

Historical Sources of Women in the Digital Humanities

- Overview and Automated Text Analysis

International Summer School, ADW Mainz, 11.10.2019

martin.prell@uni-jena.de

Schedule

- 1. Overview – Sources of Women in DH
- 2. Automated Text Analysis
 - Sample: Letters of Erdmuthe Benigna
 - Natural Language Processing - Overview
 - Preprocessing
 - Topic Modeling
- 3. Conclusion

1. Overview - Sources of Women in DH

Mark Hall, DHd2019 (1)

	2016	2017	2018									
	M	F	M&F	?	M	F	M&F	?	M	F	M&F	?
Vortrag	17	0	3	1	15	0	3	0	19	0	3	0
Poster	16	1	0	1	6	1	0	0	18	0	1	0

Tabelle 3: Anzahl an Abstracts die entweder nur männliche (M), nur weibliche (W), sowohl männliche und weibliche (M&W), oder nur unbekannte (?) Personen erwähnen.

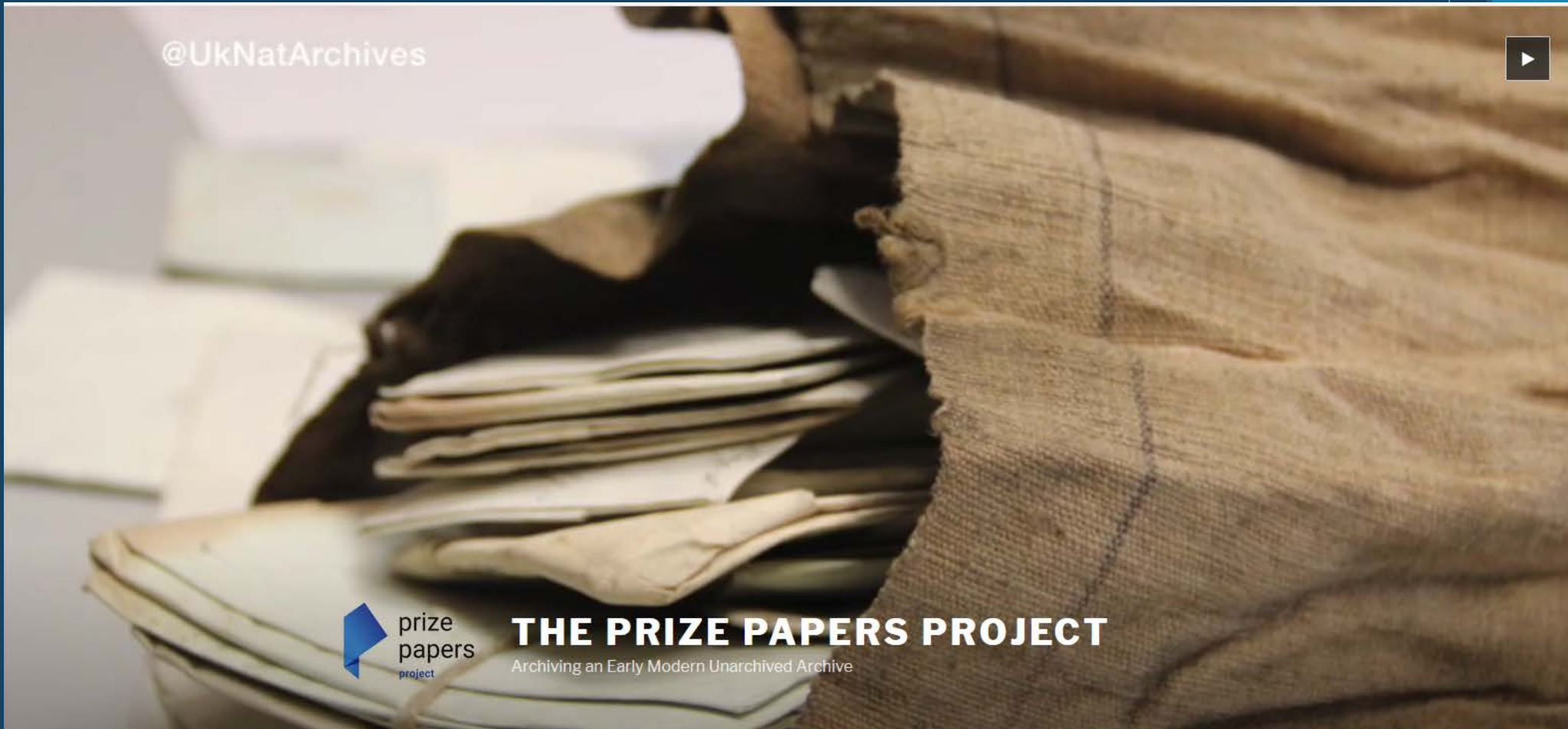
Mark Hall, DHd2019 (2)

	2016	2017	2018							
	M	F	?	M	F	?	M	F	?	
Vortrag	33	4	3	55	7	0	63	3	0	
Poster	19	1	1	6	1	0	33	1	0	


Tabelle 4: Anzahl an Erwähnungen männlicher (M), weiblicher (W), oder unbekannter (?) Personen.

