

Übungsblatt 12

Abgabe bis Dienstag, den 25. Juli um 12:00 Uhr

Aufgabe 1 (15 Punkte)

Implementieren Sie den Algorithmus von Karp-Rabin, wie in der Vorlesung erklärt. Der Algorithmus sollte alle Vorkommen von k gegebenen Mustern (jedes der Länge m) in einem gegebenen Text der Länge n finden.

Der Algorithmus soll dabei nach allen Mustern „auf einmal“ suchen, und nicht erst nach allen Vorkommen von einem Muster, dann nach allen Vorkommen von dem nächsten Muster, usw. Entsprechend sollte die Laufzeit im „typischen“ Fall $O(n + m \cdot \#matches)$ sein, und nicht $O(k \cdot n)$. Das alles wurde in der Vorlesung erklärt.

Hinweis: Implementieren Sie den Algorithmus zunächst für ein einzelnes Muster (10 Punkte) und verallgemeinern Sie ihn dann, so dass er auch für eine Liste von Mustern funktioniert (5 Punkte).

Aufgabe 2 (5 Punkte)

Auf wem Wiki finden Sie den Text einer im Jahr 2010 an einer deutschen Universität akzeptierten Informatik-Dissertation (*phd-thesis.txt*), sowie 288 Fragmente der Länge $m = 150$ aus einer in dieser Dissertation zitierten Vorarbeit von anderen Autoren (*fragments.txt*). Zum Nachlesen (bei Interesse) finden Sie auch die zugehörigen PDF Dateien auf dem Wiki.

Benutzen Sie Ihren Code aus Aufgabe 1, um alle Vorkommen dieser Fragmente (ein Fragment pro Zeile) in der Dissertation zu finden und auszugeben. Was sagt uns das über diese Dissertation?

Bemerkung: Wenn Sie Probleme mit Aufgabe 1 haben, können Sie für Aufgabe 2 auch den Code für den Knuth-Morris-Pratt Algorithmus aus der Vorlesung benutzen (er ist in Python, Java und C++ verfügbar).

Unit Tests, SVN, Jenkins, Erfahrungen wie gehabt + ein letztes Mal für dieses Semester.

Seit wann liest Donald Knuth keine E-Mails mehr und warum?