2025-11-08 1/9 stylo-ah-online Handbuch

# Inhaltsverzeichnis

stylo ah online - Handbuch	2
Ursprung: stylo ah online - stylo ohne Installation	
Links zum Thema	
Implementierte Analyse-Pipeline(s)	2
Funktionsüberblick	
Benutzung	3
GUI Konzept	3
Überblick zur Benutzung	4
1. Konfiguration des Browsers	4
2. Konfigurieren der Analyseschritte	4
2.1 Arbeitsschritt: Benennung	4
2.2 Arbeitsschritt: Configurations-Datei / Datenbank	5
2.3 Arbeitsschritt: Input / Replication	5
2.4 Arbeitsschritt: Normalisierung (Normalization)	6
2.5 Arbeitsschritt: Zerlegung (features / decomposition / token)	7
2.6 Arbeitsschritt: Zählung (Selection / Counting)	8
Arbeitsschritt: Maßanwendung	9
Arbeitsschritt: Gruppierung	
Arbeitsschritt: Export	
3. Corpus-Auswahl	9
4. Erneute Berechnung und Serien	9

2025-11-08 2/9 stylo-ah-online Handbuch

# stylo ah online - Handbuch

# Ursprung: stylo ah online - stylo ohne Installation

Im Rahmen unsere Lehrtätigkeit und unserer Forschungstätigkeit haben wir das R-Paket **stylo** eingesetzt. Wir haben dieses mittels R und Python weiterentwickelt. Vornehmlich vier Dinge sind an der R Implementierung schwierig:

- 1. Die Möglichkeit zur Selbstdokumentation der Experimente ist gering.
- 2. Die Interaktion zwischen Python und R ist problematisch.
- 3. Die Möglichkeit eines einfachen Einsatzes von Multiprocessing ist fraglich.
- 4. Die Software ist teilweise schwierig zu installieren.

Daher wollte wir die Funktionalität von **stylo** nachempfinden; ohne Installation und mit erweiterten Dokumentations- und Vergleichsmöglichkeiten ausstatten.

## **Links zum Thema**

- 1. stylo das Original in R geschrieben.
- 2. stylo ah die R / Python Kopie mit zusätzlichen Funktionen für klassische Texte.
- 3. stylo ah online die originale Kopie in JavaScript implementiert.

# Implementierte Analyse-Pipeline(s)

Die Implementierte Textanalyse verläuft in sieben Schnitten. Diese stellen die Abfolge bzw. die Pipeline der Verarbeitung dar:

- 1. Auswahl der Files (Corpus) (jedes File repräsentiert einen Text)
- Normalisierung (Formatanpassung, Zeichenvereinheitlichung, Löschung von Struktur und Metadaten, Maskierung von Wortformen)
- 3. Zerlegung in Token (Wortformen, Silben, grame, Zeichen)
- 4. Zählung / Vektorbildung (01-Kodierung, abs. / rel. Häufigkeit, TF-IDF)
- 5. Auswahl aus den Vektoren (Beschränkung der Häufigkeitsliste, Häufigkeitsfenster, Culling)
- 6. Anwendung eines Maßes
- 7. Anwendung einer Cluster-Methode

## **Funktionsüberblick**

In stylo ah online sind folgende Funktionen verfügbar:

#### Normalisierung:

Normalform Wortmaskierung (Stopworte) <u>Vereinheitichen bestimmter Zeichen</u> UV-Angleich <u>JI-Angleich</u> Markup-Löschen <u>Interpunktion löschen</u> Zeilenumbrüche löschen <u>Elision auflösen</u> Alpha

2025-11-08 3/9 stylo-ah-online Handbuch

privativum behandeln <u>Entfernen der Nummerierung</u> Worttrennungen zusammenführen <u>Vereinheitlichung des Iota subscriptum und Iota adscriptum</u> Veränderung des Abschluss-Sigmas <u>Diakritische Zeichen löschen Ligaturen</u> auflösen <u>Kleinschreibung</u> Klammern entfernen