“DH is the Study of dead Dudes”

Project „Prize Papers“



@UkNatArchives

 **THE PRIZE PAPERS PROJECT**
Archiving an Early Modern Unarchived Archive

[The Prize Papers Project](#) [Partners](#) [The Team](#) [Press & Media](#) [Events](#) [↓](#)

Conference „Questioning Models. Intersectionality in Digital Humanities“ Cologne, 11/2017

- ▶ Vera Fasshauer (Goethe-Universität Frankfurt am Main): *Faulty, Clumsy, Negligible? Revaluating the Language and Contents of Early Modern Princesses' Letters*
- ▶ Eymard Fäder, Oliver Vogels, Tilman Lenssen-Erz, Brigitte Mathiak (Universität zu Köln): *Gender Construction in the hunter-gatherer Rock Art of the Brandberg/Daureb, Namibia*
- ▶ Elena Pierazzo (Université Grenoble): *Endangered species: do we need more women writers archives?*
- ▶ Ioanna Kyvernitou (National University of Ireland): *An Ontology for Gendered Content Representation of Cultural Heritage Artefacts*
- ▶ Eleonora Carinci (Università Ca' Foscari, Venezia): *Italian Renaissance women writers: e-ditions*

Conference „Women's History in the Digital World“

“the biennial conference brings together scholars working on women's history projects with a digital component, exploring the complexities of creating, managing, researching and teaching with digital resources”

Rose's List

Women's & Gender History DH list
Datei Bearbeiten Ansicht Einfügen Format Tools Add-ons Hilfe

Dokumentenstruktur <

- List of DH Projects/Resources in ...
- Africa
- Asia
- Australia
- Caribbean
- Europe
- Transnational
- United States
- Open Access Publishing, Podcast...
- Lectures on Women's and Gender...

"La Gazette Royale – A Journey Through Haiti's Early Print Culture," accessed May 15, 2018, <http://lagazetteroyale.com/>.*

Europe

"Archaeology of Reading |." Accessed January 28, 2019. <https://archaeologyofreading.org/>.*

"BBC - 100 Voices That Made the BBC: Pioneering Women - History of the BBC." Accessed June 6, 2019. <https://www.bbc.co.uk/historyofthebbc/women-pioneers>.

"Black Central Europe." Accessed January 2, 2017. <https://blackcentraleurope.com/>. *

"Cork LGBT Archive." Accessed June 10, 2016. <http://corklgbtarchive.com/>.

"Defining Gender." Accessed December 26, 2016. <https://proxy.ulib.csuohio.edu:3793/>. **

"Digital Cavendish Project." Digital Cavendish Project. Accessed December 7, 2018. <http://digitalcavendish.org/>.

"Digital Mitford." Accessed January 10, 2017. <http://digitalmitford.org/index.html> Contact: Elisa Beshero-Bondar, U. of Pittsburgh - Greensburg. Twitter: @DigitalMitford

"Frank and Elizabeth." Accessed June 21, 2017 <https://redthirties.wordpress.com/> Contact: Anna Girling (site unavailable as of January 28, 2019- see Internet Archive link: <https://web.archive.org/web/20180502024134/https://redthirties.wordpress.com/>)

▶ <http://bit.ly/2RVz7ks>



Women's Early Modern Letters Online [WEMLO]

*Co-Directors: Professor James Daybell (Plymouth University)
and Dr Kim McLean-Fiander (University of Victoria)*

WEMLO provides a meeting place for researchers of early modern women letter writers. Created initially with [British Academy/Leverhulme](#) funding and supported by the [Cultures of Knowledge](#) project, WEMLO is a resource and discussion forum for all of the early modern women's correspondence held within the EMLO catalogue.

[Top](#)

[Launch Women's Correspondence](#)

[See all Women's Correspondence Catalogues](#)

WEMLO's Aims

Feedback obtained from a series of WEMLO workshops held in 2013 and 2014 established two principal aims for the project. First, participants requested a distinct scholarly meeting space to cater specifically for the needs of researchers of early modern women. Secondly, while researchers wanted female letter writers to be more discoverable in the historical record, they were concerned about confining them to a female-only digital silo. Rather, scholars wanted to be able to search just for female letter writers as well as to search for them alongside their male counterparts.

WEMLO Today

As a result of WEMLO's partnership with Cultures of Knowledge, it is possible now to [browse all correspondents in the EMLO union catalogue by gender](#) in any combination of senders, recipients, and individuals mentioned in the texts themselves. Visitors may also search for specific women's correspondence by us-



Woman reading a letter, by Johannes Vermeer. c. 1663. Oil on canvas, 46.5 by 39cm. (Source of image: Rijksmuseum, Amsterdam; on loan from the City of Amsterdam [A. van der Hoop Bequest])

Conclusion

- Sources of women have been destroyed
 - Sources of women are often hard to find
 - There is a cognitive bias within research
 - (Digital) Humanities have a gender imbalance
- Sources of women are underrepresented in digital repositories; the DH reinforce this trend

The „Leslie Problem“ (Blevins/Mullen 2015)

