

Algorithmen und Datenstrukturen

Suche in Texten [10 Punkte]

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) Definieren Sie Reguläre Ausdrücke für eine IP Adresse, z.B: 12.122.12.1 oder 198.168.1.1.

b) Definieren Sie Reguläre Ausdrücke für eine E-Mail-Adresse, z.B.: hans.muster@zhaw.ch

Aufgabe 3 Suche nach Muster [5 Punkte]

Es soll auf Webseiten z.B. <https://tel.search.ch/?was=Meier> nach Telefonnummern gesucht werden. Überlegen Sie sich das Muster für (Schweizer) Telefonnummern und suchen Sie auf obiger Seite alle Nummern. Das Laden der Web Seite kann mit folgendem Code durchgeführt werden.

```
import java.net.*;
import java.io.*;

String inputLine;

URL oracle = new URL("http://www.oracle.com/");
BufferedReader in = new BufferedReader( new InputStreamReader(oracle.openStream()));
while ((inputLine = in.readLine()) != null) {
    System.out.println(inputLine);
}
in.close();
```

Hinweis:

https: Siehe Beispielcode HttpClient.java

Aufgabe 4 (optional)

Erweitern Sie Ihr Programm so, dass auch URLs erkannt werden und die Seiten dieser URL rekursiv gefolgt wird (bis zu einer vorbestimmten Tiefe). Achten Sie darauf, dass Sie das die Verlinkung auch Zyklen enthalten kann und es unterschiedliche Arten von Links gibt.