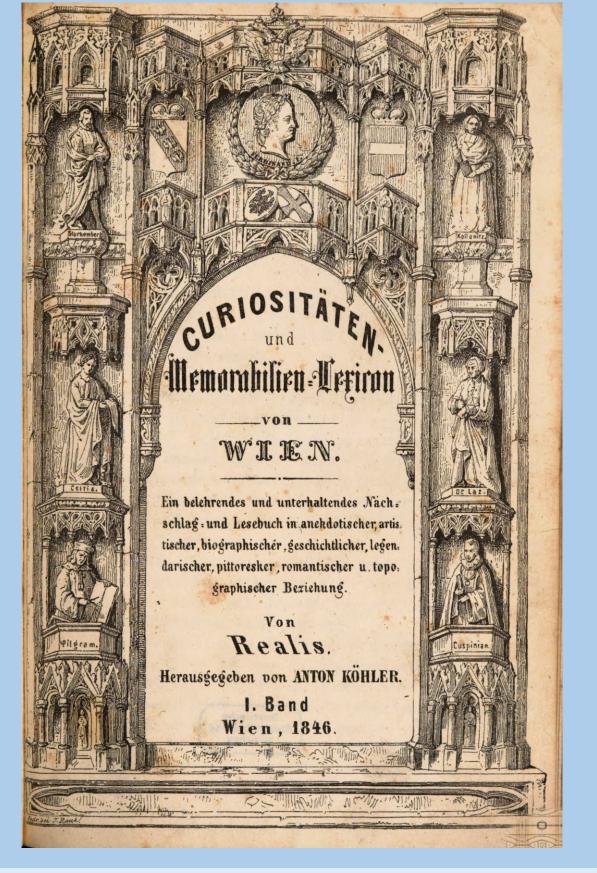




Curiositäten- und Memorabilien Lexicon von Wien Topics and Co-occurence Network of a mid-19th Century Lexicon

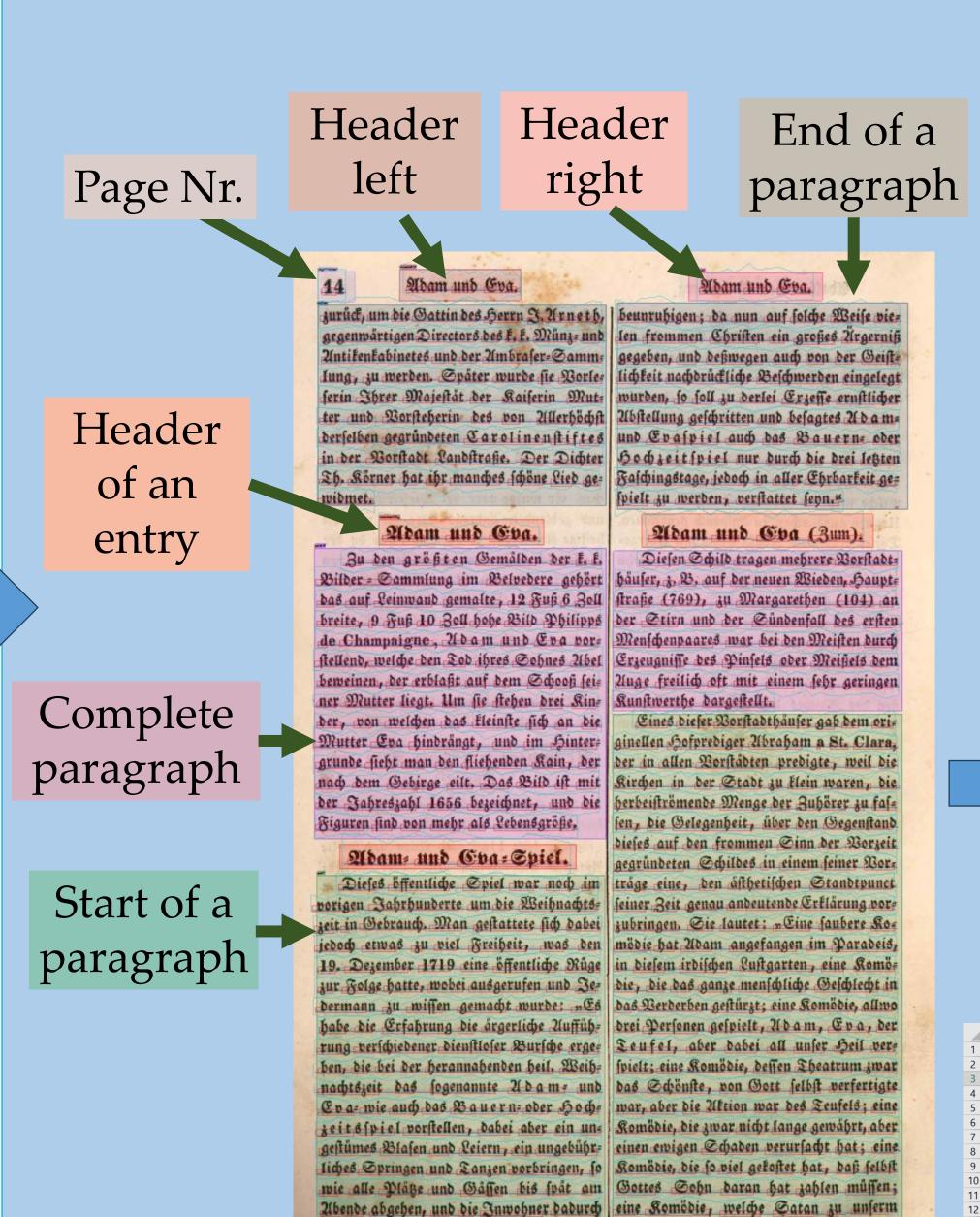
Nikola Krisztian Czindrity (University of Vienna) nikola.krisztian.czindrity@univie.ac.at

