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

Laura Sinikallio¹, Senka Drobac², Minna Tamper¹, Rafael Leal¹, Mikko Koho¹, Jouni Tuominen¹², Matti La Mela¹ & Eero Hyvönen¹²

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

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.

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


Data Aggregation and Transformation Process

Data Model

Applications: ParliamentSampo

Under development

Data Aggregation and Transformation Process

Interlinked datasets

Final transformation

RDF Knowledge graph

Parla-CLARIN XML

Postcorrection Enrichment

Gathering all data to CSV

Evaluation

Images

HTML

XML

OCR

Original minutes

Data Analysis: Using SPARQL

Applications: ParliamentSampo

Under development