NLP opdracht minor data analyse (IMINDAA)

De NLP opdracht is simpel. Kies een tekst corpus en doe daar een NLP analyse op. Je laat zien met deze analyse dat je NLP technieken beheerst en kunt toepassen. Waar je op beoordeeld wordt kun je terugvinden in het nakijkmodel. Het kiezen van een tekst corpus ben je geheel vrij in. Hieronder volgt een lijst met suggesties:

Pubmed

<https://pubmed.ncbi.nlm.nih.gov>

Project Gutenberg

<https://www.gutenberg.org>

Harry Potter: <https://github.com/amephraim/nlp/blob/master/texts/J.%20K.%20Rowling%20-%20Harry%20Potter%201%20-%20Sorcerer%27s%20Stone.txt>

Bijbel: <https://www.o-bible.com/download/kjv.txt>

Plenaire verslagen tweede kamer:

<https://www.tweedekamer.nl/kamerstukken/plenaire_verslagen>

Wikipedia: <https://en.wikipedia.org/wiki/Wikipedia:Database_download>

Zoekterm op datasets:

<https://odsc.medium.com/20-open-datasets-for-natural-language-processing-538fbfaf8e38>