1. Back to the Sources



Author: Realis (Pseuodonym for Walter Ritter von Cöckelberghe-Dützele [1786-1857])
Year of publication: 1846
Size: 2 volumes with 962 pages of content with 1295 entries (490K Tokens with 85K Types).
Source: Münchener Digitalisierungs Zentrum While some of the entries provide factual information, others include anecdotal and comical content.

2. Field Model and Transcription



The software *Transkribus* was utilised for field modeling and transcription tasks.

The field model underwent initial training with manual annotations from 30 pages. Subsequent iterations involved integrating annotations from previous models, which were manually corrected, progressively scaling up to 75 pages and then 102 pages.

For transcription, the "German Fraktur 18th Century" model (WrDiarium_M9, ID 47114; ACDH-CH) was utilized.

After ensuring that both the field and transcription models yielded favorable results, the data was extracted in TEI-XML format and tabulated using Python.

book header_id book_page_start book_page_end header_eintrag eintrag_text eintrag_text eintrag_text

C1_MDZ C1_facs_1_region_1699100518519_15 3 4 A enn man eine große Stadt vonder Ferne beschaut, so lehrenuss ['facs_1_region_17]

C1_MDZ C1_facs_2_region_1699101628768_50 4 4 AWOhäufer. Es gereicht dem gesunden Sinne der Be-wohner Wiens zur Ehre ['facs_2_region_16]

C1_MDZ C1_facs_3_region_1699102206352_41 5 5 Abaens. Der Abacus ist ein Bret oder Tisch, woraufvormals, da die Kopfr ['facs_2_region_16]

C1_MDZ C1_facs_3_region_1699102206352_41 5 5 7 Abensperg= und Traunische (Gräf=lich=)

C1_MDZ C1_facs_5_region_1699102206352_41 5 5 7 Abensperg= und Traunische (Gräf=lich=)

C1_MDZ C1_facs_5_region_1699102206352_41 7 7 8 Abgebrannte Haus (Das in der Vorstadt Wieden (447) ist ein ar=chitektonisches Monstru ['facs_5_region_16]

C1_MDZ C1_facs_6_region_1699102837851_70 8 9 Abanhme des Mondes undSternes vom S Am 12. Juli 1686 um 11 Uhr Vor=mittags, begann der Ziegeldecke ['facs_6_region_16]

C1_MDZ C1_facs_7_region_1699102892637_89 9 9 10 Abschnalzer (Der) Herr v. Tru mpf ist ein seltenes Exem=plar von Anmaßung und G ['facs_7_region_16]

C1_MDZ C1_facs_8_region_1699102941837_110 10 10 Academien. Wien zählt der wissenschaftlichen undKunst=Anstalten, die als / ['facs_8_region_16]

