Searching and Analyzing Cross-border Multilingual Legislation on the Semantic Web

Eero Hyvönen[1,2][0000−0003−1695−5840], Hien Cao[1][0009−0004−2774−8243], Rafael Leal[1][0000−0001−7266−2036], Heikki Rantala[1][0000−0002−4716−6564], and Aki Hietanen[3]

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland
[2] Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki, Finland
[3] Ministry of Justice, Finland

Abstract. This paper demonstrates how Linked Open Data (LOD) can be used for publishing, searching, and analyzing legislative documents in an international cross-broader multilingual setting. Cross-border services are needed, e.g., when one is moving from one country to another and looking for regulations for immigration, heath care, education, etc. in one's own language. The main novelty of the FinEstLawSampo demonstrator presented, based on legislation from Finland, Estonia, and the EU, is the provision of heterogeneous cross-country, multilingual, distributed legal data through multiple application perspectives for faceted searching, browsing, and for data analysis in legal informatics.

Keywords: Linked data, Law, Multilingual, Semantic portal, Data service

1 Introduction

Cross-border access to legislation published in different countries is often needed in international settings, such as the EU. Although legislation is often available openly, it is not necessarily Findable, Accessible, Interoperable, and Reusable (FAIR[4]). A specific problem is that the data and User Interfaces (UI) are typically available only in local languages that the end-user does not understand. In addition, different local keyword vocabularies for subject matter indexing and classification systems are used in different countries, which sets challenges for querying the data semantically and for precision and recall of information retrieval. Furthermore, legal documents are often available only as texts for the humans to read and not as data for machines, which makes them hard to use in applications of legal informatics[5] [3], e.g., in computational law[6].

There are two basic approaches for querying data on the Web. The *federated strategy* is to send the query to distributed local data services, collect the answers,

[4] https://www.go-fair.org/fair-principles/
[5] https://en.wikipedia.org/wiki/Legal_informatics
[6] https://law.stanford.edu/2021/03/10/what-is-computational-law/

and present them to the user. In the *centralized approach* the distributed heterogeneous datasets are first aggregated and harmonized into a global database for querying. Our demonstrator is based on the centralized approach. It is shown how legislation can be published, translated, and used as LOD, based on language-agnostic indexing schemes and/or by aligning local schemes onto each other. The data used is available as a LOD SPARQL endpoint[7] on top of which the portal FinEstLawSampo was created[8]. The software is available in Github[9].

Our data model reuses that of LawSampo[10] [6]. It is reused also for the Estonian statutes that were available in custom XML format[11], in Estonian and in English, and were transformed into RDF. The original Finnish and Estonian statutes were linked internally and externally to EU directives. NLP techniques were used for enriching the data further: 1) Legal keywords were extracted using EuroVoc[12] due to its cross-border nature (labels in 24 EU languages). 2) The documents were classified automatically using official life event classifications in use in Finland[13](9 event types) and in Estonia[14] (12 event types). The KG, published on the Linked Data Finland platform [15], contains nearly 13 000 Finnish and Estonian statutes, and 5000 EU directives from the EU Cellar[16].

2 Using the FinEstLawSampo Portal

The FinEstLawSampo portal was implemented using the Sampo-UI framework [8,10] on top of a LOD service available[17] at the Linked Data Finland platform [7]. In the Sampo-UI model the user first comes on the *landing page* with *application perspectives* to the data, based on classes of the underlying KG. By selecting a perspective, the user can filter the instances of the perspective class by using faceted semantic search. After this it is possible to analyze the result set by seamlessly integrated data-analytic tools available as tabs, or alternatively investigate and browse individual results. The facets are based on the property values of the perspective class. In our case two perspectives are provided: *Statutes* (Finnish and Estonian) and *EU directives*.

Eight facets can be used in the Statutes perspective. Fig. 1 depicts the faceted search interface for statutes with the eight facets on the left: 1) Traditional text search facet (of statute content), 2) Statute type (there are eight types in use), 3) Enforcement date, 4) EuroVoc keyword, 5) Finnish life situation (nine options), 6) Estonian life situation (eleven options), 7) EU directive (mentioned

