A Data-driven Approach to Create an Ontology of Parliamentary Work: Case Parliament of Finland on the Semantic Web

Eero Hyvönen[1,2], Petri Leskinen[1,2] and Jouni Tuominen[3,2,1]

[1]Semantic Computing Research Group (SeCo), Department of Computer Science, Aalto University, Finland
[2]Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki, Finland
[3]Helsinki Institute for Social Sciences and Humanities (HSSH), University of Helsinki, Finland

Abstract

When creating a Semantic Web application, an ontology for modeling the domain of discourse is typically first created and then populated by individual data instances. However, in many cases formulating the data model before data ingestion may not be feasible. This paper discusses this challenge when modeling the work at a parliament including organizational changes over a long period of time. Here the activities and organizational structures, such as parties, groups, and committees preparing legislation at different times, may not be known. However, there may be data available regarding the individuals participating in the parliamentary work in different roles and times. As a solution, a data-driven approach for ontology construction is proposed where the ontology is constructed automatically from data about its individuals. This idea has been applied to create an ontology of the Parliament of Finland (PoF) using a historical database of the Members of Parliament (MP) and their activities. Based on the ontology, two KGs were created and published in a Linked Open Data (LOD) service on top on which a semantic portal ParliamentSampo in use was created with thousands of end users.

Keywords

ontologies, parliamentary studies, semantic portals, linked data, digital humanities

CEUR-WS.org/Vol-3540/paper6.pdf

1. Developing an Ontology on How a Parliament Works

A foundation of constitutional democratic states is openness and transparency of parliamentary decision making. To enhance this is the ultimate goal of our work on developing the system *ParliamentSampo – Parliament of Finland on the Semantic web* [1], targeting citizen voters, media, researchers of parliamentary studies and language, and the parliamentarians themselves. As technical means for this, the ParliamentSampo initiative chose to transform and publish all the ca. 1 million speeches of the MPs of the PoF as well as their prosopograhical data as a Linked Open Data service [2] for researchers in Digital Humanities (DH) and application developers to use. On top of the LOD service, a semantic portal was created that can be utilized by a wider audience without programming skills [3].

SWODCH'23: International Workshop on Semantic Web and Ontology Design for Cultural Heritage, November 7, 2023, Athens, Greece
eero.hyvonen@aalto.fi (E. Hyvönen)
https://seco.cs.aalto.fi/u/eahyvone/ (E. Hyvönen)
© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

Our original plan was to first design an ontology of the PoF, including classes such as Person, Party, Committee, etc., and then populate it with instance data for creating a LOD service about 1) the plenary speeches of the PoF since it was established in 1907, as well as about 2) all ca. 2800 parliamentarians who have ever spoken in the plenary sessions. After having investigated in more detail on how the PoF works [4], we realized that even creating an ontology for the contemporary PoF only would be a complex task, not to mention creating an ontology covering all parliaments since 1907 and their activities and changes. Furthermore, it turned our that explicit descriptions of how the parliaments have worked in history were no readily available. However, there was XML-data available about the parlamentarians including mentions about parties, committees, and activities at different times. It was therefore decided to try to extract the ontology in a data-driven way (semi)automatically from the available datasets.

This paper discusses lesson learned in using this approach. In the following, related research on parliamentary data is first reviewed (Section 2). After this the pipeline of producing, populating, and publishing the PoF ontology is explained (Section 3). In conclusion, results of our work are summarized (Section 4) and directions of further development discussed.

This paper extends our earlier papers pertaining to ParliamentSampo [1, 5, 6, 7, 8] by focusing on the idea of constructing a parliamentary ontology in a data-driven way. As a proof-of-concept, the PoF ontology has been successfully used in DH research and as the semantic backbone of the ParliamentSampo portal[1] published on February 14th, 2023[2]. The portal has had by now over 7000 distinct users according to Matomo.org analytics, and it has been used is several studies of the leading Finnish newspaper Helsingin Sanomat, suggesting usability and feasibility of the ontology created[3]. The PoF ontology is openly available on the Web as a live LOD service and as data dumps in CSV and RDF formats.

