JOHANNES KEPLER UNIVERSITY LINZ | JKU

TNF

Faculty of Engineering and Natural Sciences

Understanding the Community Structure of Coalition and Opposition in Parliamentary Democracy on the Example of the Austrian Parliament

BACHELOR'S THESIS

submitted in partial fulfillment of the requirements
for the academic degree

Bachelor of Science

in the Bachelors Program

Computer Science

Submitted By:
Markus Hiesmair

At the:
Institut für Telekooperation

Advisor:
Gabriele Anderst-Kotsis

Linz, 20. November, 2015

Affidavit

I hereby declare that the following dissertation ”Understanding the Community Structure of Coalition and Opposition in Parliamentary Democracy on the Example of the Austrian Parliament” has been written only by the undersigned and without any assistance from third parties.

Furthermore, I confirm that no sources have been used in the preparation of this thesis other than those indicated in the thesis itself.

Linz, on December 2, 2015

Markus Hiesmair

Acknowledgment

Summary

Summary . . .

Abstract

Abstract …

Contents

Abbreviations

ÖVP Austrian People Party (german: Österreichische Volkspartei)

SPÖ Social Democratic Party of Austrian (german: Sozialdemokratische Partei Österreichs)

List of Figures

List of Tables

Chapter 1

Introduction

One of the most crucial requirements of a democracy is transparency. There are several ways how one can gain information about the current and past political activities in Austria. One of the best possibilities among them are the publicly available protocols of the national council sessions. In these protocols every word said in a session is written down and that makes up the corresponding protocol. Unfortunately, these protocols are very long and it is hard to gain meaning out of it, because of its plain and simple structure and the great amount of data.

To be able to analyze and visualize the activities and relations of the politicians and parties in a better way, and to make the structure of the political system accessible to a broader audience, analysis tools are needed. This thesis documents the methods that can be used to perform automated analysis over the available data. The protocols are being extracted, transformed, analyzed and visualized.

1.1 Research Goals

The protocols are currently available in semi-structured form - through HTML files.[1] To be able to properly persist and analyze the data, the protocols have to be transformed into a fully structured form (e.g. Java Objects). The following elements will be extracted:

- Legislative periods and their sessions
- Politicians and their mandates

[1]Until the 19. legislative period, the protocols are only available in PDF-format. These protocols cannot be extracted with reasonable quality. Therefore they won't be used in the further work.

- Parliament clubs
- Discussions and speeches during the sessions

As soon as this is done, the extracted data can be persisted into an arbitrary relational database. Furthermore, some general and network analysis should be done on the data. In the following list some interesting applications on top of the extracted data are presented:

- Create a network graph which shows the relations among politicians and parliament clubs.
- Find groups of politicians (or parliament clubs) with the same attitudes.
- Analyze how homogeneous the attitudes of politicians of the same parliament club are.
- Find the politicians which take part in the most discussions.
- Find the most absent national council members.

In the final step the results should be visualized via a web application. The focus hereby lies in making the results as easy to understand as possible.

1.2 Austrian Parliament

The analysis approaches in this thesis can basically be used for every parliament or other similar political structure, given that data is available in a similar form. As this work is done at an Austrian university and the protocols of the national council are available as open data, the show case is built up on the Austrian parliament.

The Austrian parliament basically consists of two chambers, the national council and the federal council. The national council is elected through federal elections, whereas the federal council consists of delegates of the 9 provinces. Both chambers have different responsibilities and functions, and their goal is to ensure that the decisions are in the best interest for the Austrian people [3].

1.2.1 National Council

The national council consists of 183 members, which can band together to form so called parliamentary clubs. Usually for each political party, which got elected in the national council, there is one parliamentary club, but that is no necessity. The tasks of the national council include law-making, controlling the government, seeking solutions for current problems, determining the budget and much more [3].

After every election one or more parliamentary clubs have to build the government. These parliament clubs have to have in total at least 92 mandates (more than half of the overall mandates of 183). Politicians of the government are then selected to be the federal ministers. All other parliamentary clubs, which are not in the government are in the so called opposition. In most cases the government can make laws without the opposition, but in some special cases there is $\frac{2}{3}$ majority needed [3].

