

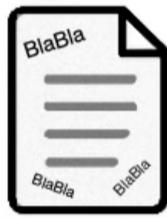


**Rauschentfernung**  
(engl. noise reduction)

**SpaCy**

- Webseite:
- Dokumentation:

**Wortbereinigung**  
(engl. word cleaning)



**Zeichenbereinigung**  
(engl. character cleaning)



**Nummernbereinigung**  
(engl. numbers cleaning)



**???**

- Webseite:
- Dokumentation:

**Normalisierung**  
(engl. normalisation)



**Standardisierung**  
(engl. standardisation)

**Rechtschreibfehlerkorrektur**  
(engl. spelling correction)



**Wortbereinigung (engl. word cleaning)**

- Stopwörter (engl. stopwords)

**SpaCy**  
• Webseite:  
• Dokumentation:  
• Anleitung:

Funktionen

**NLTK**  
• Webseite:  
• Dokumentation:  
• Anleitung:

Funktionen

**Zeichenbereinigung (engl. character cleaning)**

- Satzzeichen (engl. punctuation marks)

**SpaCy**  
• Webseite:  
• Dokumentation:  
• Anleitung:

Funktionen

**stdlib (re)**  
• Webseite:  
• Dokumentation:  
• Anleitung:

Funktionen

- Leerzeichen (engl. white space)

**SpaCy**  
• Webseite:  
• Dokumentation:  
• Anleitung:

Funktionen

**textnorm**  
• Webseite: <https://pypi.org/project/textnorm/>

Funktionen

- Sonderzeichen (engl. special character)

**SpaCy, textnorm**  
• Webseite: <https://pypi.org/project/textnorm/>

Funktionen

**stdlib (unicodedata)**  
• Webseite: <https://docs.python.org/3/library/unicodedata.html>  
• Dokumentation: <https://docs.python.org/3/howto/unicode.html#unicode-howto>  
• Anleitung:

Funktionen

**Nummernbereinigung (engl. numbers cleaning)**

- Nummern (engl. removing numbers)

**SpaCy**  
• Webseite: <https://pypi.org/project/textnorm/>  
• Dokumentation:  
• Anleitung:

Funktionen:

• .like\_num

**Normalisierung (engl. normalisation)**

- Kasusumwandlung (engl. case conversion)

**SpaCy**  
• Webseite:  
• Dokumentation:  
• Anleitung:

Funktionen

**stdlib**  
• Webseite:  
• Dokumentation:  
• Anleitung:

Funktionen

• .lower

- Formatnormalisierungen (engl. format normalisations)

**datetime**  
• Webseite:  
• Dokumentation:  
• Anleitung:

Funktionen

**Rechtschreibfehlerkorrektur (engl. spelling correction)**

**pyspellchecker**

• Webseite: <https://pypi.org/project/pyspellchecker/>  
• Dokumentation: <https://pyspellchecker.readthedocs.io/en/latest/>  
• Anleitung:

Funktionen

**contextualSpellCheck**

• Webseite:  
• Dokumentation:  
• Anleitung:

Funktionen

**PyEnchant**

• Webseite:  
• Dokumentation:  
• Anleitung:

Funktionen