2. Related Work on Parliamentary Ontologies

Lots of parliamentary debate corpora have been created from the documents of both historical and contemporary parliaments [9, 10]. Web applications and data services have been developed that allow to browse, study, and download the digitized materials.[4]

Also Linked Data and Semantic Web technologies have been used in this work. The debates of the European Parliament and the political affiliation information were connected as linked data into other datasets, such as DBpedia and the EuroVoc thesaurus, in the pioneering project Linked Data of the European Parliament (LinkedEP) [12]. The LinkedEP data was made available through a SPARQL endpoint and an online user interface. The Open Data Portal of the European Parliament provides lots of datasets as LOD and in CSV format[5]. Other examples of linked data parliament initiatives are the LinkedSaeima for the Latvian parliament [13], the Italian Parliament data[6], and the historical Imperial Diet of Regensburg of 1576 project [14]. An important XML-based initiative for harmonization and annotation of national parliamentary

[1] Available at: https://parlamenttisampo.fi
[2] Publication event homepage: https://seco.cs.aalto.fi/events/2023/2023-02-14-parlamenttisampo/
[3] See the project homepage for further detais: https://seco.cs.aalto.fi/projects/semparl/
[4] See, e.g., the LiPaD project and the Canadian Hansard, https://lipad.ca [11]
[5] https://data.europarl.europa.eu/en/datasets
[6] http://data.camera.it

corpora is the ParlaMint project part of the CLARIN infrastructure.[7] The ParlaMint project applies the TEI-based Parla-CLARIN scheme[8], and aims to create uniformly annotated multilingual parliamentary corpora with its partners. The current ParlaMint II involves 27 national parliamentary corpora [15] (see also [16]).

Methods for automatic ontology construction has been developed in various application fields as well as in domain-agnostic settings [17]. In contrast to our work, based on using biographical XML-data, these methods typically focus on mining taxonomic structures from unstructured texts [18] or from datasets of, e.g., social and mobile networks [19].

Previous applications for Finnish parliamentary data, such as [20, 21, 22, 23, 24, 25], have concerned only debates and only part of the entire time series of the Finnish parliaments, without presenting and using models of the parliamentarians and their work.

The focus of work on the parliamentary data research before has been on plenary speeches that are used in many fields of research, such as linguistics, political science, legal studies, media studies, economics, and history. In contrast, this paper focuses on the question on how to create an ontology of parliamentary work to be used in annotating, searching, exploring, and analysing parliamentary materials, such as speech corpora, as demonstrated by the ParliamentSampo system [3]. In addition, the parliamentary ontology can be used for searching, exploring, and analyzing the MPs, their organizations, and networks [26, 27] in a parliament.

3. An Ontology of the Parliament of Finland

This section describes the PoF Ontology and KG of ParliamentSampo, including MPs and other speakers in the plenary sessions PoF since 1907, parties, groups and organizations involved, and parliamentary events and proceedings in time and place.

3.1. How the Parliament of Finland Works

The organization and activities of the PoF are documented in [4]. Legislation procedures in PoF can be initiated today by a *government bill* (hallituksen esitys), by a *parliamentarian's proposal* (kansanedustajan esitys) of an MP, or as *citizen's initiative* (kansalaisaloite). The process starts with *referral discussion* (lähetekeskustelu) that sends the bill to committee in whose expertise domain the bill/proposal/initiative is related to. The committees consist of 17 or 21 MPs and vice members. At the moment, there are 16 permanent committees in PoF. Based on a report of the committee, the parliament then has a *first discussion* about the legal document in question after which still some modifications to the document can be made. Later on there is a *second discussion* where the document is finally either accepted or rejected.

The work at the PoF involves people, committees, parties, and other organizations in different roles and in relation to the discussions. Furthermore, the organizational structure has evolved in time. Creating an overarching ontology over different times is a challenge due to the dynamic nature of the PoF: lots of parties, groups and other organizational units, have been established, restructured, and vanished since 1907. Reassembling the history of the PoF from literary

[7] https://www.clarin.eu/content/parlamint-towards-comparable-parliamentary-corpora
[8] https://github.com/clarin-eric/parla-clarin

documents available was deemed infeasible, and we therefore created the ontology in a data-driven fashion based on the data available concerning the MPs and other speakers in the plenary sessions and governments making the proposals.

