ETH zürich

Analyzing 130 Years of the Swiss Parliament*

Dr. Laurence Brandenberger, *Chair of Systems Design, ETH Zürich*
Collaborators: Prof. Frank Schweitzer, Sophia Schlosser; SDSC†: Luis Salamanca, Fernando Perez Cruz

1 Project Aims

1. **Create searchable data base for 130 years of Swiss Parliament**
   - Includes all speeches, bills and votes
   - Tracking changes across chambers
   - Cosponsorship networks
   - Committee memberships/activities
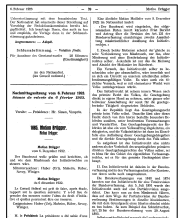2. **Analyze development of Swiss parliament**
   - How do institutional changes affect parliamentary cooperation?
   - How and why do members of parliament change their positions?
   - How does issue ownership change over time?

2 Data

- **Proceedings** of Swiss parliament (both chambers)
- **PDF documents**, courtesy of the Federal Archives
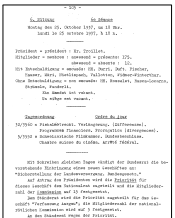
Amtliches Bulletin
- ≈ 50,000 documents

Summary of proceedings
- ≈ 500 documents

Additional protocols
- 1921-70, ≈ 30,000 doc.

- For the data from 1995-today we rely on the **online database** by the parliament and restructure it to fit the extracted data from 1891-1995
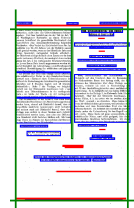
3 Pre-Processing Steps

- Pre-processing PDF scans
- Cleaning up xml-files and enriching them with structural information
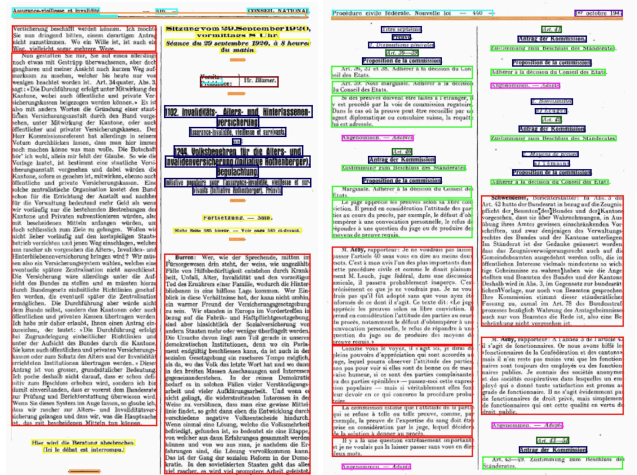
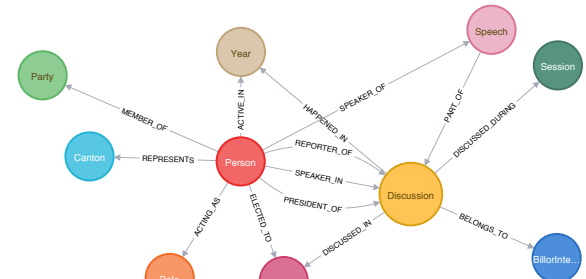identifying margins   separating columns   ordering text   enriched XML file

4 Extraction Tool

- We developed a **new tool** to classify text lines in PDFs
- Semi-supervised approach
- Distinguish between titles (dark blue), speeches (red), votes (pink), bill proposals (green), etc.

5 Building A Relational Database

- We link bills to originators, proposed amendments and speeches
- We **track bill changes** and voting outcomes
- We **group bills by issues** (using topic modelling)
- We link mentions of bills and people in speeches using Named Entity Recognition
- **Graph database:** neo4j
- Contains **entities**, **attributes**, **relations**

6 Contact and Further Information

* FNS NF Swiss National Science Foundation   † SDSC