1.2.2 Federal Council

The federal council consists of 61 members. As the members are delegates of the provinces, their main duty is to represent their province and make sure the politics in the parliament are in the interest of the province they represent. To do so, they can raise objections against legislation of the national council, but most of the time the federal council only has the power to delay legislation and not to prevent it [3].

1.2.3 Analysis Scope

In this work only the data of the national council will be analyzed because there are no openly available data sources which could be used to include the federal council in the analysis. Furthermore, the national council has a lot more responsibilities and is of greater importance for the overall democratic process in Austria.

Chapter 2

Related Work

In the context of computer science, there are only a few works on automatically analyzing political structures such as a parliament. In 2013, Renzo Lucioni [2] used publicly available voting data from the Congress of the United States of America to analyze the relationships among politicians and how distinct the two main parties, the Democrats and the Republicans, are. To achieve this, he used data from the 101^{st} Congress through the 113^{th} Congress and created network graphs which graphically showed which politicians vote similar. He also showed how the structure of the Congress developed over time by creating several graphs of the years 1989 to 2013. His results showed that the gap between the Republicans and the Democrats became larger and larger over the last decades. This means that both parties vote more and more against each other. In the context of the Austrian parliament, similar analysis can be applied, if data is available in sufficient quality. For example, it can be analyzed which parties vote similar and if there exist relations between parties which are in the coalition and parties which are in the opposition. You can find the results for the Austrian parliament in section 4.1.

An earlier work was done by Porter and Newman in 2005 [4]. They wrote a paper on network analysis of committees in the U.S. House of Representatives and tried to show the connections between representatives of the House and the committees and subcommittees. There happens a big amount of the American legislation in these committees and especially the assignment of politicians and the change of it over time are interesting subjects of analysis. In their work Porter and Newman gain information without specific knowledge of the structure of the committees, using technologies of network analysis. In particular, they tried to find communities and their connections within the network of the committees to get information about strategic assignment of politicians in important committees. Furthermore, Porter and Newman used single-linkage clustering to get clusters of communities and their connections and also visualized that with a dendrogram representing the hierarchical structure of the committees and subcommittees. Similar analysis would also be interesting for the Austrian parliament, but community detection and clustering are not included within the scope of this thesis. In the second part of Porter and Newman's paper, they also have done some analysis on the relations among politicians in the House of Representatives. The results show the most left, most right and most partisan politicians in the House. This is especially interesting because it shows that all the most left politicians are Democrats and all most right politicians are Republicans, which shows that also the House of Representatives is divided at a high degree in Democrats and Republicans.

In 2012, Amelio [1] also did a study on the voting behavior in the Italian parliament. One part of her study was analyzing party cohesion (how homogeneous all politicians of a specific party voted in the selected periods). An interesting result was that the cohesion of the parties in the opposition increased over time whereas the cohesion of the governing parties decreased and after the analyzed period the government was not reelected. Another measure taken was the parliamentarian similarity. This measure compares the voting behaviors of two parliamentarians and gives a result on how similar they voted. Based on the values obtained, Amelio did hierarchical clustering using single-linked clustering to find communities within the parliament and visualized the results in a dendrogram, similar to the result of Porter and Newman [4].

All three papers, which were discussed in this section, show that through automatic analysis of political structures, information on the structure and clustering of political systems can be gained. Furthermore, through visualizations in graphs the information can be presented in a way everybody understands it easily and therefore the visualizations can be used to improve the general understanding of political systems and the current structures of parties and politicians.

Chapter 3

Design and Implementation

In this chapter the design and implementation of the prototype for the Austrian parliament are described. First of all, in Section 3.1 the overall architecture and the different components are being discussed. The more detailed description of the implementation is divided into five sections: Section 3.2 describes the data extraction of the HTML-files, section 3.3 describes the transformation into a structured form, section 3.4 discusses the export to a relational database and the sections 3.5 and 3.6 describe the analysis and visualization of the given data.