Normalisierung

#### Zerlegung (Token):

<u>Zusätzlich ohne Konsonanten</u> Zusätzlich ohne Vokale <u>Zusätzlich lediglich kleine Wörter</u> Zusätzlich lediglich große Wörter Zerlegung in Wortformen <u>Zerlegung in Buchstaben n-grame</u> Zerlegung in Buchstaben n-grame der Worformen <u>Zerlegung in Wort n-grame</u> Zerlegung in n-grame mit Lücken <u>Zerlegung in Pseudo-Silben</u> Zerlegung in Kopf Körper und Coda <u>Zerlegungen in alle Permutationen</u> von Kopf Körper und Coda

Zerlegung

#### Zählung:

absolute Häufigkeit relative Häufigkeit 0-1-Codierung TF-IDF Beschränkung/Spannen der Frequenzlisten (nach Rang, min-max-Angabe) Culling

Zählung

#### Maße:

euclidean, chebyshev, minkowski, manhatten, canberra, soerensen, gower, soergel, lorentzian, intersection, wavehedges, motyka, ruzicka, tanimoto, innerproduct, harmonicmean, cosine, kumar hasse brook, dice, fidelity, bhattacarya 1, bhattacarya 2, hellinger, jensen, jensen shannon, topsoee, kullback divergence, jeffreys, kullback leibler, squared euclidean, pearson chi squared, neyman chi squared, squared chi squared, divergence, clarck, additive symmetric chi squared, eder simple, burrows delta, argamon linear delta, eders delta, argamons quadratic delta, wasserstein 1d

Maße

#### **Gruppierung:**

hierarchische Clusterung, multidimensional scaling (MDS), tSNR (t-distributed stochastic neigbor embedding)

Gruppierung

# **Benutzung**

## **GUI Konzept**

Jeder Verarbeitungsschritt wird in einem Konfigurationsschritt eingestellt. Jeder Konfigurationsschritt besitzt ein zusätzliches Kommentarfeld und eine der Anzeige der Zwischenergebnissen (rechte Spalte). Die GUI empfindet die Schritte der Textanalyse nach. Einzige Ausnahme bildet die Auswahl des Corpus, die ist zwar an oberer Stelle angezeigt, aber von der Benutzung her ist das der letzte Schritt! Jedes GUI-Element ist mit einer Beschriftung versehen. Sofern die Funktion einer weiteren Erklärung bedarf, dann ist zusätzlich eine kursive Erklärung angefügt.

2025-11-08 4/9 stylo-ah-online Handbuch

## Überblick zur Benutzung

- Konfiguration des Browsers
- 2. Konfigurieren der Analyseschritte
- 3. Corpus-Auswahl
- 4. Erneute Berechnung und Serien

## 1. Konfiguration des Browsers

Speicherort: Legen sie einen Ordner an in welchem die Ergebnisse von stylo-ah-online gespeichert werden können. Konfigurieren Sie ihren Browser so, dass er die Dowloads in diesem Ordner ablegt. die Anleitung für Firefox:

https://support.mozilla.org/de/kb/suchen-und-verwalten-heruntergeladener-dateien#w\_ziel-ordner-der-heruntergeladenen-dateien-andern

Datenbanken: Teilen sie dem Browser mit, dass er Datenbanken anlegen und nicht löschen soll. Dazu stellen sie sicher, dass die Chronik angelegt wird:

https://support.mozilla.org/de/kb/firefox-chronik-zeigt-ihre-besuchten-webseiten

Web-Konsole: stylo-ah-online ist in JavaScript geschrieben. Die Web-Konsole bietet eine Darstellung von Meldungen zum Programmablauf und Hinweise auf Fehler. Nach einem Fehler können sie die Seite neu laden und den Ablauf erneut starten. Es bietet sich an die Web-Konsole anzuzeigen. Für Firefox aktiviert man diese so:

https://firefox-source-docs.mozilla.org/devtools-user/tools\_toolbox/index.html

## 2. Konfigurieren der Analyseschritte

#### 2.1 Arbeitsschritt: Benennung

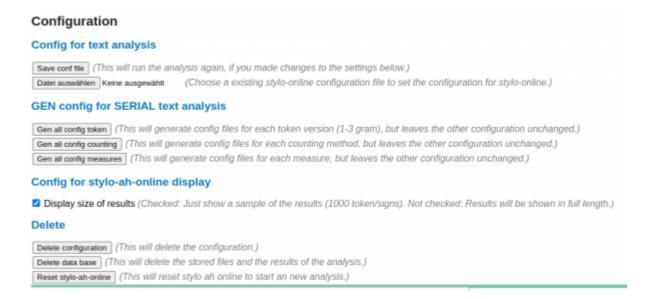
experiment 🕶	Type (Select/give a type, if this is not given by file ending.)
default	Subject (Name the subject of the file.)
unfinished 🕶	State (Select/give a state, if this is not given by file ending.)
	ID (Give the ID of the file.)
2023-07-14	Date (The actual date, erase to reset.)
	Version (Choose a version of the file.)
default	Author name (Fill in the name of the author.)
	File ending (Provide a file ending.)
experiment_default_unfinis	shed_2023-07-14_default.styloahonline
Log entry of naming se	action )

Das Benennungsmodel beinhaltet zusammengesetzte Namen für die Experimente. Dabei werden die einzelnen Teile der Benennung durch Unterstriche abgetrennt. Leerzeichen in Einträgen werden durch Bindestriche ersetzt. Die einzelnen Einträge der Benennung erlauben verschiedene

2025-11-08 5/9 stylo-ah-online Handbuch

Organisationsmuster für Experimente. Das Datum wird beim ersten Programmlauf automatisch ausgefüllt. Der einmal erstellte Name charakterisiert den Programmablauf und die Einstellungen genau. Änderungen in der Benennung führt auf ein neues Experiment, welches auch separat gespeichert wird. Die Benennung geht in die Config-Datei ein und ein erneuter Aufruf einer Config-Datei stellt alles unter dieser Bezeichnung wieder her. Auf diese Weise können mehrer Konfigurationen unterschieden werden.

#### 2.2 Arbeitsschritt: Configurations-Datei / Datenbank



Im Abschnitt "Configuration" können alle Einstellungen vorgenommen werden, die sich auf die Config-Datei beziehen. stylo-ah-online erlaubt die Speicherung der Konfiguration und genauso ihren erneuten Aufruf durch Eingabe einer Config-Datei. Für die Erstellung von Serien von Experimenten erlaubt es stylo-ah-online mehrer Config-Dateien zuerstellen und als Serie wieder zu öffnen. Die Ergebnisse werden für jede so gespeicherte Einstellung unter einem extra Namen abgespeichert. Für umfangreichere Corpora ist es nötig die Darstellung von Zwischenergebnissen zu beschränken. Das kann an dieser Stelle mit dem Haken bei "Display size of results" erzielt werden. stylo-ah-online verfügt über eine Auto-save-Funktion, außerdem werden die eingegebenen Daten gespeichert, um diese bei einem erneuten Analysedruchlauf aus der Browser eigenen Datenbank zu holen. Man kann hier die Einstellungen und die Datenbanken oder beides zurücksetzen.

#### 2.3 Arbeitsschritt: Input / Replication

2025-11-08 6/9 stylo-ah-online Handbuch

Input / Replication
Dateien auswählen Keine ausgewählt (Just choose the CORPUS FILES, than the selection below will be applyed (start analysis). Data in the data base will be overwritten.)  Re-run anaylsis (This will RUN the analysis AGAIN, if you made changes to the settings below. Data is taken from the data base.)  Dateien auswählen Keine ausgewählt (Choose MULTIPLE config files to perfom multiple analysis on one corpus.)
Note
(Log entry for the input section.)

Im Abschnitt "input / Replication" geht es um den Aufruf der Textdateien, die analysiert werden sollen. Diese Handlung löst die erste Analyse aus und speichert Ergebnisse in der Datenbank. Mit "rerun" kann man eine veränderte Konfiguration auf die vorhandenen Daten anwenden. Sollten Serienexperimente vorbereitet wurden sein, dann kann man die dazugehörigen Config-Files hier eingeben und so die Serienverarbeitung auslösen. Dazu werden die Daten aus der Datenbank verwendet. Der Arbeitsschritt der Dateieingabe muss der Letze sein, nachdem alle anderen Konfigurationen vorgenommen wurden.