C1_MDZ C1_facs_8_region_1699103941837_110 10 11 Achan. Dieser Ort wird in alten UrkundenAichowe und Eich au genannt ['facs_8_region_16]

C1_MDZ C1_facs_9_region_1699103076614_152 11 12 Acht (Die). Da Wien der Sitz beinahe aller deut-sechen Kaiser aus dem Haus (Facs_9_region_16]

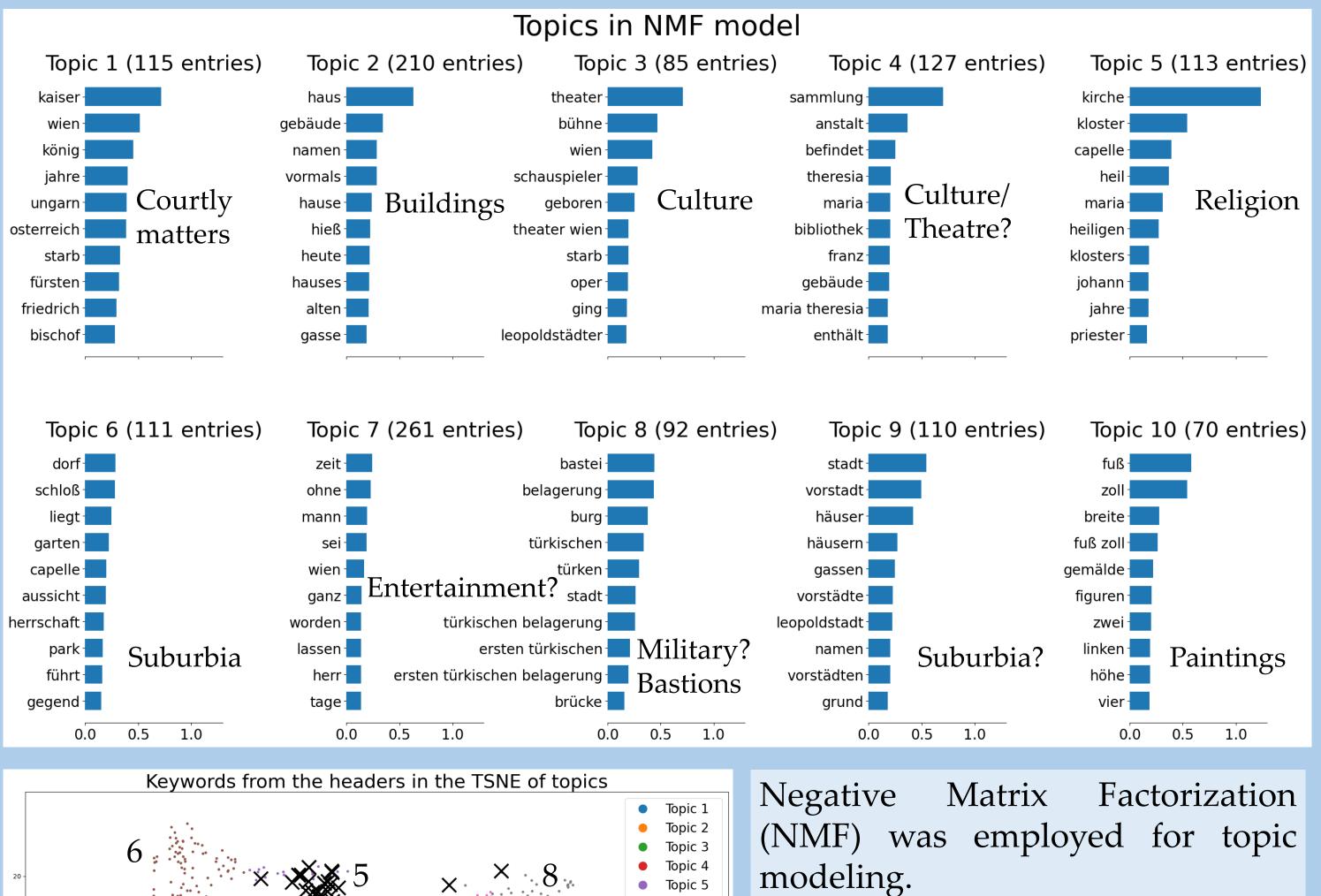
C1_MDZ C1_facs_10_region_1699103130494_171 12 13 Adalbert von Babenberg. Adalbert (Albert), dritter Markgrafvon Osterr. aus dem Haus ele ['facs_11_region_15]

