Normative texts in the City-State of Bern (1528-1795). Testing a Simple Knowledge Organisation System (SKOS) and Automatic Meta Data on a Handwritten Corpus.

Normative texts in the City-State of Bern (1528-1795).

Romein, C. Annemieke

Annemieke.Romein@Huygens.knaw.nl
KNAW Huygens Institute for the History and Culture of the Netherlands, Amsterdam; Vrije Universiteit Amsterdam

Veldhoen, Sara

Sara.Veldhoen@KB.nl
KB National Library of the Netherlands

Early modern governments created, announced and published many norms (legislation) over time. In the 1990s Michael Stolleis and Karl Härter initiated a project at the (then) Max-Planck-Institute for European Legal History (now the Max Planck Institute for Legal History and Legal Theory) that grew out to provide an overview of normative texts from 68 jurisdictions within Central Europe. Their efforts resulted in a series: Repertorium der Policeyordnungen der Frühen Neuzeit and a database. These contain data on approximately 200,000 normative texts. The hierarchical structure of systematically organized categories allowed the creation of an overview that enables (future) researchers to compare norms from one jurisdiction to another. The hierarchical structure is divided over four layers: going from very generic to very specific.

One of the studied jurisdictions within this project was the 'city-state' of Berne. Here 4932 normative texts (Mandaten) were published from 1528 until 1795 and written down in the Book of Mandates. These norms concerned various topics ranging from religious affairs to infrastructure, safety and education. The initial overview and labelling of the texts of Berne was created by Dr. Claudia Schott-Volm (2006). Her extensive overview not only includes information on the normative texts themselves but also includes the jurisdictions. Though Berne was officially a 'city-state', its lands covered most of the modern-day canton of Berne, Aargau, Vaud and some parts of some other modern-day cantons.

Using the original labels as benchmarked data, the part of the project presented here is to test whether these legal texts could be automatically labelled through Artificial Intelligence. Legal texts are, in themselves, phrased very straightforwardly and include all possible details – as confusion needs to be avoided at all costs. Using Annif which is created by the National Library of Finland to automatically categorise books; I will test the applicability to early modern handwritten texts. As Annif has various backends (e.g. TF-IDF; Omikujji, and fastText) I will test which one works best for my normative texts.

While the initial project relied on manual labelling, this poster will present some first results on a digital replication of the results. The following steps are taken to train computer models to automatically label the normative texts:

Digitizing the *Mandates* of the entire period.

Making transcriptions and creating Handwritten Text Recognition (Transkribus).

Creating a Simple Knowledge Organisation System

Reconciling the labels of the normative texts with the SKOS.

Training an AI model within Annif and testing this on the Mandaten (10-fold cross-validation with a 90/10 train/test split).

As the multilayered SKOS contains 1800 categories divided over four different levels, it may be expected that Annif performs less accurately on the lowest – most specific – topics as it may lack training data. The aim of this entire exercise is not only to test the multi-layeredness of the SKOS and the applicability of this through Annif; but (hopefully) come up with reasonable estimates how much training data is needed to enable AI/ Annif to automatically label legal texts all by itself (in the future).

Bibliography

Schott-Volm, Claudia (2006): *Repertorium der Policeyordnungen der Frühen Neuzeit (#7). Orte der Schweizer Eidgenossenschaft: Bern und Zürich* . Frankfurt am Main: Klostermann .