3.1 Architecture

Figure 3.1 shows the general architecture of the prototype which was implemented. The ETL-Application brings the data from the protocols in the database whereas the web server application visualizes the results and shows statistics and graphs for the given data. The ETL-Application is implemented using the ETL pattern. This means that there are three distinct steps: Extract - Transform - Load. First the application reads an RSS feed which contains all the protocols for one legislative period and the politician profiles (Extract). The retrieved HTML-files get parsed and are transformed into Java objects (Transform) which get loaded into a relational database[1] (Load). To visualize the then available data, the analysis engine queries the database, preforms analysis on it and converts the data in a form which can be displayed (e.g. a graph structure). Furthermore, this data is made available via RESTful web services. The Polymer web application accesses these web services and shows graphs and statistics. All the components will be described in more detailed in the following sections.

[1]in the prototype, a PostgreSQL database was used

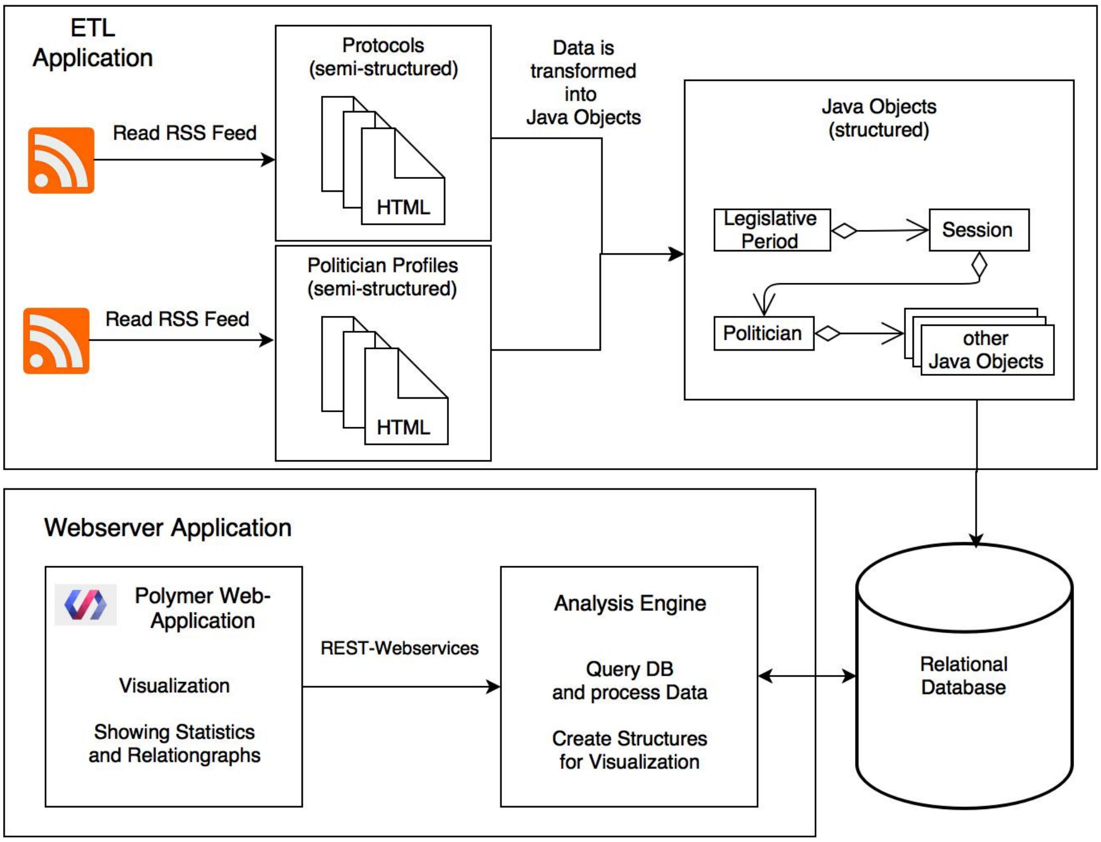

Figure 3.1: General Architecture

3.2 Data Extraction

The first step which has to be done in the ETL-Application is the extraction. The data which should be transformed has to be collected and stored. In our case the data is contained in the stenographic protocols of the national council and in politician profiles. Both the protocols and politician profiles are publicly available and can be found at the website of the Austrian parliament (See `https://www.parlament.gv.at/PAKT/STPROT/` and `https://www.parlament.gv.at/WWER/PARL/`). The protocols are available in PDF-format and since the 20^{th} legislative period also in HTML. As the transforming of the PDF-files would not result in sufficient quality, in this thesis only the HTML-files (the data since the 20^{th} legislative period) are being extracted and analyzed.

