

#163a - Wortlängenanalyse

Bei diesem Beispiel geht es darum, die Buchstabenlängen deutscher Wörter in beliebigen deutschen Texten mit einer Standardverteilung zu vergleichen.

Vorbereitungen: Lege Dir zumindest 5 Textfiles an in denen sich möglichst lange deutsche Texte befinden, speichere diese in einem bestimmten Verzeichnis Deiner Wahl als .txt-Files ab.

Aufgabe 1

Erstelle nun für alle Files Threads, mit denen die Texte parallel eingelesen werden und von jedem Wort die Anzahl der Zeichen zählt. Zähle mit, wie oft welche Wortlänge vorkommt. Relevant für die Aufgabe sind Wörter mit einer Länge von 2 bis 23 Buchstaben.

Verwende zum Auffinden aller Files eine der **System.IO.Directory.EnumerateFiles**-Methoden um ein File nach dem anderen zu durchlaufen.

Beachte: Sonderzeichen wie Punkt, Beistrich und Strichpunkt zählen natürlich nicht als Zeichen von Wörtern und dürfen nicht mitgezählt werden.

Aufgabe 2

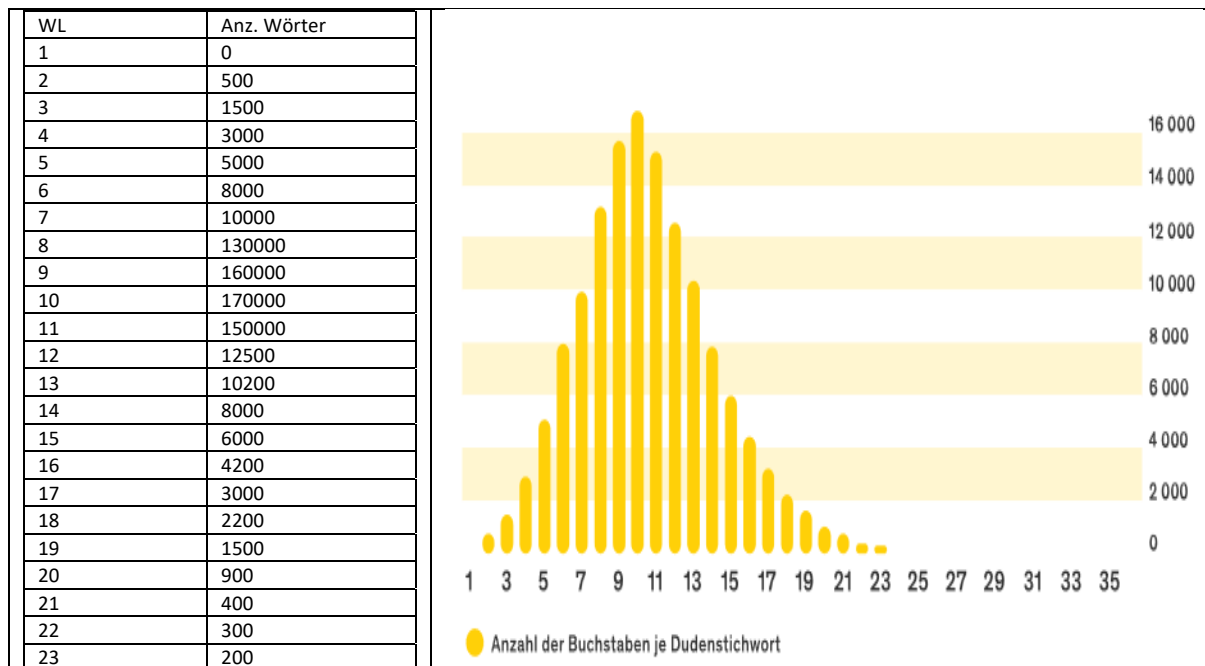
Führe die Ergebnisse der einzelnen Textdateien zusammen. Das bedeutet, es soll ermittelt werden, wie oft Wörter mit einer bestimmter Wortlänge in allen Texten vorkommt.

Gib das Ergebnis wie folgt auf der Konsole aus:

Text <1-4>: <Anzahl Wörter mit 2 Buchstaben> <Anzahl Wörter mit 3 Buchstaben> usw.

Aufgabe 3

Es gibt eine Statistik aus dem Rechtschreibduden, in welcher die Buchstabenlänge aller darin enthaltener Worte erfasst wurden.



<https://www.duden.de/sprachwissen/sprachratgeber/Durchschnittliche-Lange-eines-deutschen-Wortes>

Erstelle eine CSV-Datei, in welcher diese Werte, dem Ergebnis deiner Werte gegenübergestellt werden. Dabei sollen allerdings nicht die absoluten Zahlen herangezogen werden, sondern die prozentuelle Verteilung.

Erstelle anhand dieser CSV-Datei zwei Diagramme mittels Excel (muss nicht programmiert werden!) um das Ergebnis augenscheinlich zu beurteilen. Informiere Dich dabei, wie man in Excel csv-Dateien importiert.

Deliverables:

Results.docx: Screenshots der Konsole nach der Programmausführung und des Exceldiagramms

Results.xlsx mit dem importierten CSV-Daten und dem Excel-Diagramm