3.2. Extracting and Populating the Ontology from Actor Data

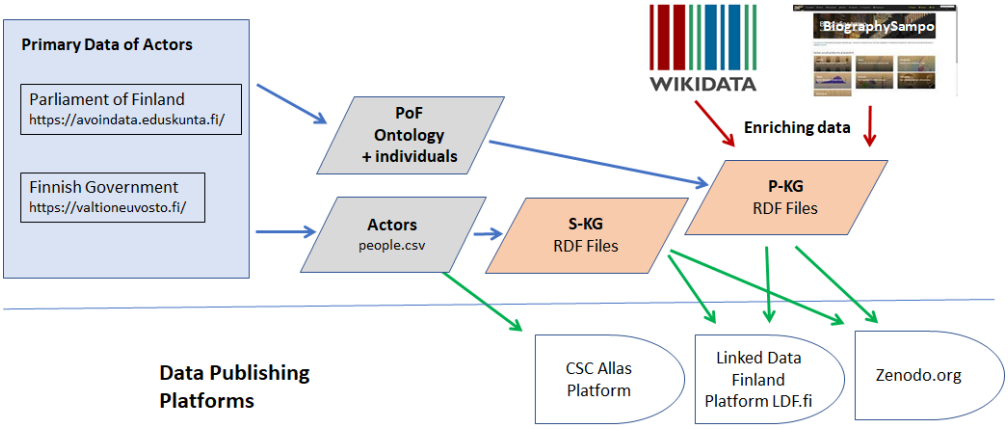

Figure 1: Pipeline of transforming data about the MPs and other speakers of the plenary sessions into the ontology of the PoF

The process of extracting the PoF ontology from data as well as populating it with instances is depicted in Figure 1. The most important data source for the PoF ontology was the database of MPs available at the Parliament Open Data portal[9]. This data is regularly updated, and contains information about 2605 MPs since 1907. The data is in a custom XML format, including basic biographical data about all MPs, such as date and place of birth, periods of time as an MP, electoral districts, memberships in parties, committees, groups, and organizations, and publications of the MPs[10]. The final data transformation result is a prosopographical knowledge graph (P-KG) published, interlinked with the speech data S-KG, as a LOD service on the Linked Data Finland platform[11] and as RDF data in Zenodo[12].

The MP XML data and some additional data about Finnish Governments were transformed into a tabular CSV table, where each row represents an MP instance (Actors.csv). This data is used as a reference to the parlamentarians when transforming their speeches into the Speech KG S-KG. The Actor data table is also published at the CSC Allas platform[13] for external users.

The MP database was enriched with data about ca. 200 additional speakers in the PoF. These additional resources are important people mentioned in the documents, such as Presidents of Finland, Ministers, or Parliamentary Ombudsmen[14] who have never been elected as MPs and

[9] https://avoindata.eduskunta.fi/#/fi/dbsearch
[10] The person data entries were available in Finnish and Swedish for all MPs, and in English for 202 cases.
[11] https://www.ldf.fi/dataset/semparl
[12] https://doi.org/10.5281/zenodo.7636419
[13] https://a3s.fi/parliamentsampo/actors/csv/
[14] https://www.oikeusasiamies.fi/en/web/guest

were therefore not included in the MP database. These people were collected into a list based on which their biographical information, e.g., occupations, times of activity, party memberships, and times and places of birth and death was extracted from external data sources such as the web pages of State Council of Finland[15], BiographySampo, and Wikidata. At this point, people, parties, and parliamentary groups were given identifiers to be used when annotating plenary session speeches in a separate data transform pipeline for the speech KG S-KG [2].

The data underlying the CSV table was also transformed into RDF instances of people. Based on literal mentions of concepts in the XML data, classes, individuals, and properties for the PoF ontology were created, too. For example, the XML data lists electoral districts of MPs as elements, such as "Uudenmaan vaalipiiri", that were included in the ontology. Furthermore, the instances of electoral districts, parliamentary groups, and committees were further distinguish by their timespan, usually the electoral terms. In RDF this means that all the temporal entities, such as "Uudenmaan vaalipiiri (1999–2019)", are instances of a common superclass, here "Uudenmaan vaalipiiri"—the dataset contains domain specific ontological categories of the parliamentary organizations. In the same way, the political activities of MPs were documented as XML elements including, e.g., committees in which the MP has been a member during a time period. Based on this data, resources with labels for different committees in different times were created in the ontology, and membership events were added in the timeline describing the activities of the MP in different roles. In short, each MP is represented in terms of the political events (s)he has participated in different roles, and in terms of basic biographical information. The person instances, parties, and parliamentary groups of P-KG act as "semantic glue" that link the Speech KG S-KG of ParliamentSampo with the P-KG. As a detail, the XML data for MPs had mentions of parliamentary groups but did not include a person's party, so a datasheet connecting the names parliamentary groups with a corresponding party was used to create the party memberships.