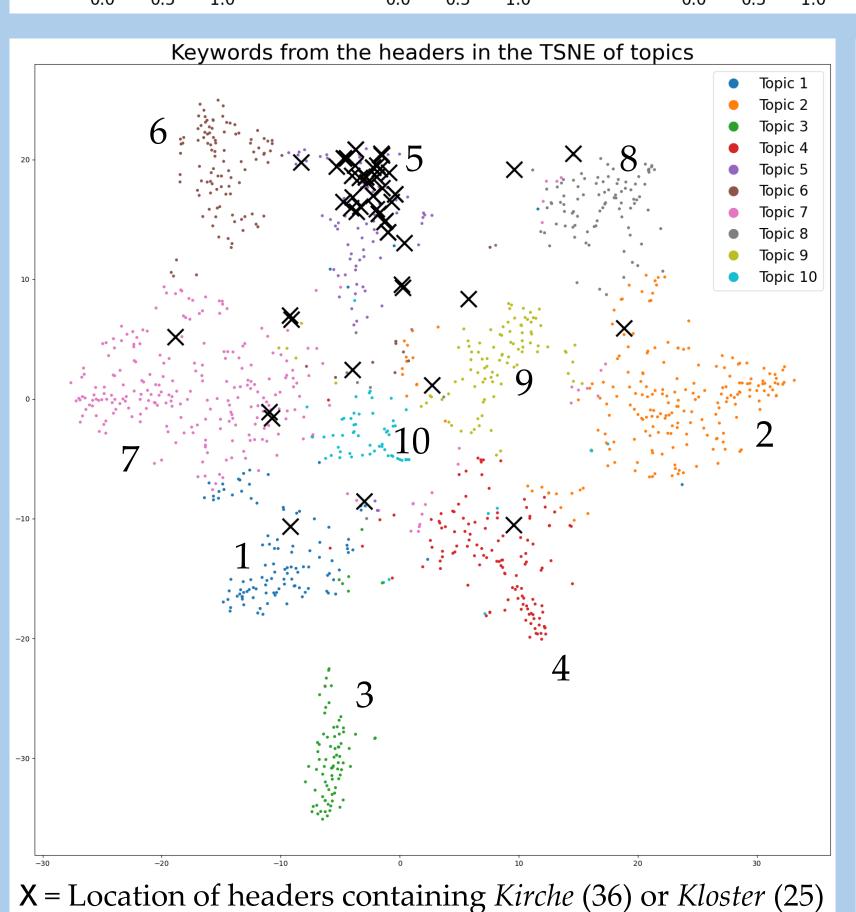
C1_MDZ C1_facs_11_region_1699103205924_208 13 14 Adam und Eva. Zu den größben Gemälden der k. k. Bilder=Sammlung im Belved (['facs_11_region_15]

C1_MDZ C1_facs_12_region_1699103248754_227 14 14 Adam und Eva. Zu den größben Gemälden der k. k. Bilder=Sammlung im Belved (['facs_11_region_15]