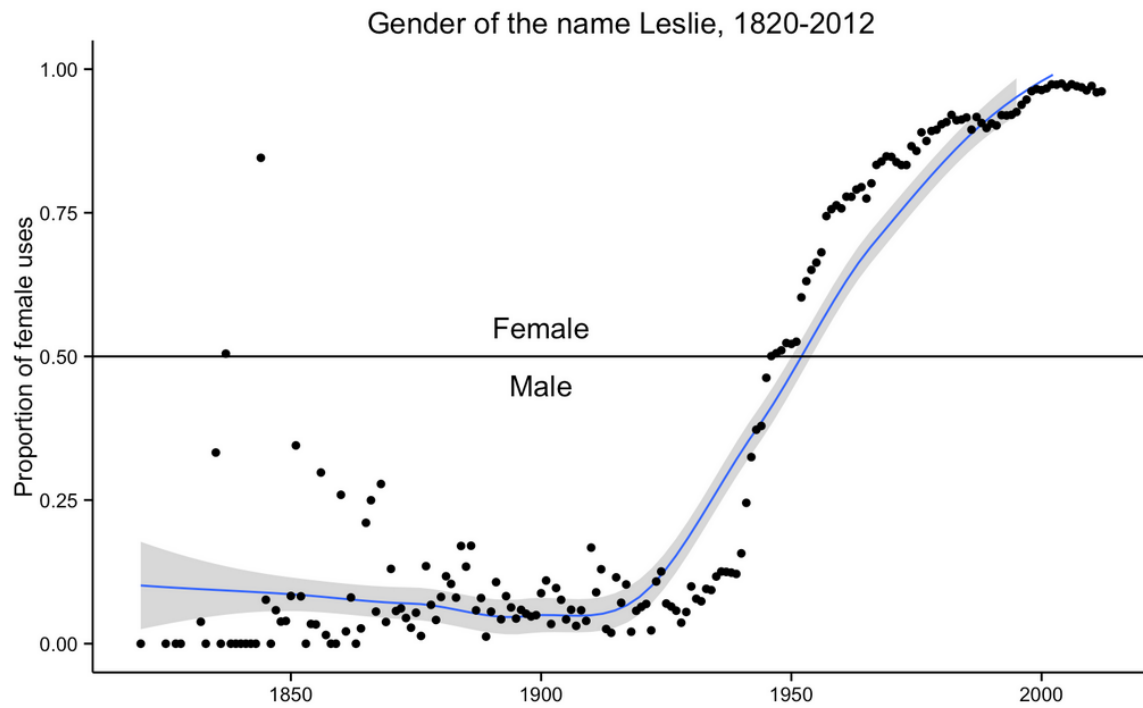


Figure 1. Gender of the name Leslie, 1820-2012

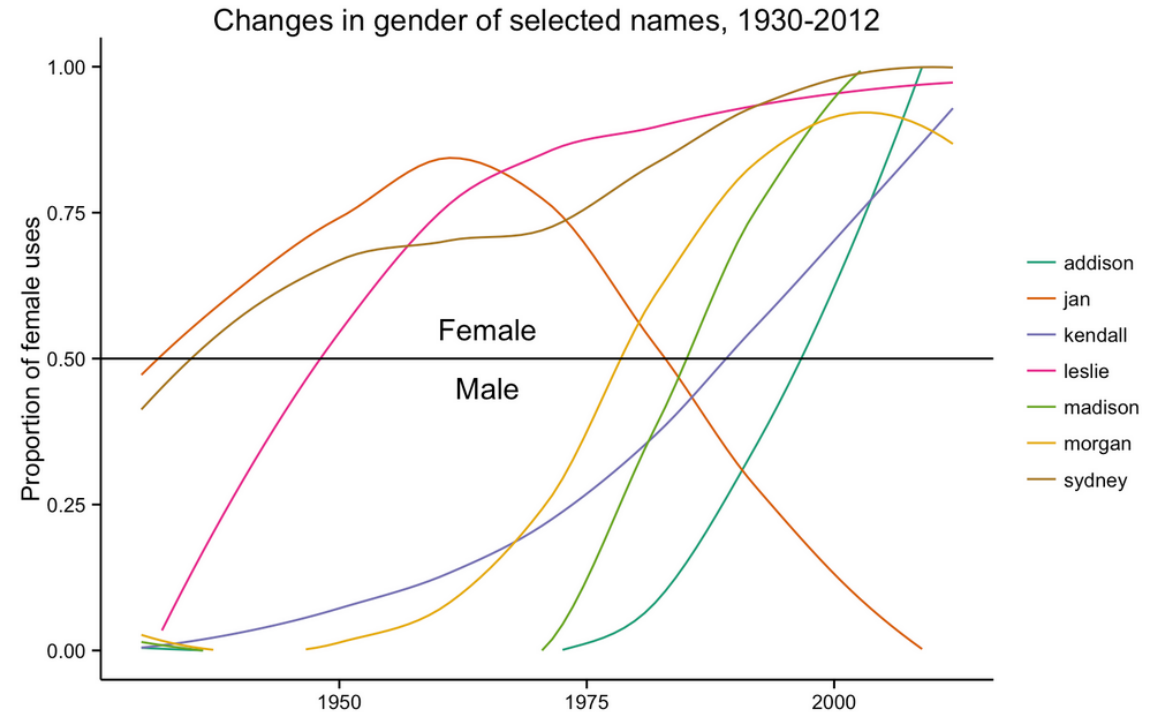


Figure 2. Changes in gender of selected names, 1930-2012

► <http://www.digitalhumanities.org/dhq/vol/9/3/000223/000223.html>

2. Automated Text Analysis

The Gender Novels Project

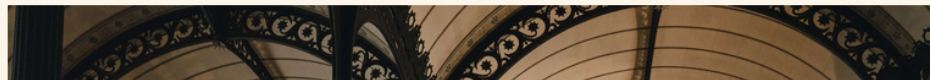
The Gender Novels Project

Home Analyses ▾ Corpus ▾ Development ▾ Our Team

Computational Reading of Gender in Novels

1770–1922

The first project of the MIT Digital Humanities Lab analyzed the description of gender and gender roles across a large repertory of novels in the English language, primarily from the nineteenth century and early twentieth century. Our goal was to build computational tools that helped us understand how conceptions of gender were expressed and changed through distant reading of thousands of books. Our database involves over 4200 novels sourced from the Gutenberg project and gives us a comprehensive look at the views of authors of the time period.