[7] LOD data service online: `https://ldf.fi/datasets/finestlaw`
[8] Portal online: `https://finestlaw.demo.seco.cs.aalto.fi/en`
[9] `https://github.com/SemanticComputing/finestlaw-web-app`
[10] Documented at the namespace URI `http://ldf.fi/schema/lawsampo/`
[11] Consolidated texts of of Estonian legislation: `https://www.riigiteataja.ee/en/`
[12] `https://data.europa.eu/data/datasets/eurovoc`
[13] Finnish classification of life events: `https://www.suomi.fi/citizen`
[14] Estonian classification of life events: `https://www.eesti.ee/en`
[15] `http://ldf.fi`
[16] `https://op.europa.eu/fi/web/eu-vocabularies/cellar`
[17] THe data service can be found at: `https://www.ldf.fi/dataset/finestlaw`

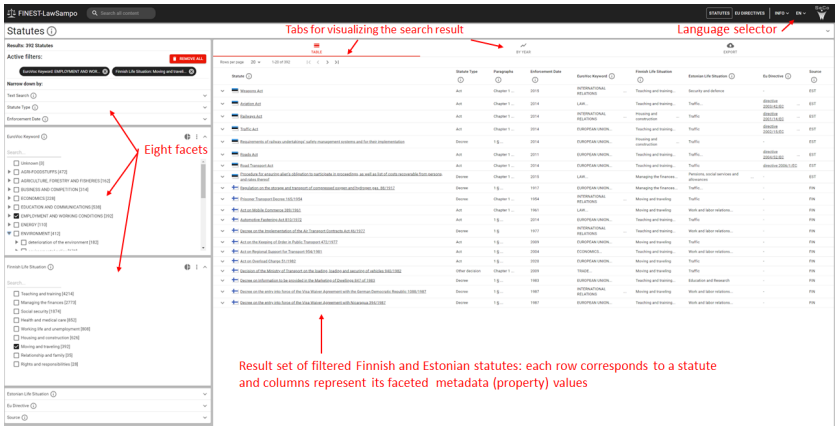

Fig. 1. Faceted search for Finnish and Estonian statutes

in the statute), and 8) Source of legislation (Finland or Estonia). The user in the figure in considering moving from Finland to Estonia for working and has therefore selected from the EuroVoc keyword facet `EMPLOYMENT AND WORKING CONDITIONS` and from the Finnish life situation facet `Moving and traveling`. The results are shown on the `TABLE` tab on the right with country flags showing the source country of the statutes. By selecting the language on the menu on the upper right corner the language for providing the statutes can be selected: English, Finnish, or Estonian can be selected. In the figure, English is selected and the Estonian statutes are provided in English (and not in their original language Estonian). By selecting Finnish the statutes are automatically translated into Finnish. Also the language of the UI is changed accordingly.

In some cases, human made official translations of the statutes are available, but if not automatic machine translation is used. For example, Estonian statutes have been published officially in English, too, but not in Finnish. When machine translations are provided the end-user is explicitly warned that the content has been translated by AI and may not be fully correct. This is important especially in the legal domain in order to avoid misunderstandings. Based on informal testing, machine translations seem pretty good and useful.

In Sampo-UI-based interfaces the faceted search results can be analyzed on the tabs. For example, if the tab `BY YEAR` in Fig. 2 is selected, the search results are projected on a timeline based on the enforcement date. This visualization shows how the number of enforced statutes have evolved in time. The tab EXPORT opens a Yasgui SPARQL editor [11] window with the query used for filtering the current results. The idea of this tab is to give a user who knows SPARQL an easy way to use the underlying LOD service directly.

The EU directives perspective makes it possible to search some 5000 EU directives based on EuroVoc keyword annotations. For each directive, links to Finnish and Estonian statutes related to the directive are provided.

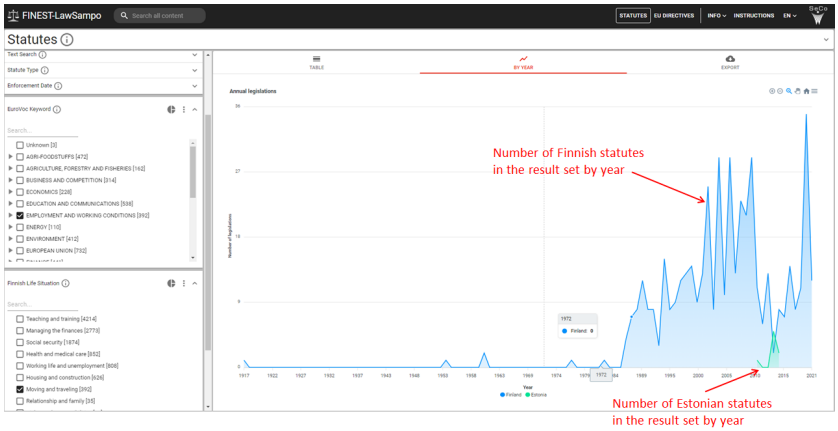

Fig. 2. Visualizing the number of statutes on a timeline by their enforcement day