C1_MDZ C1_facs_11_region_

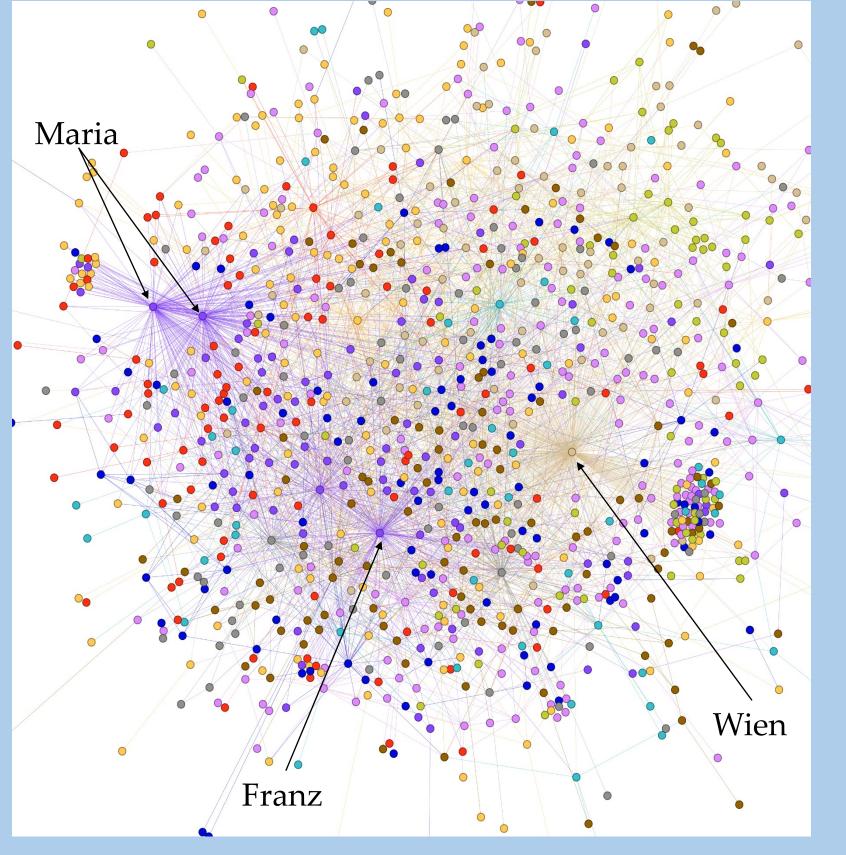
3. Topic Modeling and Co-occurence Network





Upon inspection of the maximal weight for each topic per entry and associated keywords, it is evident that NMF produces meaningful results, with the ten topics identified being well-distributed and coherent.

Further analysis of the reduced lexicon entry - topic matrix using t-SNE revealed that topics form distinct clusters, e.g., keywords such as *Kirche* and *Kloster* are prominent within Topic 5, which focuses on "Religion".

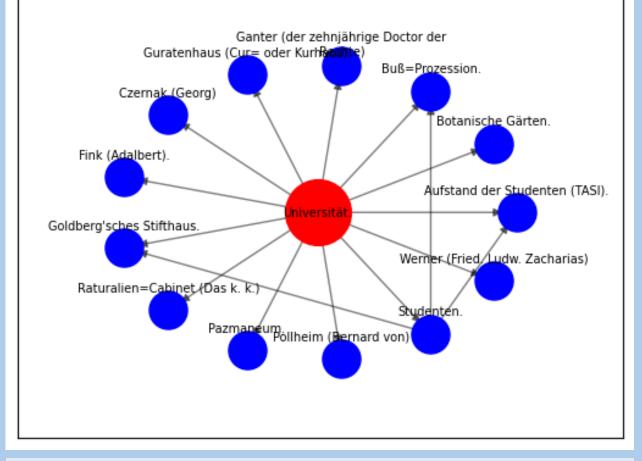


A network based on the co-occurrence of entry headers within the entries consisted of 1102 nodes and 3417 edges. Both nodes and edges are colored according to their association with topics identified by NMF.

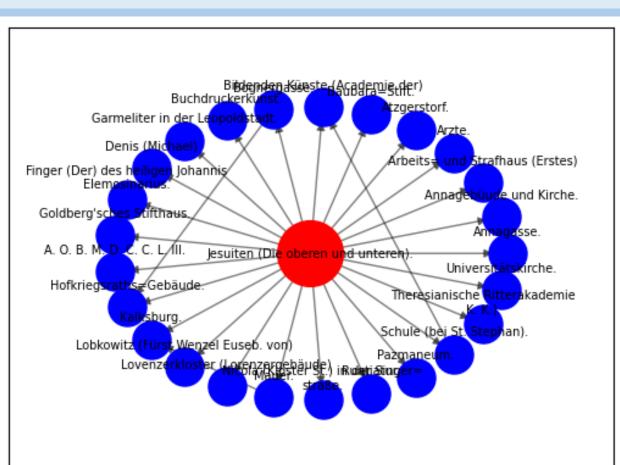
Despite efforts, it was not feasible to identify clusters using betweenness centrality or to produce meaningful interpretations between sets of connections.

However, upon inspecting subgraphs, it becomes apparent that the connections are meaningful. For instance, the entry *Universität* is connected to *Studenten* and *Botanischer Garten*.

Further investigation or different approaches are necessary to disentangle the connections.



Subgraph for the entry *Universität*



Subgraph for the entry Jesuiten



Basilisk from the entry "Basiliskenhaus"