<http://gendernovels.digitalhumanitiesmit.org/>

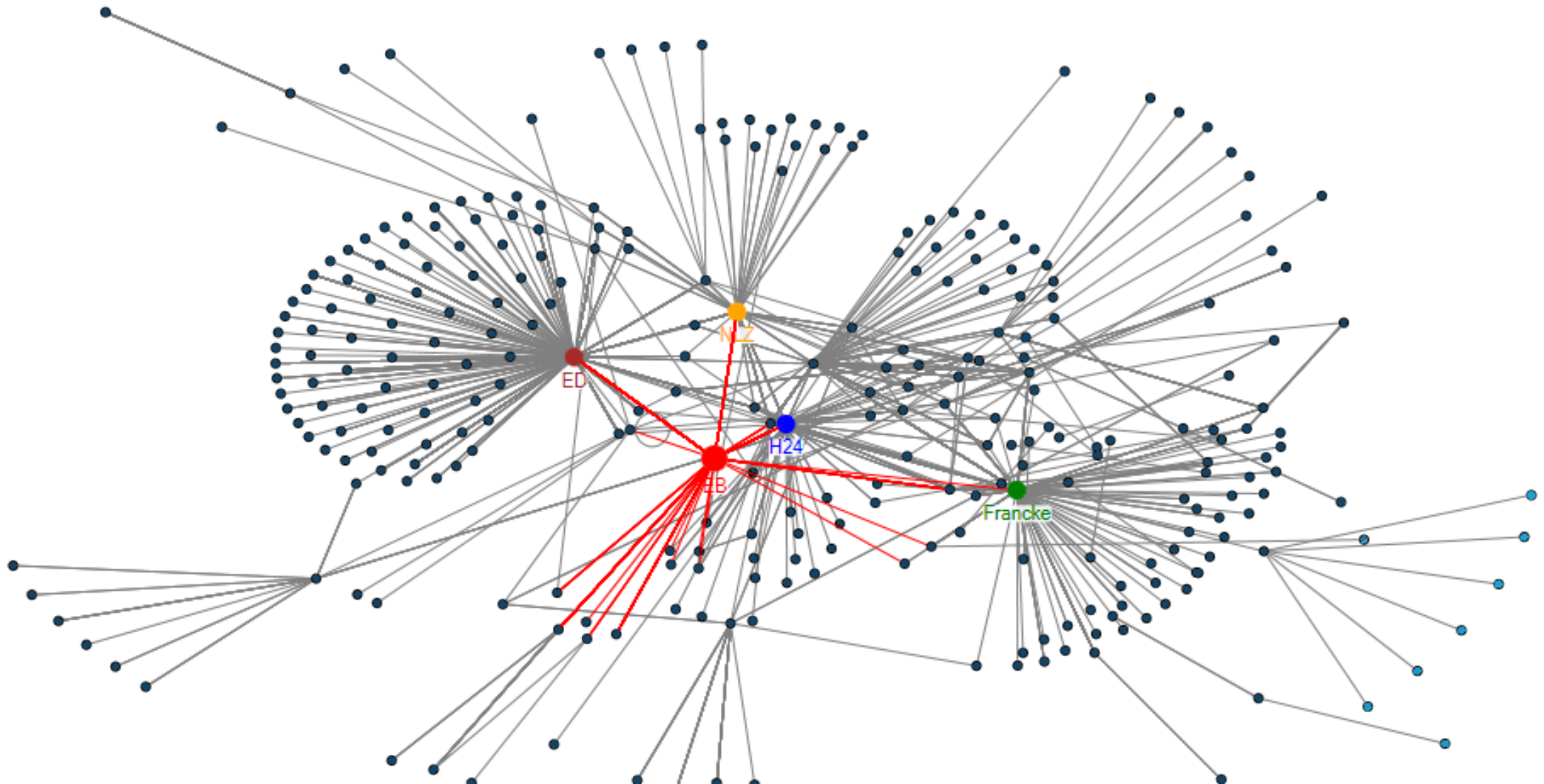
<https://msmagazine.com/2019/04/05/decoding-sex-in-the-humanities-five-key-findings-from-mits-gender-novels-project/>

Ongoing Doctoral Thesis

- ▶ Digital analysis of the correspondence network of Reussian Pietists in the 18th c.
- ▶ Methods: Historical Network Research, Natural Language Processing
- ▶ Aim: Making network positions of several agents (e.g. Erdmuthe Benigna) visible



Correspondence Network of Reussian Pietists (1696-1732)



Letters to Henry XXIV., 1711-1720 (1)



Handwritten note in the left margin: "Hochfürstliche Gnade..."

allerehrlichen H. Rathen

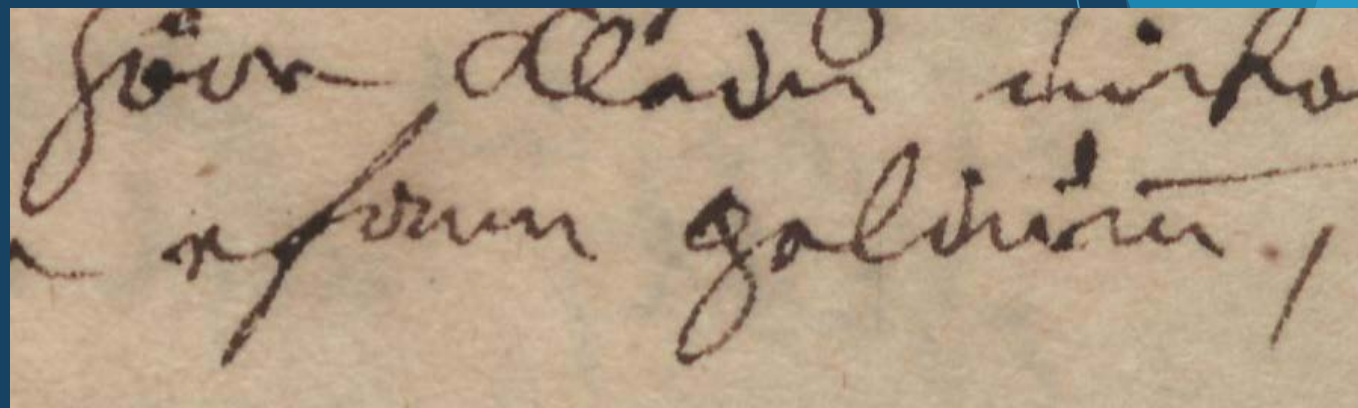
Bayreuth d. 56
October
1711

solche erwidern...
weil ich weiß...
für Bay. wider...
mir ist...
weshalb...
zu wissen ist...
dies...
wird mit...
pflicht...
für...
denn...
zu...



