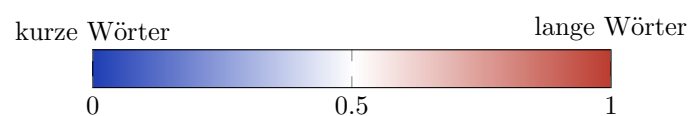


Kriterien für die Lesbarkeitsanalyse

Wortlänge

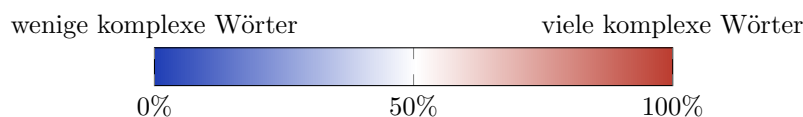
Hierfür wird zunächst die durchschnittliche Wortlänge analysiert und normiert. Sei W die Menge aller Wörter w_i im zu analysierenden Text mit Wortlänge $wl_i = |w_i|$. Die minimale Wortlänge ist 1 (bzw. 2 im Deutschen), die maximale ist $l_{max} = \max(|w_i|)$ bzgl. aller Wörter $w_i \in W$.

Der Lesbarkeitswert jedes Wortes wird normiert durch $\frac{|w_i|}{l_{max}}$ und der summierte Wert des entsprechenden (Ab-)satzes durch die Anzahl der Wörter $|W|$ geteilt. Anschließend wird der Wert z.B. auf Farbwerte zwischen blau (32, 62, 181), weiß und rot (186, 57, 44) abgetragen.



Komplexität der Vokabeln

Hier wird der Prozentanteil eines Absatzes/Satzes gemessen, der nicht in einer Liste häufig verwendeter Wörter vorkommt. Dazu kann entweder Wikipedia¹ (deutsch/englisch), ein Korpus aus Zeitungsartikeln² oder evtl. eine fachspezifische Textsammlung ausgewertet werden. Der Anteil der Wörter w_i , die nicht in der Liste L sind, wird dann durch die Anzahl $|W|$ der Wörter im zu analysierenden Text W geteilt, also $\text{Komplexität}_W = \frac{|w_i \notin L|}{|W|}$ in Prozent.



Nominalisierungen

Satzlänge

Komplexität der Satzstruktur

¹https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists#German

²<http://wortschatz.uni-leipzig.de/html/wliste.html>