Integrating Faceted Search with Data Analytic Tools in the User Interface of ParliamentSampo – Parliament of Finland on the Semantic Web

Eero Hyvönen[1,2], Petri Leskinen[2,1], and Heikki Rantala[1]

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

Abstract. This paper presents the idea and implementation of integrating faceted search and browsing with data-analytic tools in the user interface design of the new in-use semantic portal PARLIAMENTSAMPO.

Keywords: Parliamentary Studies, Linked Data, User Interfaces, Portals

1 Parliamentary FAIR Linked Data for Problem Solving

PARLIAMENTSAMPO – *Parliament of Finland on the Semantic Web* [9] is a new Linked Open Data (LOD) service[3] and an in-use semantic portal[4] on top of its SPARQL endpoint. Its main knowledge graph (KG) includes nearly million speeches of the plenary sessions of the Parliament of Finland (PoF) since its foundation in 1907 [13,8], interlinked with another KG and ontology about the 2800 Members of Parliament (MP) and other speakers, parties, and other organizations in PoF [12]. The LOD has been enriched with data from several external sources and by reasoning. Both KGs are published and are available on the Linked Data Finland [5] platform[5], including a SPARQL endpoint and other LOD services, such as content negotiation and RDF browsing. The RDF data is also openly available (CC BY 4.0) at Zenodo.org, and in Parla-CLARIN and CSV form[8]. The CSV data is updated on a daily basis.

The minutes of the plenary speeches of PoF have been available as scanned books (1907–1999), HTML pages (2000–2014), or as XML documents (2015–), but not as Findable, Accessible, Interoperable, and Re-usable FAIR data[6]. If the user knows during which parliament a speech was given, he could download the corresponding document for close-reading. But if one wants, for example, to find out the answers to the following questions, this kind of web service and research method is not a viable solution:

---

[3] LOD service available at: https//ldf.fi/dataset/semparl
[4] Portal available at: https://parlamenttisampo.fi
[5] Linked Data Finland platform: https//ldf.fi
[6] Cf. the FAIR Data initiative: https://www.go-fair.org/

1. **Question**: Who has given most regular speeches and when? **Answer**: Mr. Veikko Vennamo, SMP Party, over 12 600 speeches in 1945–1987 in total. Their distribution can be visualized on a timeline.
2. **Question**: Who was the first to speak about "NATO" in the PoF plenary sessions? **Answer**: Mr. Yrjö Enne, SKDL communist party, May 27, 1959
3. **Question**: What places are most often mentioned by the current MPs of the Swedish People's Party of Finland"? **Answer**: Can be visualized as a table or on heat maps and markers on maps.
4. **How are the MPs of the current Parliament referring to each other in their speeches** : **Answer**: Can be studied using network analysis.

PARLIAMENTSAMPO makes it possible to study parliamentary discussions and networks of politicians easily using methods Digital Humanities (DH), and find answers to questions, such as those above. Furthermore, applications can be created using LOD services, such as the PARLIAMENTSAMPO portal. It is based on the Sampo model [6] and was created using a new declarative version of the Sampo-UI framework [10] for user interface (UI) design.

2 Using the PARLIAMENTSAMPO Portal

Based on the Sampo-UI framework, the landing page of the portal contains *application perspectives* through which instances of the major classes of the underlying KG can be searched [10]. In this case, there are perspectives for speeches and people. Their instances can be searched and browsed using faceted search, and after filtering out a subset of interest 1) its individual instances can be studied by looking at their *instance pages* by browsing or 2) the whole result set can be analyzed as a whole. In both cases, a set of tabs are available for visualizing or analyzing the instance (a speech or person) or a set of them (say the speeches of MPs belonging to a party during a certain time period).

For example, in Fig. 1, the user has selected the Speeches perspective with facets Content, Speaker, Party, (Speech) Type, and others on the left. The search result, i.e., the speeches found, be visualized on the right in five ways by selecting tabs: 1) in tabular form (by default), 2) on a timeline (this tab is selected in the figure), 3) using histograms and pie charts, 4) on a map based on places mentioned in the speeches, or 5) using a heatmap. In the figure, the user has written a query "NATO*" in the Content text facet, the speech type is set to regular speeches, and then 3622 regular speeches that mention the word "NATO" in its various inflectional forms have been filtered into the search result starting from 1959. By clicking on the pie chart visualization button on the Party facet, the distribution of NATO speeches in terms of parties is shown: the most active party with 722 speeches has been the right-wing National Coalition Party Kokoomus.