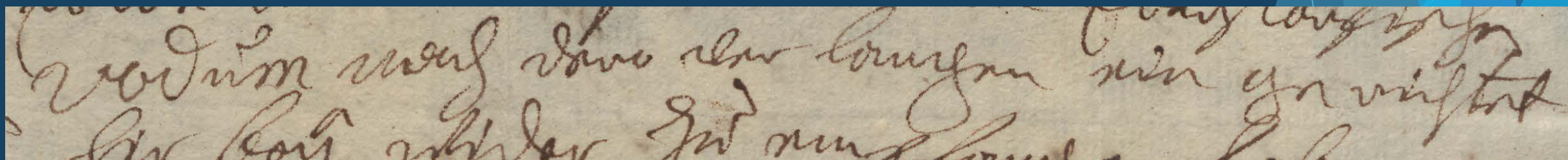
Letters to Henry XXIV., 1711-1720 (2)

Besonderheiten:



fann geliumm

„fann geliumm“



vodum nach dero ver langen ein ge richtet

„vodum nach dero ver langen ein ge richtet“

Majuskelgebrauch bei Dorothea Maria von Sachsen-Weimar

Hochgeborne^r furst freundlicher hertzliebster bruder, E l an mich gedanes schreiben So zur naumburgk geben beneben den ruben, hab ich gar wol entpfangen, vndt thue mich von wegen der selben zum freundlichsten bedancken, b ist mir auch gar lieb gewesen das E l Seindt glicklichen vndt wol zur naumburgk an kommen, wie ich auch nicht anders hofen wil E l Sollen ihre reise vollendt glicklichen volm bracht haben habe auch die grisse neben der entschuldigungk zum besten aus gericht vndt hatt mir meyn hertzliebster herr befolen E l S l wegen wiederumb freundlichen zu grissen vndt E l wollens S l nur nicht vor vbel halten das S l E l nicht beser auff gewartet haben vndt weil S l gestriges tages die ant wort vom kurfursten bekommen, thun S l Sie E l hiemitt zuschicken neben freundlicher bitt E l wollen S l derselben bedencken ferner darinnen mitt theilen des gleichen vber schicken E l S l E l auch das ingelstedische vrdel das Sich E l auch darinnen er sehen können es hetten S l E l auch gerne selber in schriften gerne ersuchen wollen So haben S l doch niemmandt vertrauliches besich ver sehen Sich S l E l werden mitt mir als einen bösen schreiber zu friden Seyn vndt es besser verstehen als es geschrieben ist vndt vnser grosser bott Seyn vndt die griß vber all zum besten aus richten vndt befel E l gottes medigen schutz vndt schirm vndt mich in E l bruder liches hertz datum alden weimar den 17 december Anno 602