To collect the needed files automatically, RSS feeds are used. There are feeds available for both, stenographic protocols and politician profiles. The HTML-files which are linked in the feeds are being downloaded and stored on the file system before they are transformed.

3.3 Transformation

The next step is the transformation of the HTML-files into Java objects (transformation of semi-/unstructured data to structured data). To be able to extract the desired information out of the HTML-files, the structure of the stenographic protocols and politician profiles was investigated. Then undesired HTML-structures were removed, because they avoided a correct information extraction. For example, the page breaks and page headers in the full text protocols were removed. Finally, the tag structure of the HTML documents and regular expressions were used to find the required data. This was not always easy, as the format of the protocols changed over time. For example, in the full text protocols since the 21^{st} legislative period, the politicians were referenced with a link to their profile, whereas in the protocols of the 20^{th} period, there were only politician names. This made it much harder to find the right politicians as the full list of politicians had to be searched for a politician with the same name and title(s). To transform the HTML-files with good quality, for the protocols of the 20^{th} period, there was an extra transformation class which could find the politicians without the link to their profiles.

3.3.1 Transformation of the Politician Profiles

For each politician who is sitting in the national or federal council, there exists a politician profile. In the first part of the transformation, all politician profiles are being transformed into a list of politician Java objects. The name (and if provided previous names), the titles, the birth date, and the political mandates are being extracted from the HTML code. Mandates are some kind of political functions like the membership in the national or federal council or the period where the politician was a federal minister. They are important especially because they include the club memberships of the politician within a specific period of time. Figure 3.2 shows the class diagram of the resulting Java objects, extracted out of the profiles.

3.3.2 Transformation of the Protocols

In the second part of the transformation, the protocols of the sessions of the national council are transformed. For each session, there exist two files: The full text protocol and the session summary. In the full text protocol there is every word which is being said in the session written down, whereas in the session summary, there is only important information like discussions and speeches summarized. Information that gets extracted out of the protocols includes sessions of legislative periods, chair men in the sessions,

