by using the drop-down menu on the left bottom in Figure 1. Here the method "Ensemble" combining several methods is selected for accurate results. Ranking with some algorithms may work better than others for certain topics, or depending on what kind of relatedness is preferred. Also the preferred result size can be specified.

3 Related Work and Contributions

Our work on legal Linked Data services was influenced by the MetaLex Document Server[13] [5] that publishes Dutch legislation using the CEN Metalex XML and ontology standards. Other Metalex ontology based implementations include legislation.gov.uk[14]

---

phySampo (2018) for prosopography, and NameSampo (2019) for toponomastic research. Cf. homepage: `https://seco.cs.aalto.fi/applications/sampo/`.

[12] `https://data.europa.eu/euodp/en/data/dataset/sparql-cellar-of-the-publications-office`

[13] `http://doc.metalex.eu`
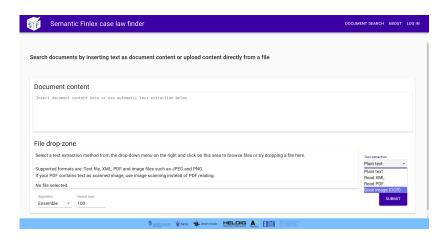
[14] `http://legislation.gov.uk`

Fig. 1. LawSampo Case Law finder for searching similar cases.

for UK legislation and Nomothesia[15] that also implements ELI-compliant identifiers. Various ELI implementations and prototypes have also been implemented and used in related existing legal information portals nationally, e.g., in Luxembourg[16], France[17], and Norway[18]. Many countries already produce ECLI-compliant case law documents to be indexed by the ECLI search engine[19]. A prominent example of publishing EU Law and publications as linked data is the CELLAR system. Previous related works in the U.S. include, e.g., the Legal Linked Data project aiming at enhanced access to product regulatory information [1].

LAWSAMPO aims to widen the focus of these related works by providing both legislation and case law to end users through intelligent user interfaces, such as semantic faceted search and document similarity-based search. The documents are automatically enriched with contextual linked data on their "homepages", and the end user is also provided with ready-to-use data-analytic tooling for analyzing the documents and their relations. In the future, we plan to expand the related enriching datasets to include, e.g., related parliamentary documents and discussions[20], in the spirit of [15]. In order to be able to publish more legal documents in a more cost-efficient way, we also work on semi-automatic pseudonymization of court judgements [9] and automatic annotation of legal documents [13].

---

[15] `http://legislation.di.uoa.gr`

[16] `http://legilux.public.lu/editorial/eli`

[17] `http://www.eli.fr/en/constructionURI.html`

[18] `http://lovdata.no/eli`

[19] `https://e-justice.europa.eu/content_ecli_search_engine-430-en.do`

[20] The ParliamentSampo system: `https://seco.cs.aalto.fi/projects/semparl/en/`.

References

1. Casellas, N., Bruce, T.R., Frug, S.S., Bouwman, S., Dias, D., Lin, J., Marathe, S., Rai, K., Singh, A., Sinha, D., Venkataraman, S.: Linked legal data: Improving access to regulations. In: Proceedings of the 13th Annual International Conference on Digital Government Research. pp. 280–281. dg.o '12, Association for Computing Machinery, New York, NY, USA (2012). https://doi.org/10.1145/2307729.2307785
2. Council of the European Union: Council conclusions inviting the introduction of the European Case Law Identifier (ECLI) and a minimum set of uniform metadata for case law. In: Official Journal of the European Union, C 127, 29.4.2011. pp. 1–7. Publications Office of the European Union (2011)
3. Council of the European Union: Council conclusions inviting the introduction of the European Legislation Identifier (ELI). In: Official Journal of the European Union, C 325, 26 October 2012. pp. 3–11. Publications Office of the European Union (2012)
4. Frosterus, M., Tuominen, J., Hyvönen, E.: Facilitating re-use of legal data in applications–Finnish law as a linked open data service. In: Proceedings of the 27th International Conference on Legal Knowledge and Information Systems JURIX 2014, Krakow, Poland. pp. 115–124. IOS Press (2014)
5. Hoekstra, R.: The MetaLex Document Server legal documents as versioned linked data. In: Proceedings of the ISWC 2011, Bonn, Germany. pp. 128–143. Springer-Verlag (2011)
6. Hyvönen, E.: Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-analysis and Serendipitous Knowledge Discovery. Semantic Web – Interoperability, Usability, Applicability **11**(1), 187–193 (2020). https://doi.org/10.3233/SW-190386
7. Hyvönen, E., Tuominen, J., Alonen, M., Mäkelä, E.: Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In: Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., Tordai, A. (eds.) The Semantic Web: ESWC 2014 Satellite Events. ESWC 2014. pp. 226–230. Springer-Verlag (May 2014), https://doi.org/10.1007/978-3-319-11955-7_24
8. Ikkala, E., Hyvönen, E., Rantala, H., Koho, M.: Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. Semantic Web – Interoperability, Usability, Applicability (2020), submitted
9. Oksanen, A., Tamper, M., Tuominen, J., Hietanen, A., Hyvönen, E.: ANOPPI: A pseudonymization service for Finnish court documents. In: Legal Knowledge and Information Systems. JURIX 2019: The Thirty-second Annual Conference. IOS Press (2019)
10. Oksanen, A., Tuominen, J., Mäkelä, E., Tamper, M., Hietanen, A., Hyvönen, E.: Semantic Finlex: Transforming, publishing, and using Finnish legislation and case law as linked open data on the web. In: Peruginelli, G., Faro, S. (eds.) Knowledge of the Law in the Big Data Age, Frontiers in Artificial Intelligence and Applications, vol. 317, pp. 212–228. IOS Press (2019)
11. Sarsa, S.: Information Retrieval with Finnish Case Law Embeddings. Master's thesis, University of Helsinki, Dept. of Computer Science (2019)
12. Sarsa, S., Hyvönen, E.: Searching Case Law Judgements by Using Other Judgements as a Query (2019), Aalto University, paper: https://seco.cs.aalto.fi/publications/2019/sarsa-et-al-csfinder.pdf
13. Tamper, M., Oksanen, A., Tuominen, J., Hietanen, A., Hyvönen, E.: Automatic annotation service appi: Named entity linking in legal domain. In: Proceedings of ESWC 2020, Posters and Demos. Springer–Verlag (June 2020), accepted
14. Tunkelang, D.: Faceted search. Morgan & Claypool Publishers, CA, USA (2009)
15. Van Aggelen, A., Hollink, L., Kemman, M., Kleppe, M., Beunders, H.: The debates of the European parliament as linked open data. Semantic Web **8**(2), 271–281 (2017)