ETH zürich

The data

The corpus is based on retro-digitised sources such as correspondence, reports, lists, lectures and minutes from the Federal Archives in Bern, as well as scientific and technical publications and internal (administrative) publications.

Digital research tool for historians

Archive holdings are becoming more and more digital (digitised or digitally born archive material). For the exploration of large digital repositories, the preparatory work of other disciplines can be used. In our case, we have chosen topic modelling.

Topic modelling

A probabilistic method (LDA) is used to identify semantic fields (topics) in the corpus. A list of constitutive words is generated for each field and the most probable topics for each document are determined. These semantic fields partly reproduce the institutional and archival transmission, but also show topics and subjects that span these structures.

topic #50

Opportunities

- The listing of cross-structural topics enables different approaches and opens up new perspectives on our source material. It supports the emancipation from the existing registry hierarchy. They enable access to the source material across the hierarchy.
- The search via topics makes it possible to refine and supplement the simple full text search. The sorting of the results (relevance) can thus be controlled.
- The search strategies learned with search engines such as Google can be applied to archive material.
- In order to make semantic shifts visible over time, we have extracted time data from the documents and carried out topic modelling on a year-by-year basis. In addition to the boom of certain terms, the restructuring of the federal administration also becomes visible.

topic #21

Challenges

- Innovative exploration strategies go hand in hand with additional competencies
- Since the result display refers to individual documents, the result (the document) is outside its context of origin, filing and collection. The indexing of the historical and archival knowledge space can be supported by digital browsing (back and forth in dossiers and folders).

Topic modelling & explorative search

Within the scope of the project "Trading zones. Computer und Swiss Federal Administration, 1960 - 1990", the Chair of History of Technology at ETH Zurich, in collaboration with ETH Scientific IT Services, is developing tools (parameterizable full-text search & topic modelling) to study a constantly growing body of sources.

Moritz Mähr, Henrike Hoffmann & Daniela Zetti (History of Technology)

D GESS