The dataset contained additional information like the name variations, abbreviations, and years of inception and dissolution of a party. This linkage was also used in linking the parliamentary speeches to the correct speaker since the dataset of parliamentary speeches had mentions of the speaker's party. Additional data regarding, e.g., family relations, events of personal biographical history, and photographs, were available from the open data sources of the Government[16], BiographySampo [28, 29], and Wikidata[17]. As an example of the enrichment with addition data, many of the MPs have been members in municipal councils. These resources, e.g., "Helsingin kaupunginvaltuusto" (The City Council of Helsinki) were automatically created based from the text using regular expressions when a recognized name of a Finnish city or municipality was followed by keywords referring to a council or town government. Similarly, the data was enriched with organizations outside of the political scope. For example, companies, schools, and other non-governmental organizations, were extracted from Wikidata when a mention was recognized in a life-time description of a MP. The P-KG data contains in total 2800 instances of the class Person.

From the XML MP data it was possible to extract also other ontological resources for the

[15] https://valtioneuvosto.fi/en/governments-and-ministers
[16] https://valtioneuvosto.fi
[17] https://wikidata.org

PoF Ontology, such as electoral districts, parliamentary groups, and committees mentioned in the data. In addition to the people, the groups mentioned in the XML data elements where extracted, disambiguated, and linked to the corresponding resources in the ontologies used. The groups contain the related parliamentary bodies and committees, governments, electoral districts, and furthermore also groups out of political fields, such as companies, schools, and colleges. Also references to vocations were identified and linked to the resources of the AMMO ontology of historical occupations [30].

As a method for knowledge extraction, patterns of regular expressions were applied to the XML data fields, especially when extracting the person name variations and expressions of time. The source data contained all terms in Finnish. In addition, also the corresponding terms in English (1710) and Swedish (5420) were extracted. In the XML only recent data entries had translations in English. Since the main XML data came from a curated database, entities could be extracted with high precision and recall.

Table 1
Resources

Resource type	Count
Timespan	10733
Label	6115
Place	4543
Person	2828
Publication	1727
Vocation	1456
School, College	670
Parliamentary Group	89
Government	76
Committee	54
Organization	54
Electoral District	46
Party	44
Parliamentary Body	38
Ministry	12
Affiliation Group	10

Table 2
Events

Event type	Count
Career Event	14756
Position of Trust	12788
Committee Membership	6669
Municipal Position of Trust	4740
Event of Education	3722
Birth	2828
Parliamentary Group Membership	2280
Electoral District Candidature	2211
Death	2071
Government Membership	1637
Governmental Position of Trust	1615
Affiliation	1397
Parliament Membership	966
Honourable Mention	537
International Position of Trust	364
Membership Suspension	25

3.3. Data Model for Parliamentary Actors, Groups, and Events

The final data model of the PoF Ontology is presented in Figure 2. It is based on the Bio CRM [31] ontology, an extension of CIDOC CRM[18] for representing biographical information based on role-centric modeling. Bio CRM makes a distinction between attributes, relations, and events,

