Plenary Debates of the Parliament of Finland as Linked Open Data and in Parla-CLARIN Markup

Laura Sinikallio[1], **Senka Drobac**[2], **Minna Tamper**[2], **Rafael Leal**[1], **Mikko Koho**[1], **Jouni Tuominen**[1,2], **Matti La Mela**[1] & **Eero Hyvönen**[1,2]

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

Problems

- All Finnish plenary session minutes 1907–2021 are publicly available but in different formats and locations, and not as *data*
- Searching, browsing, and studying the minutes is challenging
- Data is needed for Digital Humanities research and applications
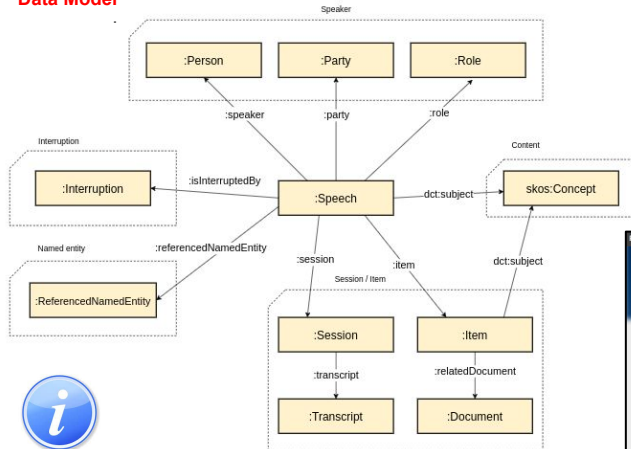- Linked Data is needed for enriching the data with related datasets

Solution

- Gathering and unifying all debates as two complete datasets: RDF knowledge graph and Parla-CLARIN XML
- Enriching and linking the speeches to different ontologies, such as Members of the Parliament Ontology and online resources
- Offering the linked data as a service for open use on the Linked Data Finland platform https://ldf.fi
- Creating a semantic portal for researching the data without programming skills – ParliamentSampo
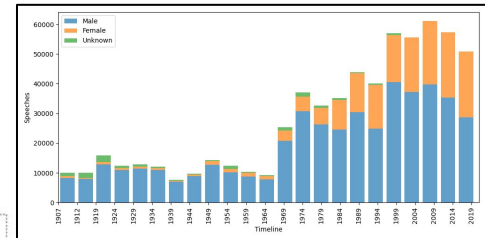
Results

- Over 900 000 individual speeches are represented in RDF and Parla-CLARIN XML formats.
- The data transformation process has created usable results and enabled new opportunities for researching the parliamentary culture and language.
- The whole history 1907–2021 of Parliament of Finland is covered.
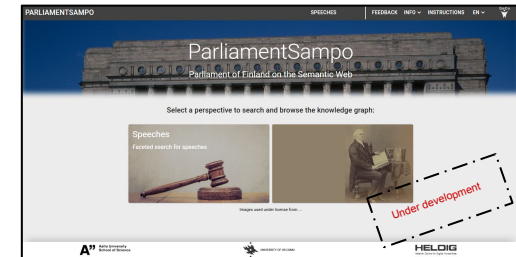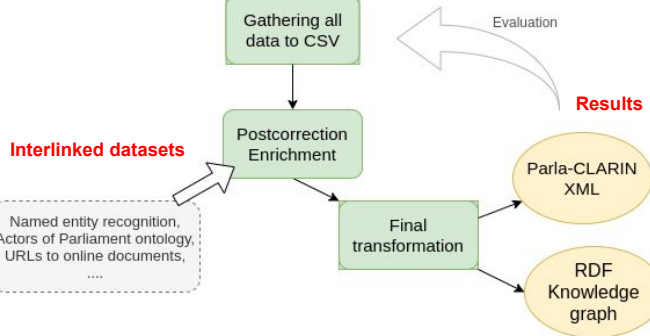- The approach is arguably applicable in other countries, too.

Data Aggregation and Transformation Process

Interlinked datasets

Data Model

Data Analysis: Using SPARQL

Applications: ParliamentSampo

More Info: https://seco.cs.aalto.fi/projects/semparl/en/