

# MALLET-BNE

October 31, 2021

```
[1]: path_to_mallet = '/home/almahartwig/Downloads/mallet-2.0.8/bin/mallet'
```

```
[2]: #!/pip install little_mallet_wrapper
     #!/pip install seaborn
```

```
[3]: import little_mallet_wrapper
     import seaborn
     import glob
     import pandas as pd
     import numpy as np
     import nltk
     from pathlib import Path

     from nltk.corpus import stopwords
```

```
[4]: directory = "/home/almahartwig/Dokumente/Studium-Master/SoSe-21/
     ↳Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt"
```

```
[5]: files = glob.glob(f"{directory}/*.txt")
```

```
[6]: files
```

```
[6]: [' /home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-
Info/Datensatz/Training-Data/training-txt/indikatoren-der-bildung-fuer-
nachhaltige-entwicklung.txt',
      ' /home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-
Info/Datensatz/Training-Data/training-txt/philosophie-einer-humanen-
bildung.txt',
      ' /home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-
Info/Datensatz/Training-Data/training-txt/zukunftsfaehigkeit-im-kindergarten-
vermitteln-kinder-staerken-nachhaltige-entwicklung-befoerdern.txt',
      ' /home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-
Info/Datensatz/Training-Data/training-txt/unesco-heute-un-dekade-bne.txt',
      ' /home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-
Info/Datensatz/Training-Data/training-txt/nachhaltigkeit-im-handel-n-tipps-fuer-
die-ausbildungs-praxis.txt',
      ' /home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-
```

Info/Datensatz/Training-Data/training-txt/nachhaltige-entwicklung-auf-kommunaler-ebene-durch-bildung-voranbringen.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/gemeinsam-fuer-unsere-zukunft.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/unesco-heute-nachhaltige-entwicklung.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/qualitaetskriterien-fuer-die-fortbildung-von-multiplikatorinnen-und-multiplikatoren.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/chancengerechtigkeit-im-deutschen-bildungssystem.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/biologische-vielfalt-und-bildung-fuer-nachhaltige-entwicklung.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/hochschulen-fuer-nachhaltige-entwicklung-erklaerung-der-hochschulrektorenkonferenz-und-der-deutschen-unesco-kommission-zur-hochschulbildung-fuer-nachhaltige-entwicklung.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/hochschulen-fuer-eine-nachhaltige-entwicklung-ideen-zur-institutionalisierung-und-implementierung.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/hochschulen-fuer-eine-nachhaltige-entwicklung-nachhaltigkeit-in-forschung-lehre-und-betrieb.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/umsetzung-der-un-dekade-bildung-fuer-nachhaltige-entwicklung-zwischenbericht-mitte-2007-bis-mitte-2010.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/ausgezeichnet-kommunen-lernorte-und-netzwerke.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/vom-projekt-zur-struktur-projekte-massnahmen-und-kommunen-der-un-dekade-bildung-fuer-nachhaltige-entwicklung.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/bonner-erklaerung-2014.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/querbeet-biologische-vielfalt-und-bildung-fuer-nachhaltige-entwicklung-anregungen-fuer-die-praxis.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/informationsblatt-zur-einreichung-von-commitments-zum-nationalen-aktionsplan-bildung-fuer-nachhaltige-entwicklung-nap-bne.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/bildung-fuer-nachhaltige-entwicklung-im-spiegel-von-kunst-kultur-und-kultureller-bildung.txt',