[18] https://cidoc-crm.org

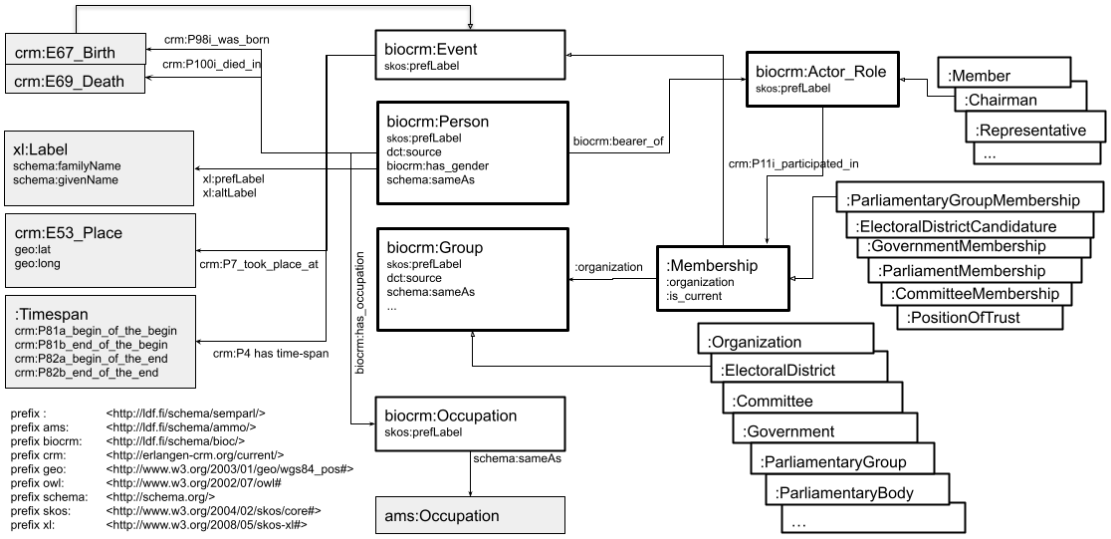

Figure 2: PoF Ontology data model [6] based on Bio CRM

where entities participate in different roles in a qualified manner. The namespaces used in the model are described in the figure on the left.

The key idea of the model is to represent an actor's (class *bioc:Person*) activities as a sequence of events (*bioc:Event*) in places (*crm:E53_Place*) and in time (*:Timespan*) with the actors participating in different roles (*bioc:Actor_Role*), such as *:Member*, *:Representative*, etc.

There are almost 200 different roles in use in the PoF Ontology. The data model has been populated by the MP database and related sources as well as by using a set of external domain ontologies, such as places based on the ontology YSO Places[19], groups and organizations (*bioc:Group*) (harvested from the data), and vocations (*bioc:Occupation*) based on the AMMO ontology.

Table 1 summarizes the number of instances of the main classes of the data model of Figure 2, and Table 2 lists the number of different event types extracted. Table 1 summarizes the number of instances of the main classes of the data model of Figure 2, and Table 2 lists the number of different event types extracted.

3.4. PoF Ontology Available Online

The PoF Ontology with RDF data are available as RDF Turtle files on Zenodo.org [32] using the CC BY 4.0 license[20]. In addition to the RDF files, the central CSV data file people.csv about the MPs and other speakers in the plenary session are available at the CSC Allas store[21].

[19] https://finto.fi/yso-paikat/en/

[20] https://doi.org/10.5281/zenodo.7636420

[21] https://a3s.fi/parliamentsampo/actors/csv/index.html

Furthermore, the linked data is available in the LDF.fi platform[22] as separate graphs interlinked with the pleanary speeches KG in a SPARQL endpoint.

For validating the P-KG data, the data model and its integrity constraints are presented in a machine-processable format using the ShEx Shape Expressions language[23]. We have made initial validation experiments with the PyShEx[24] validator. Based on the experiments, we have identified errors both in the schema and the data. We plan a full-scale ShEx validation phase integrated in the data conversion and publication process to spot and report errors in the dataset.

The data can be downloaded also through the ParliamentSampo portal that includes tools for CSV download, too. In this way the CSV data can be filtered before downloading using the faceted search of the portal https://parlamenttisampo.fi. For example, only people of a certain party during a period of time can be downloaded.

4. Discussion

This paper discussed the challenge of creating ontologies in a data-driven fashion when the domain of discourse is too complex for explicit ontology engineering or there is too little documentation of the domain available. For such cases, data-driven ontology construction was proposed, and as a case study, the event-based PoF ontology describing the work of the Parliament of Finland was discussed. It is likely that similar challenges are encountered when modelling parliaments of other countries, too.

A nice feature of the data-driven approach is that only resources actually used in the data will be considered. On the other hand, the data-driven approach means that if the data misses something, say the membership of an MP in a particular committee at a time, then the list of members in that committee instance is incomplete. It is already known in the case of PoF that the data is not fully complete. For example, the MP database for some old committees record only their chairs, not ordinary memberships. Checking and analyzing possible missing data has not been done systematically afterwards; it is assumed that the database is complete in this sense and that the user is aware about the fact that this may not always be the case. Validation could be done based on historical sources that, e.g., provide lists of members in different committees in different times if such data can be found.