In Fig. 2 the People perspective is seen, and the user has found all 507 members of the Centre Party, the active supporter of farming in the countryside, by a selection on the Party facet. For analyzing and visualizing the result, seven tabs (e.g., table, pie chart, ages, etc.) can be used. Here the Life chart visualization is
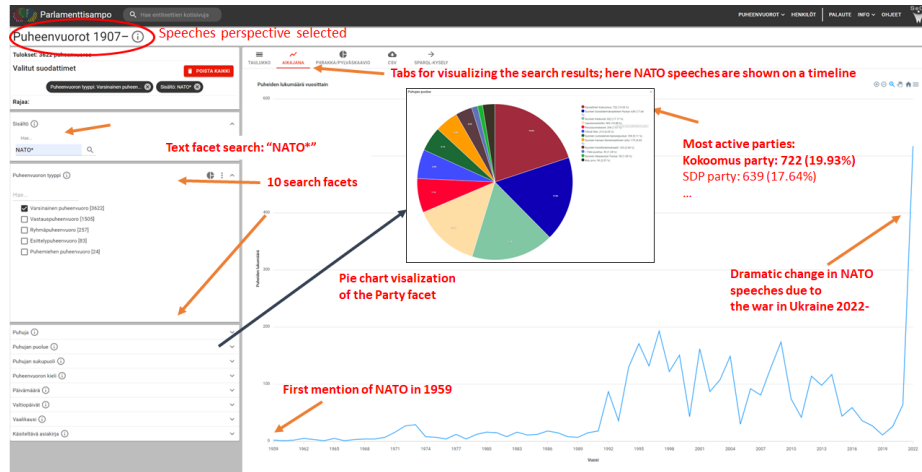
Fig. 1. Using faceted search to filter out and analyze speeches about NATO.

chosen where the MPs are shown as arcs from their places of birth (blue end) to places of death (red end). The width of the arc illustrates the number of MPs related to the arc. By clicking on an arc, links to the instance pages of the corresponding MPs are shown. The graph shows strong movement from the countryside to the capital area of Helsinki.

3 Easy Declarative Implementation Using Sampo-UI

The UI was implemented using a new declarative version of the Sampo-UI framework[7]. Here the UI with its components can be created with little programming by using a set of configuration files in JSON format in three main directories: 1) `configs`. JSON files configuring the portal and its perspectives. 2) `sparql`. SPARQL queries referred to in the `configs` files. 3) `translations`. Translations of things like menu items and labels for different locales. When creating a new UI configuration, existing UI components, such as those for the facets and visualization tabs, can be re-used, and the system can also be extended with new components. Sampo-UI has been found handy to use in practise, and it has been used to create some 15 portals in the Sampo series[8] [6] of LOD systems.

4 Related Works and Contributions

Parliamentary debate datasets have been created from the both historical and contemporary parliaments [11]. The data is typically modelled as documents
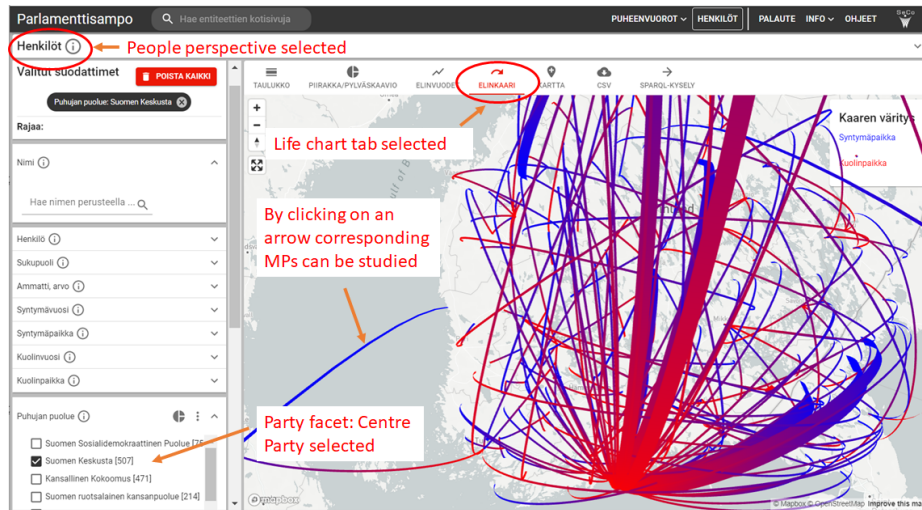

Fig. 2. Using the People perspective of PARLIAMENTSAMPO to study the movement of 507 MPs of the Centre Party from their place of birth to place of death.

