

OneTIPP Algorithmus

Version: 13.11.2015 v2.4

Bearbeiter: Sebastian Enger

1. POS-Tagging des Eingabetextes
 - a. Benutze bzw teste folgende POS-Tagger:
 - i. HunPos: <http://www.how2015.net/questions/5075725/how-to-correctly-set-hunpos-tagger-in-nltk-for-pos-tagging-in-english>
 - ii. Stanford POS: <http://textminingonline.com/tag/nltk-stanford-pos-tagger>
 - iii. Senna: <http://www.nltk.org/modules/nltk/tag/senna.html>
 - iv. TreeTagger: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
 1. ALS STTS Modul ansprechen
 - v. ParZu: <https://github.com/rsennrich/parzu>
 - vi. Spacy: <https://github.com/honnibal/spaCy>
2. Definiere Logiken zum Zusammenfassen und für den Austausch:
 - a. Zum Beispiel: „Ich lief schnell“ (PP-VV-ADJ)
 - b. Zum Beispiel: „Ich aß Eis“ (PP-VV-NN)
 - c. Zum Beispiel:
 - i. Nomen nur mit Nomen tauschen
 - ii. Adverb nur mit Adverb tauschen
 - iii. VV-NN-ADJ → ADV NN VV ADJ (nur Beispiel)
 - d. Beim Austausch von Wort (NN) mit einem Synonym oder Akronym baue nach dem Satz einen Füllsatz, passend zum Thema, mit dem originalen Wort drin ein.
 - i. Dies maximal 1 mal pro 5 Sätze
3. Definiere Logiken, von Worten, die nicht ausgetauscht werden soll:
 - a. ADJA - attributives Adjektiv
 - b. ADV - Adverb
 - c. APZR - Zirkumposition rechts
 - i. Und weitere andere
 - d. Zahlwörter (pos-tag=CARD) ausschreiben: <https://pypi.python.org/pypi/num2words>
 - e. Fremdwörter (pos-tag=FM) ins Deutsche Übersetzen (Version 1) und austauschen (Version 2)
4. Sinnhafte Kernelemente jedes Satzes herausfinden
 - a. Kernelemente sind Sinn-tragende Hauptelemente
 - i. Zum Beispiel: „ich [aß Eis]“
 - ii. Zum Beispiel: „ich [ging spazieren] und [traf Harry Potter]“
 - iii. Logiken für Kernelemente bestimmen:
 1. Zum Beispiel: „Verb Nomen“, „Verb Adjektiv“, „Verb Adjektiv Nomen“
 - b. Sinnhafte Kernelemente werden passend Sinnverwand getauscht (Synonym, Antonym):
 - i. Zum Beispiel: „ich [aß ein Eis]“
 - ii. Kernelement: VERB-POSSESSIVPRONOMEN-NOMEN
 - iii. Bestimme von „ich [aß ein Eis]“:
 1. Zeitform („Präteritum“)
 2. Person („1. Person Singular“)
 3. Geschlechterbestimmung eines/des folgenden NOMEN
 - a. [EIS] → männlich
 - iv. Austauschverhalten:
 1. Nimm ein Verb aus der Synonym Datenbank und bringe es in die Zeitform, Person aus (Punkt 3. b. iii.)
 2. Nimm ein Nomen aus Synonym DB und passe bzw. bringe es zu dem Geschlecht passend aus (Punkt 3. b. iii. 2.) dar

3. Adverb nur mit Adverb tauschen
4. Nomen nur mit Nomen tauschen
5. Adjektiv nur mit Adjektiv tauschen

5. Qualitätstests nach jedem Austausch:

- a. Pos-tagge jeden Ausgabesatz
- b. Der „Pos-getaggte“ Ausgabesatz muss sich an den definierten Logiken für „Sinnhafte Kernelemente“ halten
- c. Der „Pos-getaggte“ Ausgabesatz muss sich an den definierten Logiken für „Zusammenfassen und für den Austausch“ aus Punkt 2. Halten
- d. Hidden Markov Model (HMM) zur Berechnung der Wahrscheinlichkeiten eines Ausgabesatzes -> Vergleich mit Ursprungssatz -> wenn zu Grosse Differenz: Neu manipulieren (<http://www3.cs.stonybrook.edu/~ychoi/cse628/lecture/06-hmm.pdf>)
- e. Nimm Lesbarkeitsstatistik von Ein und OneTIPP Ausgabe -> wissenschaftliche Texte bleiben wissenschaftlich und umgangssprachliche bleiben gleiches Niveau

Ortsnamen: <http://www.ling.uni-potsdam.de/~kolb/download.html>

Namensbestimmung: <http://www.heise.de/ct/ftp/07/17/182/>

Todo: unsere Synonymdatenbank POS-TAGGEN, dazu den passenden POS Tagger finden.