

Algorithmen und Datenstrukturen

Suche in Texten

Aufgabe 1 Levenshtein Distanz [2 Punkte]

Die Levenshtein Distanz gibt an, wie viele Editieroperationen zum Überführen des einen Strings in den andern notwendig sind. Bestimmen Sie die Levenshtein Distanz folgender String Paare.

AUSTAUSCH – AUFBAUSCH
 BARBAREN – BARBARA
 COCACOLA – COCAINA

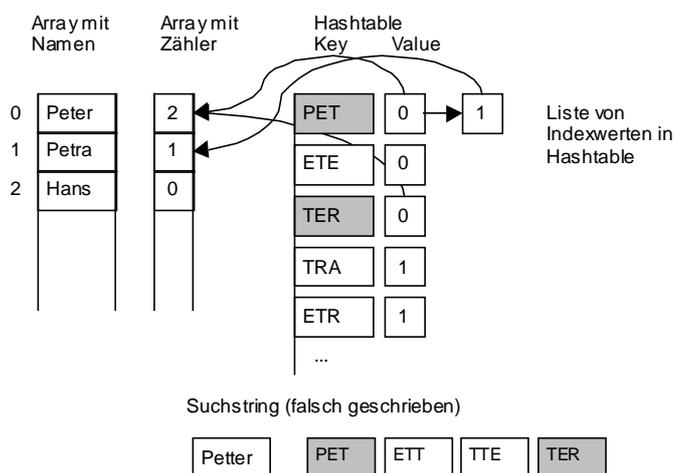
Aufgabe 2 Reguläre Ausdrücke [3 Punkte]

a) Lösen Sie die Aufgabe in den Vorlesungsunterlagen.

b) Definieren Sie Reguläre Ausdrücke für:
 Eine IP Adresse: 12.122.12.1 oder 198.168.1.1
 Eine E-Mail Adresse: hans.muster@zhaw.ch

Aufgabe 3 Fehlertolerante Suche [5 Punkte]

Es soll eine effiziente fehlertolerante Suche nach dem Namen (der Rangliste) implementiert werden. Für diesen Zweck wird die sogenannte Trigramm Methode angewandt. Die Idee dabei ist, dass der String in 3-Buchstaben Gruppen unterteilt wird, z.B. sind das bei „Peter“, die 3-er Gruppen „PET“, „ETE“, „TER“. Diese 3-er Gruppen werden für alle vorkommenden Namen gebildet und wie in folgender Abbildung gezeigt abgespeichert.



Die 3-er Gruppen werden in eine Hashtabelle als Schlüssel abgelegt; der dazugehörige Wert entspricht einer Liste von (Array-/Listen-)Indexwerten der entsprechenden Namen, z.B. PET kommt in Peter und Petra vor, demnach enthält die Liste 0 und 1 (Arrayindex von Peter und Petra).

Bei der Suche wird der Suchstring ebenfalls in 3-er Gruppen unterteilt. Diese Gruppen werden nun in der Hashtabelle gesucht und der Zähler des entsprechenden

Namens um eins inkrementiert. Das wird für alle 3-er Gruppen des Suchstrings gemacht. Derjenige Namen mit dem höchsten Zählwert am Schluss ist dann der gesuchte.

Aufgabe: Implementieren Sie einen **FuzzySearchServer** dem Sie eine Rangliste übergeben und einen Trigram Index erzeugen können.

Es soll dann nach einem Namen mit z.B. "Kip Dan" gesucht werden können und es sollte "Kiptum Daniel" gefunden werden.

Hinweise:

- Der Name wird zuerst noch "normalisiert", i.e. alles klein geschrieben.
- Map<String,List> zur Speicherung der Trigramme verwenden.
- Ergänzen Sie das Programmgerüst

Abgabe

Praktikum: ADS10.3

Filename: FuzzySearchServer.java