### 2.4 Arbeitsschritt: Normalisierung (Normalization)

2025-11-08 7/9 stylo-ah-online Handbuch

### NORMALIZATION

(Please check <a href="http://ecomparatio.net/~khk/NORM-DECOMP-DIST/textnorm.html">http://ecomparatio.net/~khk/NORM-DECOMP-DIST/textnorm.html</a> to see some examples of how the selection would work.)

Word masking / stop words
Use Word masking (Give back the string without stop words.)  Show stop word list (Check this to apply stop word removal.)  Datei auswählen Keine ausgewählt (Choose a existing stop word file (CSV format, divider: ;;).)
Sign equalization
<ul> <li>✓ Disambiguate discritica</li> <li>✓ Disambiguate dashes</li> <li>☐ Text output latin u-v (repaces all u with v)</li> <li>☐ Text output latin j-i (repaces all j with i)</li> <li>☐ lota sub to ad (takes greek utf8 string and repleces iota subscriptum with iota ad scriptum)</li> <li>☐ Text output tailing sigma uniform (equalize tailing sigma)</li> <li>☐ Text output without discritics (replaces discritics)</li> <li>☐ Text output without some signs (delete some to the programmer unknown signs: †, *, &lt;, #, §, *)</li> <li>☐ Text output without ligature (takes a string, return string with ligatures turned to single letters)</li> <li>✓ Text output equal case (input a string and get it back with all small case letters)</li> <li>✓ Text output no brackets (input string and get it back with no brackets)</li> </ul>
Markup / Format
<ul> <li>✓ Without markup (input a string and get it back with markup (html / xml) removed)</li> <li>✓ Delete punctuation (takes string and returns the string without punctuation)</li> <li>✓ Without newline (input string and get it back with linebreaks removed)</li> </ul>
Word level conversions
<ul> <li>□ Elision expansion (elusion it will be expanded)</li> <li>□ Alpha privativum / copulativum (takes utf8 greek and splits the alpha privativum and copulativum fron wordforms)</li> <li>□ Text output without numbering (takes string, return string without the edition numbering i.e. [2])</li> <li>☑ Text output no hypenation (removes hyphenation)</li> </ul>
Combinations
✓ (Select one of the combined normalization functions (none of the single steps is used).)
Translitteration
✓ (Select one of the transliterations.)

2.5 Arbeitsschritt: Zerlegung (features / decomposition / token)

2025-11-08 8/9 stylo-ah-online Handbuch

## FEATURES / DECOMPOSITION / TOKEN

(The word level decomposition and the gram decomposition will be combined. Check <a href="http://ecomparatio.net/~khk/NORM-DECOMP-DIST/zerl.html">http://ecomparatio.net/~khk/NORM-DECOMP-DIST/zerl.html</a> for some examples to see how decomposition will work.)

decomposition will work.)	NOTAM DEGOME - DIGITZEN. Hum	
Word level decompo	osition	
General N-Gram dec	omposition	
Word level (groups of words)	✓ Gram-level	
1	N (gram-size, set to one means for example word statistics)	
1	M (gap-size, for gap n-gram	
▼ Padding (used for	sign level of words)	
Note		
I on ontry for the taken co	notion 1	
SELECTION / CO	UNTING ct the counting methode)	
Most frequent token	/ words (per text)	
0	Min value (position in frequency ordered list)	
100	Max value (position in frequency ordered list)	
Culling (per corpus)		
	Min value (per cent of presents of a token in all texts)  Max value (per cent)	
Text length normaliz	eation	
□ compare fractions of te	exts (smallest text gives the length; none other methode is applied)	
Note		

2025-11-08 9/9 stylo-ah-online Handbuch

**Arbeitsschritt: Maßanwendung** 

**Arbeitsschritt: Gruppierung** 

**Arbeitsschritt: Export** 

## 3. Corpus-Auswahl

1. Aufruf von

## 4. Erneute Berechnung und Serien

From:

http://replicatio.science/dokuwiki/ - documentatio replicationis

Permanent link:

http://replicatio.science/dokuwiki/doku.php/de/styloahonline/handbuch?rev=1699468061

Last update: **2023-11-08** 