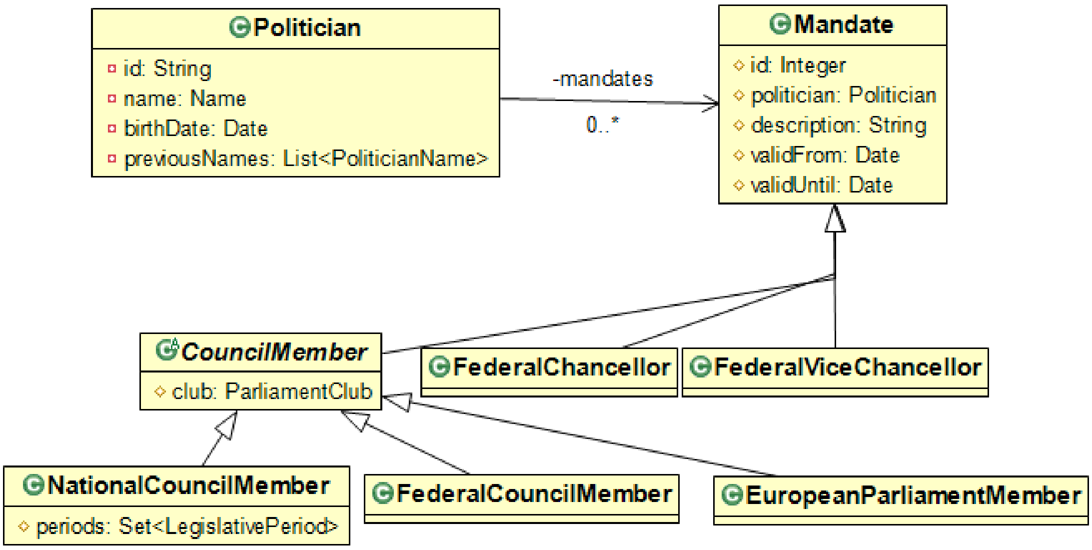

Figure 3.2: Class Diagram of the result of the Profile Transformation Step

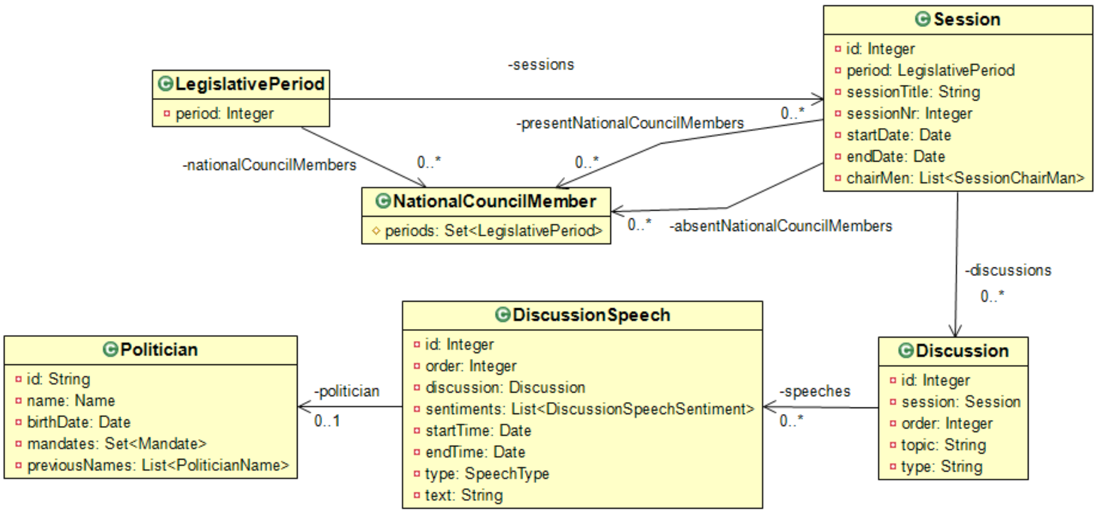

Figure 3.3: Class Diagram of the result of the Protocol Transformation Step

politician absences and presences, discussions and speeches of politicians. The list of politicians which was the result of the first transformation part is used to find the politicians in the protocols and to reference them in speeches. Figure 3.3 shows the class diagram of the resulting objects of the protocol transformation step.

3.4 Export into a Database

The export to a database is the loading part of the ETL-application. The in the previous step created Java objects are being persisted into a relational database. To stay independent of specific databases the Java Persistence API and Hibernate are used.

Name	Description
Overall Absence per Period	A percentage of the absence of all national council members in one legislative period.
Absence of a Politician per Period	A percentage of the absence of one national council member in one legislative period.
Absence of a Parliamentary club per Period	A percentage of the absence of all members of one parliamentary club in one legislative period.
Count of Speeches of politicians per Period	The count of speeches a politician held in the national council during one legislative period.

Table 3.1: Simple Analysis Measures

Using OR-mapping brings the advantages that no SQL-statements have to be written and changes on the tables/objects are easily made using Java annotations.

3.5 Analysis

The first step of the analysis phase was to determine which analysis could be applied on the given data. The results can be classified into two categories: simple analysis and network analysis. The simple analysis takes just simple measures like how many speeches a politician held during a legislative period or how often was he absent. These measures are taken just via SQL queries on the database built up by the ETL Application. Table 3.1 shows all simple analysis measures taken. The network analysis will be discussed in the following section.