In spite of some limitation in the data, the PoF has been used successfully in DH research and in creating the semantic portal ParliamentSampo in use in Finland with thousands of users. Using the LOD service and portal is discussed in more detail, e.g., in [3, 27]. However, the user has to have data literacy [33] to understand the possible limitations of the data. This requirement is typical in DH research and concerns any system based on the data.

Planned future development of ParliamentSampo includes using and extending the system in parliamentary research studies, correcting the historical data based on user feedback that is collected, e.g., using the portal, validating the data using ShEx shape expressions, and maintaining the data services as part of the national FIN-CLARIAH research infrastructure

[22] https://www.ldf.fi/dataset/semparl
[23] https://shex.io
[24] https://github.com/hsolbrig/PyShEx

program[25].

Acknowledgements Thanks to the numerous participants of the ParliamentSampo project and Ari Apilo, Sari Wilenius, and Päivikki Karhula of PoF for collaborations. Our work was funded by the Academy of Finland in the projects Semantic Parliament[26] and CLARIAH, by CLARIN.eu in the ParlaMint II project[27]. Our work is also related to the EU project InTaVia[28] and the EU COST action Nexus Linguarum[29] on linguistic linked data data resources and analysis. Thanks to Finnish Cultural Foundation for the Eminentia Grant of the first author. The project uses the computing resources of the CSC – IT Center for Science.

References

[1] E. Hyvönen, L. Sinikallio, P. Leskinen, et al., Finnish parliament on the semantic web: Using ParliamentSampo data service and semantic portal for studying political culture and language, in: Digital Parliamentary data in Action (DiPaDA 2022), Workshop at the 6th Digital Humanities in Nordic and Baltic Countries Conference, long paper, CEUR Workshop Proceedings, Vol. 3133, 2022, pp. 69–85.

[2] E. Hyvönen, L. Sinikallio, P. Leskinen, S. Drobac, R. Leal, M. L. Mela, J. Tuominen, H. Poikkimäki, H. Rantala, Plenary speeches of the parliament of finland as linked open data and data services (2023). URL: https://seco.cs.aalto.fi/publications/2023/hyvonen-et-al-ps-text2kg-2023.pdf, forth-coming.

[3] E. Hyvönen, H. Rantala, P. Leskinen, Integrating faceted search with data analytic tools in the user interface of ParliamentSampo – Parliament of Finland on the Semantic Web, in: Proceedings of ESWC 2023, Poster and Demo Papers, Springer, 2023. URL: https://seco.cs.aalto.fi/publications/2023/hyvonen-et-al-ps-eswc-2023.pdf.

[4] M. Hidén, H. Honka-Hallila, Miten eduskunta toimii, Edita Publishing, Helsinki, 2006.

[5] L. Sinikallio, S. Drobac, M. Tamper, R. Leal, M. Koho, J. Tuominen, M. L. Mela, E. Hyvönen, Plenary debates of the Parliament of Finland as linked open data and in Parla-CLARIN markup, in: 3rd Conference on Language, Data and Knowledge, LDK 2021, Schloss Dagstuhl- Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing, 2021, pp. 1–17. URL: https://drops.dagstuhl.de/opus/volltexte/2021/14544/pdf/OASIcs-LDK-2021-8.pdf.

[6] P. Leskinen, E. Hyvönen, J. Tuominen, Members of Parliament in Finland knowledge graph and its linked open data service, in: Further with Knowledge Graphs. Proceedings of the 17th International Conference on Semantic Systems, 6-9 September 2021, Amsterdam, The Netherlands, IOS Press, 2021, pp. 255–269. doi:10.3233/SSW210049.

[7] E. Hyvönen, L. Sinikallio, P. Leskinen, S. Drobac, J. Tuominen, K. Elo, M. La Mela, M. Koho, E. Ikkala, M. Tamper, R. Leal, J. Kesäniemi, Parlamenttisampo: eduskunnan aineistojen linkitetyn avoimen datan palvelu ja sen käyttömahdollisuudet, Informaatiotutkimus 40 (2021). doi:10.23978/inf.107899.

