

Aufgabenblatt 12: Dynamische Datenstrukturen

Kapitel 1: Praktikumsaufgaben

12.1.1: Textanalyse (Liste) ✓

Erstellen Sie ein Programm, das in einer Textdatei die 20 am häufigsten vorkommenden Wörter findet, indem es eine Liste aller verschiedenen Wörter aufbaut und in dieser Liste zählt, wie oft das jeweilige Wort bereits vorgekommen ist.

Sie können davon ausgehen, dass in der Datei alle Umlaute und auch 'ß' durch Zeichen des englischen Alphabets ersetzt sind (ä->ae, ö->oe, ü->ue, ß->ss).

Im Anhang finden Sie eine Datei (Buddenbrocks.txt), die den vollständigen Text des Romans "Die Buddenbrocks" von Thomas Mann enthält, und in der bereits die erforderlichen Ersetzungen durchgeführt sind.

Bei der Analyse dieser Datei findet mein Programm die folgenden Worthäufigkeiten:

Der Text enthaelt insgesamt 227956 (24918 verschiedene) Woerter
Die 20 am haeufigsten vorkommenden Woerter sind:

und	9582 (4.20 Prozent aller Woerter)
die	5716 (2.51 Prozent aller Woerter)
der	4906 (2.15 Prozent aller Woerter)
er	3720 (1.63 Prozent aller Woerter)
in	3376 (1.48 Prozent aller Woerter)
mit	3237 (1.42 Prozent aller Woerter)
zu	3215 (1.41 Prozent aller Woerter)
sie	3177 (1.39 Prozent aller Woerter)
das	2593 (1.14 Prozent aller Woerter)
sich	2590 (1.14 Prozent aller Woerter)
den	2414 (1.06 Prozent aller Woerter)
es	2224 (0.98 Prozent aller Woerter)
ich	2205 (0.97 Prozent aller Woerter)
nicht	2153 (0.94 Prozent aller Woerter)
ein	2137 (0.94 Prozent aller Woerter)
war	2048 (0.90 Prozent aller Woerter)
von	1685 (0.74 Prozent aller Woerter)
auf	1639 (0.72 Prozent aller Woerter)
dem	1569 (0.69 Prozent aller Woerter)
dass	1459 (0.64 Prozent aller Woerter)

Diese Woerter machen 27.04 Prozent der Woerter des Textes aus.

Siehe auch: http://de.wikipedia.org/wiki/Liste_der_h%C3%A4ufigsten_W%C3%B6rter_der_deutschen_Sprache

Feedback

Du hast am 22.12.16, 20:38 folgendes Feedback zu dieser Aufgabe abgegeben:

bearbeitet: komplett

Schwierigkeit: 7/10

Spaß: 10/10

Zeit: 45min

12.1.2: Textanalyse (Baum) ✓

Erstellen Sie eine modifizierte Version des Programms Textanalyse (Liste) , indem Sie statt einer Liste einen Baum verwenden.

Versuchen Sie die erforderlichen Operationen auf dem Baum teils mit und teils ohne Rekursion zu implementieren.

Messen Sie die Tiefe des sich ergebenden Baums, um ein Gefühl für die Effizienz der Baumoperationen im Vergleich zu den Listenoperationen zu bekommen.

Feedback

Du hast am 22.12.16, 23:41 folgendes Feedback zu dieser Aufgabe abgegeben:

bearbeitet: komplett

Schwierigkeit: 5/10

Spaß: 10/10

Zeit: 15min

Text: einfach, da nur die Datenstruktur geändert werden muss.

Kapitel 2: Vertiefung und Selbsttest

12.2.1: Textanalyse (Hashing)

Erstellen Sie eine weitere Variante der Textanalyse (Baum), indem Sie eine Hashtabelle verwenden.

12.2.2: Textanalyse (Vergleich)

Vergleichen Sie Ihre drei Implementierungen zur Textanalyse

- Textanalyse (Liste)
- Textanalyse (Baum)
- Textanalyse (Hashing)

bezüglich Ihrer Laufzeit. Verwenden Sie dazu einen Profiler.