3.5.1 Network Analysis

All the speeches held in the Austrian parliament are tagged with a annotation. A majority of them are tagged with either pro or con, speeches with other annotations are less interesting in the context of this thesis and won't be considered in the following analysis. The pro-con annotations of the speeches give the possibility to create a relation graph of all the politicians (and parliamentary clubs) based on their position (pro or con) on the topics of the discussions they held speeches in. Such a graph shows visually how related two politicians are. If two politicians have a strong positive relationship (if they have the same attitudes in the discussions) they will be displayed close together, but if they have a strong negative relationship (mainly contrary attitudes), they will be displayed far away from each other.

To be able to display such a graph the strength of a relationship (the link weight w_{p_1,p_2}) between two politicians p_1 and p_2 has to be calculated. This is done via the following formula:

$$w_{p_1,p_2} = \frac{\sum_{i=0}^{D} w_{d_i,p_1,p_2}}{\sum_{i=0}^{D} |w_{d_i,p_1,p_2}|}$$

where D is the total number of discussions and d_i is the i^{th} discussion. w_{p_1,p_2} is a normalized number between $+1$ and -1, where $+1$ means that the two politicians had the same opinion on every topic and -1 means that they have totally contrary opinions. w_{d_i,p_1,p_2} is the weight of the politicians p_1 and p_2 in the discussion d_i and can be described by the following formula:

$$w_{d_i,p_1,p_2} = \begin{cases} +1 & \text{if } p_1 \text{ and } p_2 \text{ have the same attitude in } d_i \\ -1 & \text{if } p_1 \text{ and } p_2 \text{ have contrary attitudes in } d_i \\ 0 & \text{if } p_1 \text{ or } p_2 \text{ did not speak in } d_i \end{cases}$$

The result of this formula is a normalized number between $+1$ and -1, where $+1$ means that the two politicians had the same opinion on every topic and -1 means that they have totally contrary opinions.

To be able to create such a graph, first the data has to be brought in a form in which it can be queried with little effort. To have the data in the necessary form, immediately after the loading of the ETL-Application the discussion speeches get analyzed and the relations of the politicians were pre-processed. The algorithm iterates over all discussions and investigates the discussion speeches. For each politician pair, which takes part in the discussion, the relation is computed. If they have the same position on the topic of the discussion (either if both are pro or both are con), the weight is $+1$, else the weight is -1. So, for every discussion there are $\frac{N*(N-1)}{2}$ relations computed. If the relation of two politicians should be computed, the following formula is applied:

$$w_{p_1,p_2} = \frac{\sum_{i=0}^{D} w_{d_i,p_1,p_2}}{\sum_{i=0}^{D} |w_{d_i,p_1,p_2}|}$$

where D is the total number of discussions, d_i is the i^{th} discussion and w_{d_i,p_1,p_2} is the weight of the politicians p_1 and p_2 in the discussion d_i. If not both politicians spoke in the discussion d_i the weight is zero. The result of this formula is a normalized number between $+1$ and -1, where $+1$ means that the two politicians had the same opinion on every topic and -1 means that they have totally contrary opinions.

To create a graph, like Lucioni [2] did in his study on the Congress of the United States of America, the relations of every politician pair in the parliament have to be calculated. This was done separately for every period, which was analyzed, so that all periods itself could be analyzed. In such a graph the nodes are the politicians p_i which held speeches in the corresponding period. Two nodes p_i and p_j are connected via a link if both of them spoke at least once in the same discussion.

3.6 Visualization

Chapter 4

Results and Discussion

4.1 Relations of Parliament Clubs

Graph + Explanations

4.2 Relations of Politicians

Graph + Explanations

Chapter 5

Conclusions and Future Work

Bibliography

[1] A. Amelio and C. Pizzuti. Analyzing voting behavior in italian parliament: Group cohesion and evolution. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, pages 140–146, Aug 2012.

[2] Renzo Lucioni. *Senate Voting Relationships*, 2013. `http://www.renzolucioni.com/senate-voting-relationships/`.

[3] Austrian Parliament. *The Austrian Parliament*, 2015. `http://www.parlament.gv.at/ENGL/PERK/PARL/`.

[4] Mason A. Porter, M. E. J. Newman, Peter J. Mucha, and Casey M. Warmbrand. A network analysis of committees in the u.s. house of representatives. *PNAS*, Jan 2005.