

## Übungsblatt 3

Abgabe bis Dienstag, den 22. November um 16:00 Uhr

### Aufgabe 1 (15 Punkte)

Auf dem Wiki finden Sie einen Link zu einem Auszug aus dem *Query Log* einer großen Websuchmaschine (3,558,412 Anfragen).

In der Vorlesung haben wir zusammen die Klasse *QueryLogAnalyzer* geschrieben, die in einem gegebenen Query Log die Anzahl der Vorkommen jedes vorkommenden Anfragestrings zählt (Methode *computeQueryFrequenciesSimple*), und die  $k$  häufigsten Anfragen und deren Häufigkeit zurückgibt, für ein gegebenes  $k$  (Methode *computeMostFrequentQueries*).

Erweitern Sie die Klasse um zwei Methoden *computeQueryFrequenciesUsingSorting* und *computeQueryFrequenciesUsingMap*, die beide dasselbe Ergebnis berechnen sollen wie das in Vorlesung geschriebene *computeQueryFrequenciesSimple*, aber einmal mit Hilfe von Sortieren und einmal mit Hilfe eines assoziativen Arrays (nehmen Sie in Java *HashMap*, in C++ *hash\_map*). Schreiben Sie selbstverständlich einen (guten) Test für die beiden neuen Methoden.

Schreiben Sie darüber hinaus ein Programm *QueryLogAnalyzerMain*, das die Laufzeit der beiden neuen Methoden für verschiedene Dateigrößen vergleicht (ohne die Zeit für *computeMostFrequentQueries*), und erstellen Sie ein Schaubild mit beiden Laufzeitkurven. Verschiedene Dateigrößen bekommen Sie zum Beispiel dadurch, indem Sie von der gegebenen Datei einfach nur die ersten  $x$  Zeilen lesen.

### Aufgabe 2 (5 Punkte)

Committen Sie Ihre Lösung in das SVN, den Code in einem neuen Unterordner *src/uebungsblatt\_3* und das Schaubild in einem neuen Unterordner *non-code/uebungsblatt\_3*. Schauen Sie, dass auf Jenkins (unserem Build-System, siehe Link auf Ihrer Daphne Seite) alles ohne Fehler durchläuft.

Schreiben und committen Sie außerdem wie gehabt (in *non-code/uebungsblatt\_3*) eine Datei *erfahrungen.txt*, in der Sie Ihre Erfahrung mit dem Übungsblatt und, wenn Sie möchten, mit der Vorlesung beschreiben, insbesondere in Bezug auf den Zeitaufwand und etwaige Verständnisprobleme.