

Übungsblatt 4

Abgabe bis Dienstag, den 23. Mai um 12:00 Uhr

Aufgabe 1 (10 Punkte)

Implementieren Sie eine Klasse *HashMap* mit Methoden für *insert* (Einfügen bzw. Verändern eines Eintrages) und *lookup* (Nachschlagen ob ein Eintrag enthalten ist und wenn ja mit welchem Wert). Eine genaue Spezifikation der beiden Methoden finden Sie in der Datei *HashMap.TIP* auf dem Wiki. Ihr Design *muss* objekt-orientiert sein. Insbesondere: auch wenn Sie Python benutzen, sollten Sie also die o.g. Funktionen im Rahmen einer Klasse implementieren!

Sie können annehmen, dass die Schlüssel vom Typ *string* und die Werte vom Typ *int* sind. Sie können außerdem annehmen, dass die Größe der Hashtabelle bei der Erzeugung des Objektes gegeben ist und sich dann nicht mehr ändert.

Zur Auflösung von Kollisionen können Sie Verkettung oder offene Adressierung benutzen. Wenn Sie möchten und sich mit dem Vorlesungsstoff sehr sicher fühlen, können Sie auch Cuckoo Hashing implementieren.

Aufgabe 2 (10 Punkte)

Schreiben Sie ein Programm namens *WordCount* bzw. *word_count*, das Ihre *HashMap* aus Aufgabe 1 benutzt, um in einer gegebenen Textdatei die Anzahl Vorkommen (Häufigkeiten) der einzelnen Worte zu zählen und nach Häufigkeit sortiert auszugeben. Eine genaue Spezifikation zusammen mit einem Unit Test finden Sie in der Datei *WordCount.TIP* auf dem Wiki.

Lassen Sie Ihr Programm auf der auf dem Wiki verlinkten Textdatei laufen. Plotten Sie die 500 größten Worthäufigkeiten (in sortierter Reihenfolge) in einem Schaubild. Das Schaubild sollte ebenfalls Teil der Abgabe sein, wie beim ÜB1. Was Sie sehen sollten, ist das sogenannte Zipfsche Gesetz: die Worthäufigkeiten in einem (fast) beliebigen Textdokument in einer (fast) beliebigen Sprache verhalten sich ungefähr so wie die Funktion $1/x$ bzw. $1/x^\alpha$ für eine Konstante $\alpha > 0$.

Hinweis: Weil wir so nett sind, finden Sie auf dem Wiki in allen drei Sprachen Code zum Einlesen einer Textdatei und Zerlegen des Textes in Worte.

[erst dran glauben und dann wenden]

Committen Sie Ihren Code (samt Unit Tests) und die Schaubilder in das SVN, in einen neuen Unterordner *blatt-04*. Nennen Sie den Unterordner nicht *Blatt-04* und auch nicht *blatt_04* und auch nicht *blatt-4* oder *blatt-vier*. Er soll wirklich *blatt-04* heißen, so wie es im ersten Satz steht.

Erzählen Sie uns wie gehabt in Ihren *erfahrungen.txt*, wie lange Sie ungefähr gebraucht haben und ob es Probleme gab und wenn ja, welche.

Was sind die Ursachen von Mathe-Phobie?