

Talk of Europe - Linking European Parliament Proceedings

Max Kemman and Astrid van Aggelen

Political institutions increasingly open up their parliamentary proceedings to the general public. This results in an abundance of data that are useful in disciplines such as political science, social science, and history. However, this information is generally published in non-machine readable formats such as HTML or PDF, restricting research activities to manual analyses on datasets of limited size.

In recent years, efforts have been made to curate parliamentary debates as structured data (Gielissen & Marx, 2009) and represent them as linked data (Juric, Hollink & Houben, 2012). By using the machine-readable linked data format Resource Description Framework (RDF) to describe metadata items like politicians, parties and speeches, advanced information needs involving combinations of and relations between these items become resolvable. The PoliMedia project, which resulted in a unified search system of topics covered in the national parliament, radio broadcasts and newspapers (Kleppe et al., 2014), demonstrated that linked political data allow for interesting applications.

The current project, Talk of Europe (www.talkofeurope.eu), builds upon PoliMedia, broadening its scope to the European Union. This project has two ambitions. The first is to curate all data on the plenary debates of the European Parliament from 1996 onwards, available in 21 languages, and convert these to RDF. To ensure wide applicability in political research, the data are enriched with biographical and political information on the speakers, building upon the work of Høyland, Sircar & Hix (2009). The second aim is to showcase the potential of this interlingual, enriched and interlinked political dataset in research and in application development and to stimulate pan-European collaboration. To this end, developers and scientists of different countries and disciplines are brought together to explore possible links with other datasets and to create proof-of-concept applications.

The outcomes of this project will facilitate, among others, research on the influence of the European Union on national politics and media. By connecting the European parliamentary data with that of PoliMedia, it becomes possible to investigate the relation between topics covered in the European and Dutch parliaments and in Dutch media. Since the European parliamentary data is available in 21 languages, such investigations can be performed and compared across all nations of the European Union.

The presentation will present the data model and its embedding in existing Linked Open Datasets, showing the current semantic framework and inviting peers to provide feedback. Furthermore, workshop visitors will be enabled to interact with the dataset through a SPARQL endpoint, an RDF-based search application.