E l

alzeitt getreue vndt willigeSchwester bis in tod

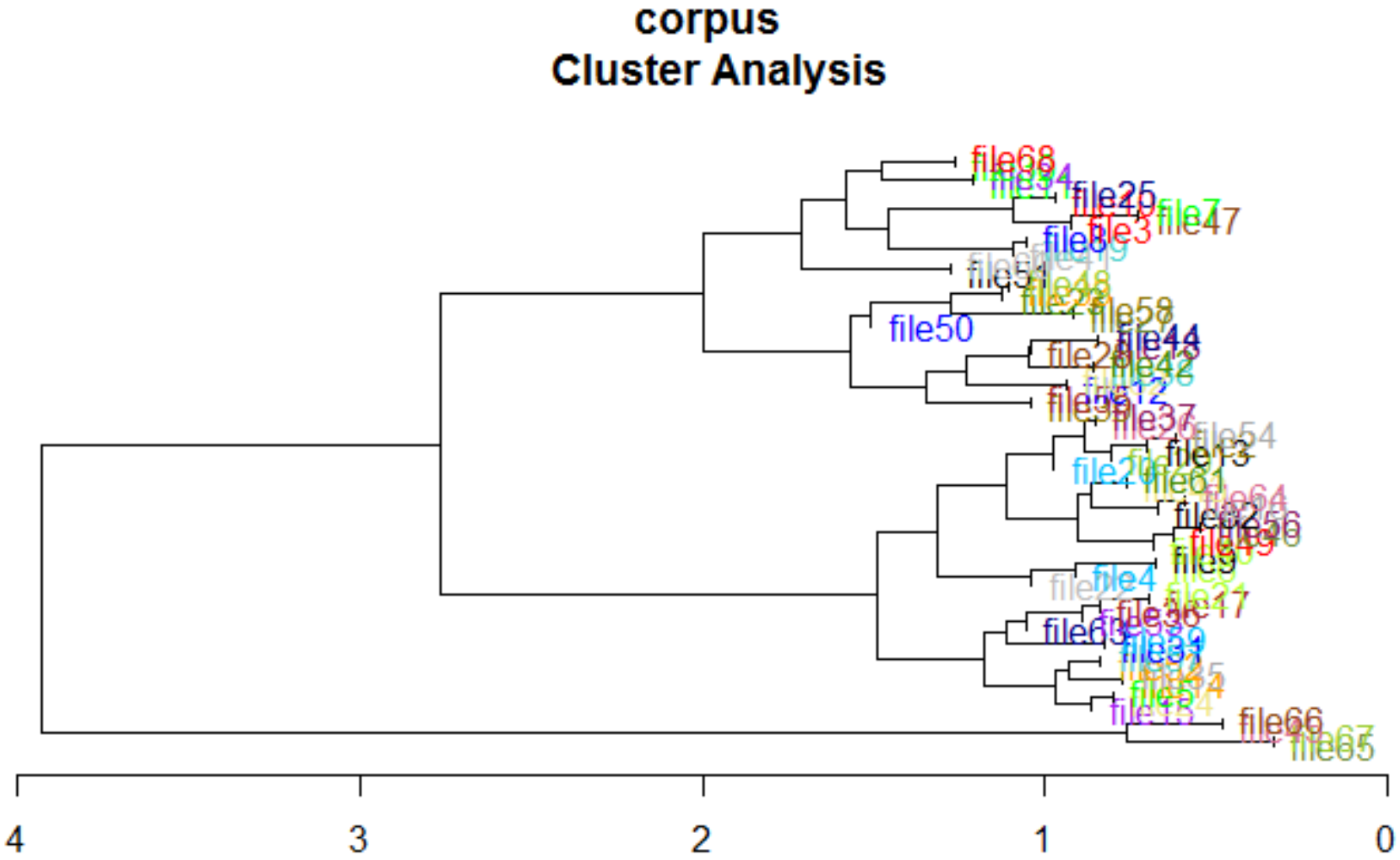
Dorothea Maria h z Sachsen

Match	... norm
Seyn	... sein
Sonsten	... sonst
Sol	... soll
Selber	... selber
So	... so
Sonsten	... sonst
Eltesten	... ältesten
Son	... Sohn
Sol	... soll
Sonsten	... sonst
Söne	... Söhne
Septembris	... September
Schwester	... Schwester
Dorothea	... Dorothea
Maria	... Maria
Sachsen	... Sachsen
Schwester	... freundschwesterlichen
Schwester	... schwesterlichem
Solches	... solches
So	... soviel
Schwester	... Schwester
Sich	... sich
Colegial	... Kollegial-
Sönne	... Söhne
Sache	... Sache
Solches	... solches
Curfürstliche	... kurfürstliche
Colegium	... Kollegium
Selbige	... dasselbige

Natural Language Processing

- ▶ Sentiment Analysis
- ▶ Stylometry

Stilometrie: Letters to Henry XXIV., 1711-1720



100 MFW Culled @ 0%
Classic Delta distance

Natural Language Processing

- ▶ Sentiment Analysis
- ▶ Stylometry
- ▶ Named Entity Recognition
- ▶ Topic Modeling
- ▶ ...

3. Preprocessing of Erdmuthe's Letters

Bayreuth d. 56
Oktob. 1717

aller Liebster Herr Vater

Abgrüßen für alle die ich nicht weiß
den mit der besten Liebe

solch erwidern die ich erwidern Bayreuth
weidlich auch dass der Lungen ein zu viel
für Bay. wider zu nun Lungen haben
mir ist lieb das bald ich zu sollen
wird dasson das durch das den für
zu sehen ist, das aber solch der untern
das so zu mir da mit zu was Lungen
das wenn das durch der Ordnung
und mit Prodestation wider zu viel zu
pflicht, solta ich aber alleine nicht der ganz
für möglich, denn warum wenn, weil das
denn fast auf die der Ordnung so für
halten bey den, halten sie das feine Actum
Proserium in zu nicht können der sehen
und halten so deutlich immer so Prodest,
insoweit Prodestinen in zu auf so mich so
möglich zu werden können, und wenn

Please compare this xml-file with the corresponding „file10.txt“. What types of preprocessing can you identify?

```
230 <text>
231 <body>
232 <pb·fac="PK15_56_0001.tif"·n="56r"/>
233 <div·type="letter">
234 <fw·type="folNum"·place="top-right"·hand="#archivist_hand">56</fw>
235 <opener>
236 <dateline·rend="align-right"><placeName·ref="http://www.geonames.org/2933921/ebersdorf.html"·corresp="#ebersdorf"·
236 resp="#MP">Eber&#x017F;torf</placeName>·<hi·rend="latin"><choice><abbr>d</abbr><expand·resp="#MP">den</expand></choice></hi>·<date·type="sent"·
236 when="1717-10-08">8<lb/>
237 <hi·rend="latin">ocktober</hi><lb/>
238 </date>
239 </dateline><lb/>
240 <salute>
241 <choice><orig>aller·lieb&#x017F;ter</orig><reg·resp="#MP">allerlieb&#x017F;ter</reg></choice>·<choice><abbr>H</abbr><expand·
241 resp="#MP">Herr</expand></choice>·<rs·type="person"·ref="http://d-nb.info/gnd/130132365"·corresp="#heinrichxiv_koestritz"·resp="#MP">vetter</rs>
242 </salute>
243 </opener><lb/>
244 <space·dim="vertical"·unit="lines"·quantity="1"/>
245 <p><choice><abbr>Erltd</abbr><expand·resp="#MP">Erlaucht</expand></choice>·werden·d<subst><del·rend="overwritten">ie</del><add·
245 place="across">a&#x017F;</add>/subst>·<choice><orig>ver·enderde</orig><reg·resp="#MP">verenderde</reg></choice>·<rs·type="place"·
245 ref="http://www.geonames.org/2933921/ebersdorf.html"·corresp="#ebersdorf"·resp="#MP">Eber&#x017F;torfi&#x017F;che</rs><lb/>
246 <hi·rend="latin"·xml:lang="lat">vodum</hi>·nach·dero·<choice><orig>ver·langen</orig><reg·resp="#MP">verlangen</reg></choice>·
246 <choice><orig>ein·ge·richtet</orig><reg·resp="#MP">eingerichtet</reg></choice><lb/>
247 <choice><orig>hir·b<choice><unclear·reason="illegible"·cert="high"·resp="#MP">e</unclear><unclear·reason="illegible"·cert="low"·
247 resp="#MP">o</unclear></choice>&#x00FF;</orig><reg·resp="#MP">hirbe&#x00FF;</reg></choice>·wider·zu·empfangen·haben,<lb/>
248 <mir·i&#x017F;t·lieb·d<choice><unclear·reason="illegible"·cert="high"·resp="#MP">a</unclear><unclear·reason="illegible"·
248 cert="low"·resp="#MP">o</unclear></choice>·B·<choice><abbr>Erltd</abbr><expand·resp="#MP">Erlaucht</expand></choice>·&#x017F;ich·<choice><orig>ge·
248 fallen</orig><reg·resp="#MP">gefallen</reg></choice>·laße<choice><orig>&#x02F7;</orig><reg·resp="#MP">n</reg></choice><lb/>
```

