

Climate Change Glossary

Text Mining in the Social Sciences, Prof. Dr. Stede
Student: Anna-Janina Goecke



Project Information

General

- Creation of a German glossary containing loaded climate change terms

Methods

- Online Research
- Web Scraping
- Text Analysis Tools



eco-anxiety

climate change denial

climate change fanatics

Klimaglossar

climate change lie

climate change hysteria

Klimabegriffe

Klimaphobie, Klimapsychose, Klimafundamentalismus, Klimafanatismus, Klimapopulismus, Klimagerechtigkeit, Klimawahn, Klimaleugner, Klimalüge, Klimaalarm, Klimasünde, Klimagefahr, Klimarisiko, Klimabelastung, Klimamüdigkeit, Klimaskeptiker, Klimahysteriker, Klimabeeinflussung, Klimawandelleugner, Klimareligion, Klimakiller, Klimakrise, Klimanotstand, Klimakatastrophe, Klimanotstand, Klimaproblem, Klimamissstand, Klimabedrohung, Klimasorge, Klimaangst, Klimaversagen, Klimahysterie, Klimaschwindel, Klimalockdown, Klimarelativismus, Klimaschwindel, Klimalügner, Klimasekte, Klimapsychose, Klimawandelfeindlichkeit, Klimaextremismus, Klimafundamentalismus, Klimafanatismus, Klimapopulismus, Klimaterror, Klimairrsinn, Klimaangst, Klimapanik, Klimaverdiener, Klimakonservative, Greta-Effekt, Klimaindustrie, Klimalobbyisten

Web Scrapping

Tool: Trafilatura

- Command-line tool and Python package
- very easy and intuitively to use

Steps:

- Get links of main page and all subpages
- Retrieve text segments from all links

<u>Websites</u> EIKE Klimareporter Fridays For Future
--

Textanalysis

Tool: Quanteda

Preprocessing

- Lowercasing
- Remove stopwords and punctuation
- Tokenization
- Optional: Stemming
- Create a document-feature matrix

Document-Feature Matrix

Tool: Quanteda

Fridays for Future

```
Document-feature matrix of: 6 documents, 4,956 features (94.01% sparse) and 2 docvars.  
features  
docs      climat island åland albania afghanistan action strike futur unit world  
1jhpGg9y.txt      18     158     157      150           146       4       2      24     63       5  
3dzTRw5p.txt       1       0       0       0           0       4       2       0       0       3  
3gNVM4c5.txt       4       0       0       0           0       1       2       1       0       2  
5ZdQ7PMb.txt       9       0       0       0           0      10       7       3       0       5  
7Pk2VLdn.txt       0       0       0       0           0       0       0       1       0       0  
8PB7y1o7.txt      20       0       0       0           0      11       6       2       1       6  
[ reached max_nfeat ... 4,946 more features ]
```

