

Inhaltsverzeichnis

stylo ah online - Handbuch	2
<i>Ursprung: stylo ah online - stylo ohne Installation</i>	2
<i>Links zum Thema</i>	2
<i>Implementierte Analyse-Pipeline(s)</i>	2
<i>Funktionsüberblick</i>	2
<i>Benutzung</i>	3
GUI Konzept	3
Überblick zur Benutzung	4
Die einzelnen Schritte der Textanalyse benutzen	4
Arbeitsschritt: Browser Voreinstellungen	4
Arbeitsschritt: Benennung	4
Arbeitsschritt: Configurations-Datei / Datenbank	4
Arbeitsschritt: Input	4
Arbeitsschritt: Normalisierung	4
Arbeitsschritt: Zerlegung	4
Arbeitsschritt: Zählung	4
Arbeitsschritt: Maanwendung	4
Arbeitsschritt: Gruppierung	4
Arbeitsschritt: Export	4
Serien / Replikation	4
Browser	5

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 Schritten. Diese stellen die Abfolge bzw. die Pipeline der Verarbeitung dar:

1. Auswahl der Files (Corpus) (jedes File repräsentiert einen Text)
2. 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 \(Stopworte\)](#) [Vereinheitlichen bestimmter Zeichen](#) [UV-Angleich](#) [JI-Angleich](#) [Markup-Löschen](#) [Interpunktion löschen](#) [Zeilenumbrüche löschen](#) [Elision auflösen](#) [Alpha](#)

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

Normalisierung

Zerlegung (Token):

Zusätzlich ohne Konsonanten Zusätzlich ohne Vokale Zusätzlich lediglich kleine Wörter Zusätzlich lediglich große Wörter Zerlegung in Wortformen Zerlegung in Buchstaben n-gramme Zerlegung in Buchstaben n-gramme der Wortformen Zerlegung in Wort n-gramme Zerlegung in n-gramme mit Lücken Zerlegung in Pseudo-Silben Zerlegung in Kopf Körper und Coda Zerlegungen in alle Permutationen 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, manhattan, canberra, soerensen, gower, soergel, lorentzian,
intersection, wavehedges, motyka, ruzicka, tanimoto, innerproduct, harmonicmean, cosine, kumar
hasse brook, dice, fidelity, bhattacharya 1, bhattacharya 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 neighbor 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.

Überblick zur Benutzung

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

Die einzelnen Schritte der Textanalyse benutzen

Arbeitsschritt: Browser Voreinstellungen

Speicherort, Web-Konsole, RAM

Arbeitsschritt: Benennung

1. das Benennungsmodell
2. Benennungen in der Config-Datei
3. Import Export der Config-Datei
4. mehrerer Konfigurationen speichern
5. ...

Arbeitsschritt: Configurations-Datei / Datenbank

Arbeitsschritt: Input

Arbeitsschritt: Normalisierung

Arbeitsschritt: Zerlegung

Arbeitsschritt: Zählung

Arbeitsschritt: Maanwendung

Arbeitsschritt: Gruppierung

Arbeitsschritt: Export

Serien / Replikation

1. Aufruf von

Browser

Welche Browser werden unterstützt?

From:
<http://replicatio.science/dokuwiki/> - **documentatio replicationis**

Permanent link:
<http://replicatio.science/dokuwiki/doku.php/de/styloahonline/handbuch?rev=1699461856>

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