Automated Text Normalization with CAB

<https://kaskade.dwds.de/demo/cab/>

DTA::CAB Web Service v1.110



Query:

Analyzer: default ▾

Format: Text ('vertical' attributes) ▾

Flags: pretty clean exlex tokenize

Options:

URL: <https://kaskade.dwds.de/demo/cab>

[File Upload](#) | [Analyzers](#) | [Formats](#) | [Status](#) | [Documentation](#)

DTA::CAB::Server::HTTP

DTA::CAB v1.110 (30756)

jurish@bbaw.de

[Imprint](#) · [Privacy](#)

If you use this service in an academic context, please include the following citation in any related publications:

Jurish, B. *Finite-state Canonicalization Techniques for Historical German*. PhD thesis, Universität Potsdam, 2012 (defended 2011). URN [urn:nbn:de:kobv:517-opus-55789](https://nbn-resolving.org/urn:nbn:de:kobv:517-opus-55789).

See [here](#) for a list of other CAB-related publications.

CAB - Limitations (file10.txt)

Eberstorf den 8 Oktober 1717 allerliebster Herr Vetter Erlaucht werden das veränderte Eberstorfische Votum nach deren verlangen eingerichtet hierbei wider Zu empfangen haben , mir ist lieb daß Erlaucht sich gefallen lassen was wegen des Dankfestes von hier aus geschehen ist , daß aber Erlaucht vermeinen obschon sie weil es nun geschehen endlich auch damit zufrieden wören daß es genug damit gewesen wäre , daß Mann des Körpers verordnugen Zerschnitten und mit Protestation wider Zurück geschickt , halte ich aber alleine nicht vor ganz k Mann , doch daß Dankfest auf die verordnug so hier h ten sie doch Einen actum Bosesorium in Zukunft kö ten es künftig immer so probiert , unseres prote muß es endlich Zur Sprache kommen , und wird sich recht Zeichen es muß doch pichen oder brechen , so t lassen geschehen A. Superintendent betreff hätten die anderen Herren in geistlichen Dingen mehr in unseren Lande Zu sagen und wider unseren Willen Zu verordnen als wir selbst der teifel ist recht los lässt sie toben witten lachen , Gott unser heil wird in Eile sie Zu Schanden machen , unverzagt und ohne grauen soll ein Christ wo er ist stets sich lassen schauen , es muß ja erst recht finster ps die lorel keße ich herzlichwerten ehe es lichte wird , Gott wird ihnen schon endlich ein Gebiß ins maul legen , Erlaucht haben vergangen mir geschrieben wie es doch Köm daß wir jetzt