using XML-based formats, such as parla-CLARIN[4] and ParlaMint [3]. Linked data has been used before for modelling and publishing speeches of the European Parliament (LinkedEP) [14], the Latvian parliament [2], the Italian Parliament[9], and in the historical Imperial Diet of Regensburg of 1576 project [1]. The contribution of our paper is to demonstrate novel ideas in the UI design for parliamentary DH research: 1) seamless integration of faceted search with data analytic tools for DH problem solving and 2) easy implementation of the UI based on declarative configurations and re-use of existing components. Published on February 14, 2023, the portal was used by some 3000 users in ten days. More information, links, videos, and publications related to PARLIAMENTSAMPO data services and the portal can be found on the project homepage[10]. PARLIAMENTSAMPO is based on and is part of the Finnish Linked Open Data infrastructure LODI4DH[11] [7].

References

1. Bleier, R., Zeilinger, F., Vogeler, G.: From early modern deliberation to the semantic web: Annotating communications in the records of the Imperial Diet of 1576. In:

---

[9] http://data.camera.it

[10] https://seco.cs.aalto.fi/projects/semparl/en/

[11] https://seco.cs.aalto.fi/projects/lodi4dh/

Proceedings of the Digital Parliamentary Data in Action (DiPaDA 2022). vol. 3133, pp. 86–100. CEUR Workshop Proceedings (2022)

2. Bojārs, U., Darģis, R., Lavrinovičs, U., Paikens, P.: LinkedSaeima: A linked open dataset of Latvia's parliamentary debates. In: Semantic Systems. The Power of AI and Knowledge Graphs. SEMANTiCS 2019. pp. 50–56. Springer (2019)

3. Erjavec, T., Ogrodniczuk, M., Osenova, P., et al.: The ParlaMint corpora of parliamentary proceedings. Lang Resources & Evaluation (2022), https://doi.org/10.1007/s10579-021-09574-0

4. Fišer, D., Eskevich, M., Lenardič, J., de Jong, F. (eds.): ParlaMint II: The Show Must Go On. European Language Resources Association, Marseille, France (June 2022), https://aclanthology.org/2022.parlaclarin-1.1.pdf

5. 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: The Semantic Web: ESWC 2014 Satellite Events, Revised Selected Papers. pp. 226–230. Springer-Verlag (2014), https://doi.org/10.1007/978-3-319-11955-7_24

6. Hyvönen, E.: Digital humanities on the Semantic Web: Sampo model and portal series. Semantic Web – Interoperability, Usability, Applicability (2022), accepted, https://seco.cs.aalto.fi/publications/2021/hyvonen-sampo-model-2021.pdf

7. Hyvönen, E.: How to create a national cross-domain ontology and linked data infrastructure and use it on the semantic web. Semantic Web – Interoperability, Usability, Applicability (2022), https://seco.cs.aalto.fi/publications/2022/hyvonen-infra-2022.pdf, under review

8. Hyvönen, E., Leskinen, P., Sinikallio, L., Drobac, S., Leal, R., Mela, M.L., Tuominen, J., Poikkimäki, H., Rantala, H.: ParliamentSampo infrastructure for publishing the plenary speeches and networks of politicians of the Parliament of Finland as open data services. In: Proceeding: International Workshop on Knowledge Graph Generation from Text (TEXT2KG), co-located with ESWC 2023 conference. CEUR Workshop Proceedings (2023), https://seco.cs.aalto.fi/publications/2023/hyvonen-et-al-ps-data-2023.pdf, forth-coming

9. Hyvönen, E., Sinikallio, L., Leskinen, P., 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. pp. 69–85. CEUR Workshop Proceedings, Vol. 3133 (May 2022)

10. 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 (2022)

11. La Mela, M., Norén, F., Hyvönen, E. (eds.): Digital Parliamentary Data in Action (DiPaDA 2022): Introduction, vol. 3133. CEUR WS (March 2022)

12. Leskinen, P., Hyvönen, E., Tuominen, J.: 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. pp. 255–269 (2021)

13. Sinikallio, L., Drobac, S., Tamper, M., Leal, R., Koho, M., Tuominen, J., Mela, M.L., Hyvönen, E.: 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. pp. 1–17. Schloss Dagstuhl- Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing (August 2021)

14. Van Aggelen, A., Hollink, L., Kemman, M., Kleppe, M., Beunders, H.: The debates of the European Parliament as Linked Open Data. Semantic Web – Interoperability, Usability, Applicability **8**(2), 271–281 (2017)