

2. Aufgabenblatt zum Softwarepraktikum “Muster- und Bildererkennung” 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, initialisiert und nach dem in der Vorbesprechung gegebenen Muster sortiert.

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.

Beachten Sie, daß die Satzzeichen ‘ ’ _ - , ; : ! ? . () [] \{ \} separat betrachtet werden müssen.

Aufgabe 2:

- a) Schreiben Sie eine lineare Suchroutine, die feste Wortfolgen im Text unter Verwendung des Suffix-Arrays auffindet.
- b) Verbessern Sie die in a) implementierte Suchfunktion durch das Ausnutzen der Sortierung des Suffix-Arrays.

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

Aufgabe 3:

Schreiben Sie eine weitere, verbesserte Suchfunktion, die die Gemeinsamkeiten benachbarter Suffix-Array-Einträge berücksichtigt. Vergleichen Sie die Laufzeiten der verschiedenen Suchverfahren auf den gegebenen Anfragen.

Die Testdateien und Anfragen finden Sie auf der Webseite des Praktikums (<http://www-i6.informatik.rwth-aachen.de/web/Teaching/LabCourses/SS08/Softwarepraktikum/>)

Abgabe:

Abnahmetermin: **Donnerstag, 29. Mai**

Abnahmeort: **CIP-Pool**

Bitte senden Sie den Quellcode der Lösungen **spätestens einen Tag vor Abnahme** per e-mail an die Betreuer **{plahl,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!