



Dr. Maximilian Hadersbeck
Leonie Weißweiler

Ludwig-Maximilians-Universität
Centrum für Informations- und Sprachverarbeitung

Abgabe: 12.1.2017

10. Übung zur Vorlesung Einführung in die Programmierung für Computerlinguisten

Bitte achten sie bei ihren Lösungen darauf, dass die Groß- und Kleinschreibung der Nutzereingaben keine Rolle spielen darf - die Eingabe "Spam and Eggs" soll also das gleiche Ergebnis liefern wie die Eingabe "spam and eggs".
Ausnahmen sind im Angabentext eindeutig gekennzeichnet.

Verwenden sie für die folgenden Aufgaben die Markdown Datei des Kapitels "Linux" des Skriptes, die sie hier herunterladen können: <http://www.cip.ifi.lmu.de/weissweiler/files/Blatt10/Linux.md>

Aufgabe 10-1

Suchen Sie alle Überschriften im Text. In Markdown-Dateien werden Überschriften mit # gekennzeichnet (#Große Überschrift, ##kleinere Überschrift etc.)

Aufgabe 10-2

Finden sie alle Wörter oder Passagen, die als fett markiert sind (**fett** oder `__fett__` steht für fett, aber *kursiv* oder `_kursiv_` für kursiv)

Aufgabe 10-3

Extrahieren Sie alle www-Links und speichern Sie sie in der Datei links.txt. WWW-Links sind in Markdown wie folgt gekennzeichnet: `<http://www.cis.uni-muenchen.de>`

Aufgabe 10-4

Geben Sie alle Zeilen aus, in denen Linux erwähnt wird.

Aufgabe 10-5

Suchen Sie im Text alle Wörter, in denen zwei Buchstaben doppelt hintereinander vorkommen, wie das Wort "vorkommen" zum Beispiel.

Aufgabe 10-6

Holen Sie sich mit wget den Wikipedia-Artikel über Gerhard Polt

Aufgabe 10-7

Extrahieren sie mit lynx dump den Text aus der html datei

Aufgabe 10-8

Erzeugen Sie eine Frequenzliste aller großgeschriebenen Wörter aus der Datei polt.txt, die länger als 5 Buchstaben sind und geben Sie die Anzahl der Keys aus.

Aufgabe 10-9

Schreiben Sie eine Funktion, die eine Zeile als Argument bekommt und die Wörter in umgekehrter Reihenfolge ausgibt, ohne slicing zu benutzen.

Aufgabe 10-10

Schreiben Sie eine Funktion, die eine Liste von Wörtern bekommt und jedes Wort zusammen mit der Position innerhalb der Wortliste ausdrückt. Ausgabe: Wort 1 = spam Wort 2 = and Wort 3 = eggs

Aufgabe 10-11

Lesen Sie sich den Wikipedia-Artikel zu dem Gedicht "Fünfter Sein" von Ernst Jandl durch. Im Gedicht ändert sich nur ein einziges Wort, was es möglich macht, es mit wenigen Schleifenanweisungen von einem Programm ausgeben zu lassen. Schreiben Sie das Programm.