

Übung zur Vorlesung “Einführung in die Computerlinguistik und Sprachtechnologie”

Wintersemester 2018/2019, Prof. Dr. Udo Hahn, Luise Modersohn

Übungsblatt 5 vom 06.12.2018

Abgabe bis 12.12.2018, 23.59 Uhr; per Email (PDF-Format) an

luise.modersohn@uni-jena.de

Aufgabe 1 : Recherche (2 Punkte)

Warum wird ein Algorithmus wie der PorterStemmer in der Computerlinguistik eingesetzt? Warum wird nicht einfach eine Lemmatisierung durchgeführt?

Aufgabe 2 : Anwendung (4 Punkte)

a)

Schreibe die folgenden Beispiele als Folge von Vokalen V und Konsonanten K (wie in der Vorlesung gezeigt).

Beispiel: troubles = KVKVK.

1. Fuchsschwanzsäge
2. Schifffahrtsgesellschaft
3. Ballsaal
4. Angst

b)

Führen Sie für die unten genannten Wörter sowohl eine Lemmatisierung, als auch ein Stemming durch (siehe Vorlesung).

1. Krankheiten
 2. gesprochen
 3. verheiratet
 4. Nieren
-

Aufgabe 3 : Silbentrennung

5 Punkte

Definieren Sie eine Funktion, die geeignet ist eine vereinfachte Version der 1. Regel für Silbentrennung im Deutschen anzuwenden. Diese lautet (vereinfacht):

Es wird vor jedem Konsonanten getrennt, auf den ein Vokal folgt.

Betrachten Sie dabei die Liste `vokale` als gegeben, die alle Vokale (auch die Umlaute) des Deutschen als Strings enthält.

Ihre Funktion soll einen String entgegennehmen und alle Stellen finden, an denen laut dieser Regel getrennt werden könnte. Schreiben Sie dazu einfach an den passenden Stellen Trennstelle eintragen.

Anmerkungen:

- Sie können für diese Übung mit `String[n-1]` den n-ten Buchstaben eines Strings abfragen, also etwa für `"Mama" [0]` das "M".
- Zudem können Sie mit `length(String)` die Länge eines Strings abfragen, im obigen Beispiel also 4.
- Sie können mit `in` bzw. `not in` testen, ob etwas (nicht) in einer Liste ist. So würde etwa ein eingerückter Block unter `if "k" not in vokale` ausgeführt werden.
- Wörter können nicht vor dem ersten Buchstaben getrennt werden!