Flexion

Syntax

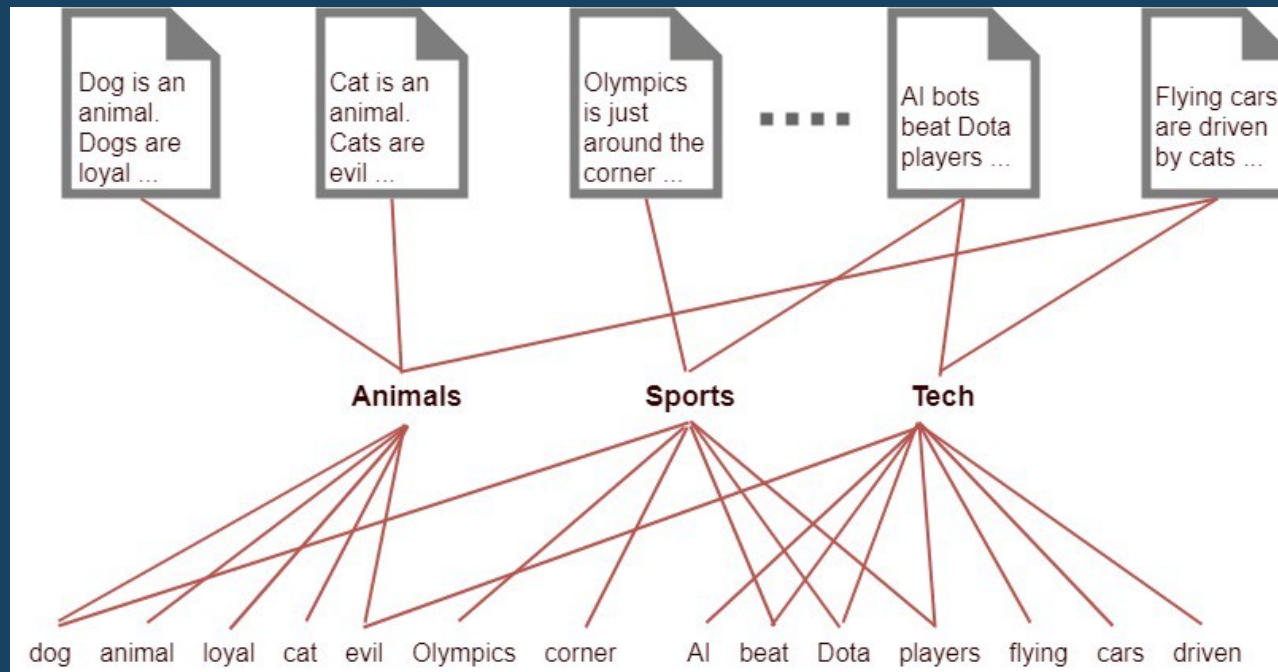
Punctuation

Topic Modeling - Hands On

Topic Modeling


► LDA (Latent Dirichlet Allocation):

“Each document can be described by a distribution of topics and each topic can be described by a distribution of words”



Voyant

<https://voyant-tools.org/>



VOYANT
see through your text

Add Texts 📄 🔍 ?

Type in one or more URLs on separate lines or paste in a full text.

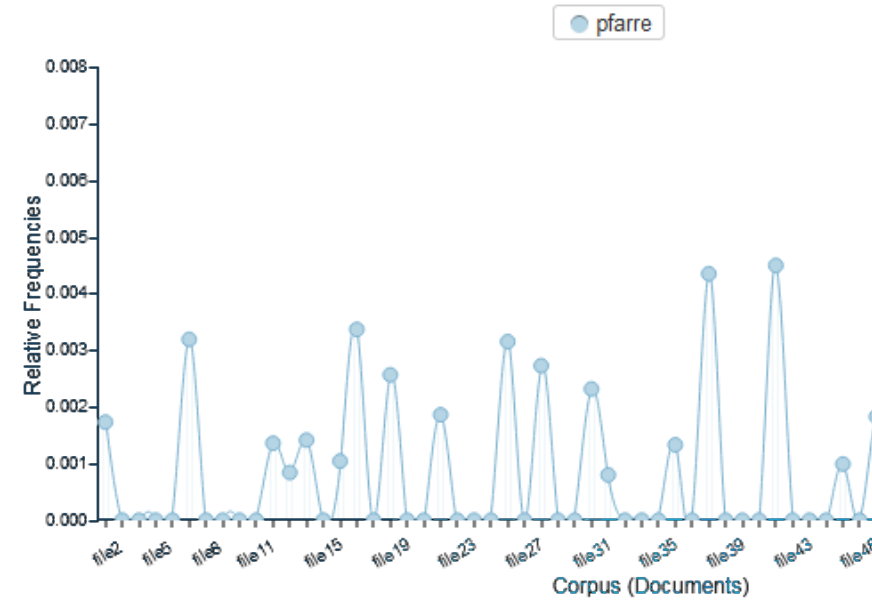
Voyant Tools is a web-based reading and analysis environment for digital texts.

Voyant Tools , Stéfán Sinclair & Geoffrey Rockwell (© 2019) Privacy v. 2.4 (M29)

Result

Topic	Scores
kanzler kinder votum regierung herzen reiße schaden gnaden zukunft gemeinschaft	
pfarre halle consistorio consistorium gelt seele kinder ehre game babst	
groschen florentiner gulden untertanen meinug resolution übel watzdorf bedenken rechnug	
meinug taler herschafft gewissen bier hoffen pfarre kindern verantworten gnädig	

file2
allerliebster Herr
Väter Ich habe
nun alles getan
was in meinem
vermögen
gewesen
, um Erlaucht der
lässt der
Vormundschaft Zu
überheben , bin
letz deswegen
nach der
Burg gefahren ,
und habe am 13



Dariah Topics Explorer

A total of 10 topics in 66 documents

Topics are probability distributions over the whole vocabulary of a text corpus. One value is assigned to each word, which indicates how *relevant* the word is to that topic (to be exact, how *likely* one word is to be found in a topic). After sorting those values in descending order, the first n words represent a topic.

Below the topics are ranked by their *numerical dominance* in the corpus; each bar displays a topic's dominance score. You can start exploring the topic model by **clicking on the topic bars** – or select one of the other tabs in the above menu.



TOP 15: RELATED WORDS

halle
meinug
kinder
pfarre
consistorio
seele
versprochen
mitvormund
amt
gottes
befördern
ehre
ostern
pfarrsache
heßlern

TOP 3: SIMILAR TOPICS

SCHADEN, ÜBEL, FÖRSTER, ...
PFARRE, CONSISTORIUM, FROMMEN, ...
KANZLER, REGIERUNG, VOTUM, ...

TOP 10: RELATED DOCUMENTS

FILE3
FILE21
FILE18
FILE25
FILE2
FILE31
FILE28
FILE60
FILE19
FILE11

Conclusion

- ▶ We need more sources (i.e. scholarly editions) of underrepresented agents with different social and regional backgrounds as digital, easily accessible, freely available, widely visible and well-structured data
- ▶ As scholars of humanities, we need first and foremost a diplomatic transcription without tacit smoothing in order to a) document female writing practice and b) protect the original source
- ▶ As digital humanists, we also need a normalized transcription that helps to overcome at least basic hurdles of automated processing
- ▶ We need pre-processing algorithms for pre-modern texts that do not yet belong to the trained text canon. E.g. further development of automated orthographic normalizers