3 Related Works and Discussion

Legislation is widely published online [9] as portals, such as legislation.gov.uk for the legislation for the UK, Scotland, Wales, and Northern Ireland[18], and EU level systems, such as HUDOC[19], EUR Lex [20], the EU Cellar[21], and the ECLI Search Engine[22] for the case law. Our work was influenced by the MetaLex Document Server[23] [4], the EU Cellar, LawSampo [6], and national legal online services in Greece, Luxemburg[24], France, Norway[25], and the U.S. [2].

A major point of comparison of our work is the cross-border N-Lex system[26], based on federated text search using local national legal web services. In FinEstLawSampo the challenges of the federated text search are addressed by harmonizing and aggregating all data, including available man-made translations, into a global KG on top of which the portal was developed. The contents were made semantically interoperable using ontologies. In this way arguably better search results are obtained but the price is more complex implementation.

Usability of the Sampo model has been evaluated in some other Sampo portals [1] suggesting feasibility of the model in general. However, formal end user

[18] https://www.legislation.gov.uk
[19] https://hudoc.echr.coe.int/
[20] https://eur-lex.europa.eu/
[21] https://data.europa.eu/euodp/en/data/dataset/sparql-cellar-of-the%2Dpublications-office
[22] https://e-justice.europa.eu/content_ecli_search_engine-430-en.do
[23] http://doc.metalex.eu
[24] http://legilux.public.lu/editorial/eli
[25] http://lovdata.no/eli
[26] https://n-lex.europa.eu/n-lex/

testing of FinEstLawSampo remains a topic of further research. The scalability of the Sampo-UI framework depends on the complexity of the data model used. The number of instances to be searched for in largest Sampo systems [5] has been around one million.

Acknowledgments This research was funded by the Nordic Council of Ministers and the Cross-border Digital Services (CDBS) programme. CSC – IT Center for Science provided computational resources for the project.

References

1. Burrows, T., Pinto, N.B., Cazals, M., Gaudin, A., Wijsman, H.: Evaluating a Semantic Portal for the "Mapping Manuscript Migrations" Project. DigItalia **2**, 178–185 (2020), http://digitalia.sbn.it/article/view/2643
2. 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: Proc. of the 13th Annual International Conf. on Digital Government Research (dg.o '12). pp. 280–281. Assoc. for Comp. Machinery (2012)
3. Erdelez, S., O'Hare, S.: Legal informatics: Application of information technology in law. Annual Review of Information Science and Technology **32** (01 1997)
4. Hoekstra, R.: The MetaLex Document Server legal documents as versioned linked data. In: Proceedings of the ISWC 2011. pp. 128–143. Springer (2011)
5. Hyvönen, E.: Digital humanities on the semantic web: Sampo model and portal series. Semantic Web – Interoperability, Usability, Applicability **14**(4), 729–744 (2022). https://doi.org/10.3233/SW-190386
6. Hyvönen, E., Tamper, M., Oksanen, A., Ikkala, E., Sarsa, S., Tuominen, J., Hietanen, A.: LawSampo: A semantic portal on a linked open data service for Finnish legislation and case law. In: The Semantic Web: ESWC 2020 Satellite Events. Revised Selected Papers. pp. 110–114. Springer (2019)
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: ESWC 2014 Satellite Events. pp. 226–230. Springer (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 **13**(1), 69–84 (January 2022). https://doi.org/10.3233/SW-210428
9. van Opijnen, M., Peruginelli, G., Kefali, E., Palmirani, M.: Online publication of court decisions in europe. Legal Information Management **17**, 136—-145 (2017). https://doi.org/10.1017/S1472669617000299
10. Rantala, H., Ahola, A., Ikkala, E., Hyvönen, E.: How to create easily a data analytic semantic portal on top of a SPARQL endpoint: introducing the configurable Sampo-UI framework. In: VOILA! 2023 Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs 2023. CEUR Workshop Proceedings, Vol. 3508 (2023), https://ceur-ws.org/Vol-3508/paper3.pdf
11. Rietveld, L., Hoekstra, R.: The YASGUI family of SPARQL clients. Semantic Web – Interoperability, Usability, Applicability **8**(3), 373–383 (2017). https://doi.org/10.3233/SW-150197