Textanalysis

Tool: Quanteda

Most Frequent Words

Fridays for Future



Klimareporter

**EIKE**

Textanalysis

Tool: Quanteda

Most Frequent „Climate“ Words

Klimareporter



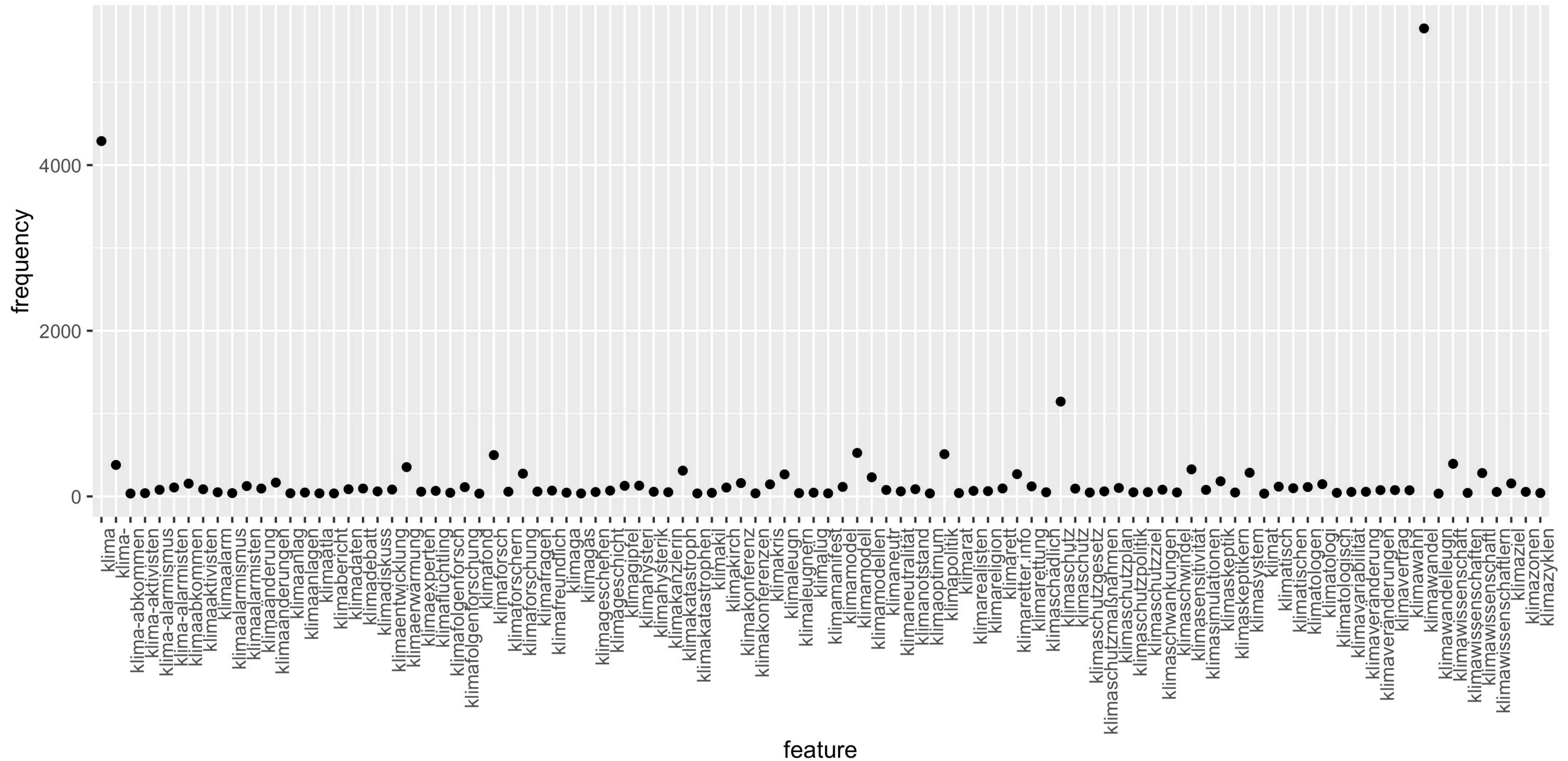
EIKE



Textanalysis

Tool: Quanteda

Most Frequent „Climate“ Words - EIKE



Textanalysis

Tool: Quanteda

Keyword-in-Context

- Returns a list of a keyword in its immediate context (preceding and following words)

Keyword-in-Context

	docname	from	to	pre	keyword	post
1	1jhpGg9y.txt	206	206	considered the importance of rapid	climate	action otherwise . Social media
2	1jhpGg9y.txt	299	299	shift the narrative of the	climate	crisis is a reality -
3	1jhpGg9y.txt	321	321	that shows the reality of	climate	change to the world's most
4	1jhpGg9y.txt	452	452	Passing the mic will enable	climate	change to be tackled from
5	1jhpGg9y.txt	485	485	bare the root causes of	climate	change will achieve climate and
6	1jhpGg9y.txt	489	489	of climate change will achieve	climate	and social justice for all
7	1jhpGg9y.txt	628	628	unity in the fight against	climate	injustice . We have to
8	1jhpGg9y.txt	3811	3811	Philippines , Youth Advocates for	Climate	Action Philippines , Fridays for
9	1jhpGg9y.txt	4045	4045	Fridays for Future Calgary ,	Climate	Strike Canada - Ismael Canizalez
10	1jhpGg9y.txt	4304	4304	, Belgium , Youth for	Climate	Belgium - Rachel Heckerman ,
11	1jhpGg9y.txt	4345	4345	Ines Vilhena , Portugal ,	climate	activist - Roxana Borda ,
12	1jhpGg9y.txt	4393	4393	Future Bogotá , Latinas for	Climate	- Karin Watson Ferrer ,
13	1jhpGg9y.txt	4448	4448	Future Chile & Latinas For	Climate	- Emily Mall , United
14	1jhpGg9y.txt	5049	5049	, Northern Ireland , Youth	Climate	Association NI / Fridays for

Textanalysis

Tool: Quanteda

Collocations

- Words that co-occur
- Can be used to identify English „climate“ compounds (e.g. „climate change“)

Collocations

Klimareporter

	collocation	count	count_nested	length
1	klimawandel möglich	29	0	2
2	in den	86	0	2
3	macht unabhängigen	29	0	2
4	° wird	30	0	2
5	mit dem	46	0	2
6	wenn sie	34	0	2
7	bis 2030	26	0	2
8	für die	101	0	2
9	gute wille	19	0	2
10	in der	110	0	2
11	gibt es	34	0	2
12	sie unabhängigen	29	0	2
13	dieser website	16	0	2
14	in deutschland	39	0	2
15	zu energiewende	29	0	2
16	aus dem	28	0	2
17	auf dem	34	0	2
18	negative emissionen	15	0	2

Outlook

Next Steps:

- Create custom stopwords lists
- Extraction of climate change terms
- Sentiment Analysis
- Topic Modeling
- Collaboration with BBAW (Berlin-Brandenburgische Akademie der Wiss.)
- Analysis of results

References

Julia Silge and David Robinson. 2017. Text Mining with R: A Tidy Approach (1st. ed.). O'Reilly Media, Inc.

<http://inhaltsanalyse-mit-r.de/grundlagen.html>

<https://quanteda.io/articles/quickstart.html>

<https://trafilatura.readthedocs.io/en>

<https://trafilatura.readthedocs.io/en/latest/tutorial0.html>