[25] https://seco.cs.aalto.fi/projects/fin-clariah/
[26] https://seco.cs.aalto.fi/projects/semparl/
[27] https://www.clarin.eu/parlamint
[28] https://intavia.eu
[29] https://nexuslinguarum.eu

[8] S. Drobac, L. Sinikallio, E. Hyvönen, An OCR pipeline for transforming parliamentary debates into linked data: Case ParliamentSampo – Parliament of Finland on the semantic web, in: Digital Humanities in the Nordic and Baltic Countries, 7th Conference, 2023. URL: https://seco.cs.aalto.fi/publications/2022/drobac-et-al-ocr-2022.pdf, accepted.

[9] M. La Mela, F. Norén, E. Hyvönen (Eds.), Digital Parliamentary Data in Action (DiPaDA 2022): Introduction, volume 3133, CEUR WS, 2022. URL: http://ceur-ws.org/Vol-3133/paper00.pdf.

[10] D. Fišer, M. Eskevich, J. Lenardič, F. de Jong (Eds.), Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022. URL: https://aclanthology.org/2022.parlaclarin-1.0.

[11] K. Beelen, T. A. Thijm, C. Cochrane, K. Halvemaan, G. Hirst, M. Kimmins, S. Lijbrink, M. Marx, N. Naderi, L. Rheault, R. Polyanovsky, T. Whyte, Digitization of the Canadian parliamentary debates, Canadian Journal of Political Science 50 (2017) 849–864. doi:10.1017/S0008423916001165.

[12] A. Van Aggelen, L. Hollink, M. Kemman, M. Kleppe, H. Beunders, The debates of the European Parliament as Linked Open Data, Semantic Web – Interoperability, Usability, Applicability 8 (2017) 271–281. doi:10.1007/s42001-019-00060-w.

[13] U. Bojārs, R. Darģis, U. Lavrinovičs, P. Paikens, LinkedSaeima: A linked open dataset of Latvia's parliamentary debates, in: Semantic Systems. The Power of AI and Knowledge Graphs. SEMANTiCS 2019, Springer, 2019, pp. 50–56. doi:10.1007/978-3-030-33220-4_4.

[14] R. Bleier, F. Zeilinger, G. Vogeler, From early modern deliberation to the semantic web: Annotating communications in the records of the Imperial Diet of 1576, in: M. La Mela, F. Norén, E. Hyvönen (Eds.), Proceedings of the Digital Parliamentary Data in Action (DiPaDA 2022) Workshop co-located with 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022), volume 3133, CEUR WS, 2022, pp. 86–100. URL: http://ceur-ws.org/Vol-3133/paper06.pdf.

[15] M. Ogrodniczuk, P. Osenova, T. Erjavec, D. Fišer, N. Ljubešic, Çagrı Çöltekin, M. Kopp, K. Meden, ParlaMint II: The show must go on, in: D. Fišer, M. Eskevich, J. Lenardič, F. de Jong (Eds.), Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 1–6. URL: https://aclanthology.org/2022.parlaclarin-1.1.pdf.

[16] A. Pancur, T. Erjavec, The siParl corpus of Slovene parliamentary proceedings, in: Proceedings of the Second ParlaCLARIN Workshop, European Language Resources Association, 2020, pp. 28–34. URL: https://www.aclweb.org/anthology/2020.509parlaclarin-1.6.

[17] S. Elnagar, V. Yoon, M. Thomas, An automatic ontology generation framework with an organizational perspective, in: Proceedings of the 53rd Hawaii International Conference on System Science, Scholar Space, 2020. URL: http://hdl.handle.net/10125/64339.

[18] A. Ko, S. Gillani, A research review and taxonomy development for decision support and business analytics using semantic text mining, International Journal of Information Technology & Decision Making 19 (2023) 97–126.

[19] L. Rajaonarivo, T. Mine, Y. Arakawa, Automatic generation of event ontology from social network and mobile positioning data, in: IEEE/WIC/ACM International Conference on

Web Intelligence and Intelligent Agent Technology, WI-IAT '21, Association for Computing Machinery, New York, NY, USA, 2022, pp. 87–94. URL: https://doi.org/10.1145/3486622.3493933. doi:10.1145/3486622.3493933.

