

Übung 12

Worthäufigkeiten

Themen: *set*, *defaultdict*, *requests*, *Zeitmessung*, *Effizienz*, *matplotlib*

Schreiben Sie ein Programm, welches die Worthäufigkeiten aus *Alice in Wonderland* extrahiert und graphisch darstellt.

Schritte:

- Laden Sie eine Stoppwortliste mit dem Modul `requests` von folgender Adresse herunter: <https://tinyurl.com/evpxs6u8>
- Speichern Sie die Stoppwörter in einer Liste.
- Lesen den Text von *Alice in Wonderland* ein.
- Ersetzen Sie Großbuchstaben durch Kleinbuchstaben.
- Entfernen Sie Satzzeichen, Klammern etc.
- Spalten Sie den Text in Wörter auf mit Whitespace-Symbolen als Wortgrenzen.
- Entfernen Sie aus der erhaltenen Wortliste die Stoppwörter.
- Messen Sie mit `timeit.timeit`, wie lange das Ausfiltern dauert.
- Speichern Sie dann die Stoppwörter in einem Set.
- Messen Sie noch einmal, wie lange das Ausfiltern dauert.
- Berechnen Sie die Worthäufigkeiten.
- Sortieren Sie die Wörter absteigend nach Häufigkeit und plotten Sie die Häufigkeiten mit der `matplotlib`-Bibliothek.