'/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/hochschulen-fuer-eine-nachhaltige-entwicklung-netzwerke-foerdern-bewusstsein-verbreiten.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/nationaler-aktionsplan-bildung-fuer-nachhaltige-entwicklung-2017.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/der-beitrag-der-un-dekade-2005-2014-zu-verbreitung-und-verankerung-der-bildung-fuer-nachhaltige-entwicklung.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/un-dekade-mit-wirkung-10-jahre-bne-in-deutschland.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/zukunftsstrategie-2015-positionspapier-des-nationalkomitees-der-un-dekade.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/vom-projekt-zur-struktur-strategiepapier-der-ag-berufliche-aus-und-weiterbildung.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/lehrmaterialien-zum-jahresthema-mobilitaet.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/deutsche-unesco-kommission-jahrbuch-2016-2017.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/roadmap-zur-umsetzung-des-weltaktionsprogramms.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/zukunftsfaehige-kommunen-chancen-durch-bildung-fuer-nachhaltige-entwicklung.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/tagungsbericht-un-dekade-bildung-fuer-nachhaltige-entwicklung-der-beitrag-europas.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/zwischenbilanz-zum-nationalen-aktionsplan-bildung-fuer-nachhaltige-entwicklung.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/strukturen-staerken-ausgezeichnete-kommunen-lernorte-und-netzwerke-des-unesco-weltaktionsprogramms-bildung-fuer-nachhaltige-entwicklung-im-portraet.txt',  
 '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/Hausarbeit-Trends-Info/Datensatz/Training-Data/training-txt/referenzrahmen-fuer-die-fruehkindliche-bildung.txt']

```
[7]: stopwords = nltk.corpus.stopwords.words('german')
newStopWords = ['www', 'hrsg', 'pdf', 'bzgl', 'duk', 'for' ]
stopwords.extend(newStopWords)
#print(stopwords)

[8]: #little_mallet_wrapper.process_string(text, numbers='remove',
    ↪stop_words=stopwords)

[9]: training_data = []
for file in files:
    text = open(file, encoding='utf-8').read()
    processed_text = little_mallet_wrapper.process_string(text,
    ↪numbers='remove', stop_words=stopwords)
    training_data.append(processed_text)

[10]: original_texts = []
for file in files:
    text = open(file, encoding='utf-8').read()
    original_texts.append(text)

[11]: obit_titles = [Path(file).stem for file in files]

[12]: obit_titles

[12]: ['indikatoren-der-bildung-fuer-nachhaltige-entwicklung',
'philosophie-einer-humanen-bildung',
'zukunftsfaehigkeit-im-kindergarten-vermitteln-kinder-staerken-nachhaltige-
entwicklung-befoerdern',
'unesco-heute-un-dekade-bne',
'nachhaltigkeit-im-handel-n-tipps-fuer-die-ausbildungs-praxis',
'nachhaltige-entwicklung-auf-kommunaler-ebene-durch-bildung-voranbringen',
'gemeinsam-fuer-unsere-zukunft',
'unesco-heute-nachhaltige-entwicklung',
'qualitaetskriterien-fuer-die-fortbildung-von-multiplikatorinnen-und-
multiplikatoren',
'chancengerechtigkeit-im-deutschen-bildungssystem',
'biologische-vielfalt-und-bildung-fuer-nachhaltige-entwicklung',
'hochschulen-fuer-nachhaltige-entwicklung-erklaerung-der-
hochschulrektorenkonferenz-und-der-deutschen-unesco-kommission-zur-
hochschulbildung-fuer-nachhaltige-entwicklung',
'hochschulen-fuer-eine-nachhaltige-entwicklung-ideen-zur-institutionalisierung-
und-implementierung',
'hochschulen-fuer-eine-nachhaltige-entwicklung-nachhaltigkeit-in-forschung-
lehre-und-betrieb',
'umsetzung-der-un-dekade-bildung-fuer-nachhaltige-entwicklung-zwischenbericht-
mitte-2007-bis-mitte-2010',
'ausgezeichnet-kommunen-lernorte-und-netzwerke',
```

```
'vom-projekt-zur-struktur-projekte-massnahmen-und-kommunen-der-un-dekade-
bildung-fuer-nachhaltige-entwicklung',
'bonner-erklaerung-2014',
'querbeet-biologische-vielfalt-und-bildung-fuer-nachhaltige-entwicklung-
anregungen-fuer-die-praxis',
'informationsblatt-zur-einreichung-von-commitments-zum-nationalen-aktionsplan-
bildung-fuer-nachhaltige-entwicklung-nap-bne',
'bildung-fuer-nachhaltige-entwicklung-im-spiegel-von-kunst-kultur-und-
kultureller-bildung',
'hochschulen-fuer-eine-nachhaltige-entwicklung-netzwerke-foerdern-bewusstsein-
verbreiten',
'nationaler-aktionsplan-bildung-fuer-nachhaltige-entwicklung-2017',
'der-beitrag-der-un-dekade-2005-2014-zu-verbreitung-und-verankerung-der-
bildung-fuer-nachhaltige-entwicklung',
'un-dekade-mit-wirkung-10-jahre-bne-in-deutschland',
'zukunftsstrategie-2015-positionspapier-des-nationalkomitees-der-un-dekade',
'vom-projekt-zur-struktur-strategiepapier-der-ag-berufliche-aus-und-
weiterbildung',
'lehrmaterialien-zum-jahresthema-mobilitaet',
'deutsche-unesco-kommission-jahrbuch-2016-2017',
'roadmap-zur-umsetzung-des-weltaktionsprogramms',
'zukunftsfaeheige-kommunen-chancen-durch-bildung-fuer-nachhaltige-entwicklung',
'tagungsbericht-un-dekade-bildung-fuer-nachhaltige-entwicklung-der-beitrag-
europas',
'zwischenbilanz-zum-nationalen-aktionsplan-bildung-fuer-nachhaltige-
entwicklung',
'strukturen-staerken-ausgezeichnete-kommunen-lernorte-und-netzwerke-des-unesco-
weltaktionsprogramms-bildung-fuer-nachhaltige-entwicklung-im-portraet',
'referenzrahmen-fuer-die-fruehkindliche-bildung']
```

```
[13]: little_mallet_wrapper.print_dataset_stats(training_data)
```

```
Number of Documents: 35
Mean Number of Words per Document: 10442.0
Vocabulary Size: 42658
```

```
[ ]: num_topics = 17
```

```
[ ]: training_data = training_data
```

```
[ ]: output_directory_path = '/home/almahartwig/Dokumente/Studium-Master/SoSe-21/
↳Hausarbeit-Trends-Info/Datensatz/Training-Data/Ergebnisse-BNE'
```

```
#No need to change anything below here
```

```
Path(f"{output_directory_path}").mkdir(parents=True, exist_ok=True)
```

```
path_to_training_data = f"{output_directory_path}/training.txt"
```

```
path_to_formatted_training_data = f"{output_directory_path}/mallet.training"
path_to_model                    = f"{output_directory_path}/mallet.model.
    ↳{str(num_topics)}"
path_to_topic_keys              = f"{output_directory_path}/mallet.topic_keys.
    ↳{str(num_topics)}"
path_to_topic_distributions     = f"{output_directory_path}/mallet.
    ↳topic_distributions.{str(num_topics)}"
```

```
[ ]: little_mallet_wrapper.quick_train_topic_model(path_to_mallet,
                                                    output_directory_path,
                                                    num_topics,
                                                    training_data)
```

```
[ ]: topics = little_mallet_wrapper.load_topic_keys(path_to_topic_keys)

for topic_number, topic in enumerate(topics):
    print(f"Topic {topic_number}\n\n{topic}\n")
```

```
[ ]: topic_distributions = little_mallet_wrapper.
    ↳load_topic_distributions(path_to_topic_distributions)
```

```
[ ]: topic_distributions
```