[20] M. La Mela, Tracing the emergence of nordic allemansrätten through digitised parliamentary sources, in: M. Fridlund, M., Oiva, P. Paju (Eds.), Digital histories: Emergent approaches within the new digital history, Helsinki University Press, 2020, pp. 181–197. doi:10.33134/HUP-5-11.

[21] K. Kettunen, M. La Mela, Semantic tagging and the nordic tradition of everyman's rights, Digital Scholarship in the Humanities 37 (2021). doi:10.1093/llc/fqab052.

[22] M. Andrushchenko, K. Sandberg, R. Turunen, J. Marjanen, M. Hatavara, J. Kurunmäki, T. Nummenmaa, M. Hyvärinen, K. Teräs, J. Peltonen, J. Nummenmaa, Using parsed and annotated corpora to analyze parliamentarians' talk in Finland, Journal of the Association for Information Science and Technology 185 (2021) 1–15. doi:10.1002/asi.24500.

[23] S. Simola, A century of partisanship in Finnish political speech, 2020. URL: https://sites.google.com/site/sallasimolaecon/home/research.

[24] E. Lillqvist, I. K. Kavonius, M. Pantzar, "velkakello tikittää": Julkisyhteisöjen velka suomalaisessa mielikuvastossa ja tilastoissa 2000—2020, Kansantaloudellinen Aikakauskirja 116 (2020) 581––607. URL: https://journal.fi/politiikka/article/view/77163.

[25] K. Elo, J. Karimäki, Luonnonsuojelusta ilmastopolitiikkaan: Ympäristöpoliittisen käsitteistön muutos parlamenttipuheessa 1960–2020, Politiikka 63 (2021). doi:10.37452/politiikka.109690.

[26] H. Poikkimäki, P. Leskinen, M. Tamper, E. Hyvönen, Analyses of networks of politicians based on linked data: Case ParliamentSampo – Parliament of Finland on the Semantic Web, in: New Trends in Database and Information Systems, Springer International Publishing, 2022, pp. 585–592. doi:10.1007/978-3-031-15743-1_53.

[27] H. Poikkimäki, P. Leskinen, E. Hyvönen, Applying network and bibliometric analyses to mentions of politicians in plenary speeches: Case ParliamentSampo - Parliament of Finland on the semantic web (2023). Submitted for evaluation.

[28] E. Hyvönen, P. Leskinen, M. Tamper, H. Rantala, E. Ikkala, J. Tuominen, K. Keravuori, BiographySampo – publishing and enriching biographies on the semantic web for digital humanities research, in: The Semantic Web. 16th International Conference, ESWC 2019, Proceedings, Springer, 2019, pp. 574–589. doi:10.1007/978-3-030-21348-0.

[29] M. Tamper, P. Leskinen, E. Hyvönen, R. Valjus, K. Keravuori, Analyzing biography collection historiographically as linked data: Case national biography of finland, Semantic Web – Interoperability, Usability, Applicability 14 (2023) 385–419. doi:10.3233/SW-222887.

[30] M. Koho, L. Gasbarra, J. Tuominen, H. Rantala, I. Jokipii, E. Hyvönen, AMMO Ontology of Finnish Historical Occupations, in: Proceedings of the The First International Workshop on Open Data and Ontologies for Cultural Heritage (ODOCH'19), volume 2375, CEUR Workshop Proceedings, 2019, pp. 91–96. URL: http://ceur-ws.org/Vol-2375/.

[31] J. Tuominen, E. Hyvönen, P. Leskinen, io CRM: A data model for representing biographical data for prosopographical research, in: Proceedings of the Second Conference on Biographical Data in a Digital World 2017 (BD2017), volume 2119, CEUR Workshop Proceedings, 2018, pp. 59–66. URL: http://ceur-ws.org/Vol-2119/paper10.pdf.

[32] L. Sinikallio, P. Leskinen, S. Drobac, M. Tamper, R. Leal, J. Tuominen, H. Rantala, M. La Mela,

M. Koho, E. Hyvönen, ParliamentSampo knowledge graph, 2023. doi:10.5281/zenodo.7636420.
[33] T. Koltay, Data literacy for researchers and data librarians, Journal of Librarianship and Information Science 49 (2017) 3–14. doi:10.1177/0961000615616450.