

2. Aufgabenblatt zum Softwarepraktikum “Muster- und Bilderkennung” im Grundstudium

Aufgabe 1:

Schreiben Sie ein Programm, das einen Text aus einer Datei einliest, ein einfaches Lexikon erstellt und eine Suffix-Array-Struktur auf Basis der Einträge im Lexikon bildet und initialisiert.

Aufgabe 2:

Sortieren Sie das Suffix-Array nach dem in der Vorbesprechung gegebenen Muster und schreiben Sie eine einfache Suchroutine, die feste Wortfolgen im Text unter Verwendung des sortierten Suffix-Arrays auffindet.

Beim Sortieren sollen nicht nur die einzelnen Zeichen des Textes sortiert werden, sondern alle Zeichenfolgen, d.h. wenn 20 mal Zeichen 'a' vorkommt, dann werden die 20 'a'-s mit Hilfe ihrer Nachfolger sortiert. Hat man dann z.B. noch 5 mal 'ab' gefunden, sortiert man diese wiederum nach deren Nachfolgebuchstaben.

Ausgabe der Suche zu jeder Anfrage ist jeweils die Position des ersten Auftretens im Corpus und die Anzahl der Vorkommen.

Aufgabe 3:

Schreiben Sie eine verbesserte Suchfunktion, die die Eigenschaften der Sortierung des Suffix-Arrays ausnutzt. Vergleichen Sie die Laufzeiten der beiden Suchverfahren mit einer einfachen (linearen) Suche auf den gegebenen Anfragen.

Die Testdateien und Anfragen finden Sie auf der Webseite des Praktikums.

Abgabe:

Abnahmetermin: **Donnerstag, 3. Mai**
Abnahmeort: **CIP-Pool**

Bitte senden Sie den Quellcode der Lösungen **spätestens einen Tag vor Abnahme** per e-mail an die Betreuer **{mauser,vilar}@i6.informatik.rwth-aachen.de**. Achten Sie auch darauf, daß Ihre Programme unter Linux kompilierbar sind, bei keiner Eingabe abstürzen, jeden Bedienungsfehler mit einer sinnvollen Fehlermeldung quittieren und beim Aufruf mindestens eine kurze Hilfe ausgeben!