



Dr. Maximilian Hadersbeck
Leonie Weißweiler

Ludwig-Maximilians-Universität
Centrum für Informations- und Sprachverarbeitung

Abgabe: 26.1.2017

12. 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.

Aufgabe 12-1

Schreiben Sie eine Funktion, die rekursiv die Fakultät einer übergebenen Zahl berechnet.

Aufgabe 12-2

Schreiben Sie eine Funktion, die für eine Zahl x rekursiv die Summe der Zahlen $= x + (x-1) + (x-2) + \dots + 1$ berechnet (z.B. $\text{summe}(4) = 4+3+2+1$).

Aufgabe 12-3

Schreiben Sie eine Funktion, die eine Zahl und eine Liste übergeben bekommt, rekursiv überprüft, ob die Zahl in der Liste vorhanden ist und das Ergebnis ausgibt.

Aufgabe 12-4

Schreiben Sie eine Funktion, die rekursiv das Produkt zweier Zahlen berechnet, indem sie sie auf die Addition zurückführt: $2*3 = 2 + 2 + 2$.

Aufgabe 12-5

Schreiben Sie eine Funktion, die eine Liste bekommt und jedes zweite Element auf der Konsole ausgibt.

Aufgabe 12-6

Schreiben Sie eine Funktion, die eine Liste bekommt, jede zweite Zahl daraus entfernt, die Reihenfolge der Zahlen umdreht und dann die Liste zurückgibt.

Aufgabe 12-7

Die Mailbox, voll mit SPAM - Mail, finden sie in der komprimierten Datei hier: <https://www.cip.ifi.lmu.de/~weissweiler/files/10MB_spammax-verkettet.txt.zip>

a) Kopieren Sie die Datei in Ihr Verzeichnis und entpacken Sie mit einem UNIX-Befehl die Datei in 10MB_spammax-verkettet.txt.

Schreiben Sie für die nächsten Aufgaben jeweils eine Funktion und dazu ein Hauptprogramm, das die Funktionen testet.

b) Lesen Sie im Hauptprogramm die gesamte Mailbox in einen String ein.

c) Schreiben Sie eine Funktion, der der String als Argument übergeben wird und die zählt, wie viele Emails in der Mailbox sind.

d) Schreiben Sie eine Funktion, der der String als Argument übergeben wird und die alle Absender("From") und den Betreff ("Subject:") der einzelnen Emails in einer Datei zeilenweise zusammen speichert.

e) Schreiben Sie eine Funktion, der der String übergeben wird und die alle Emailadressen der Mailbox extrahiert und sie mit ihrer Häufigkeit in einem dictionary einträgt und das dictionary zurück gibt.

f) Kopieren Sie die letzte Funktion und ändern Sie sie so ab, dass bei jeder Emailadresse der Username entfernt wird, also nur noch die Domain übrig ist und tragen Sie dann die Domains in das dictionary ein und das dictionary zurück gibt.

g) Schreiben Sie eine Funktion, der die Mailbox als Argument übergeben wird und die alle Wörter (ohne Sonder- oder Satzzeichen oder Ziffern) aus der Mailbox extrahiert und in der Datei "www.txt" speichert.

h) Schreiben Sie ein Programm, das die im vorherigen Programm erzeugte Datei www.txt liest und mit Hilfe eines dictionary eine Frequenzliste der Wörter erzeugt. Geben Sie die 30 häufigsten Wörter, die keine Stoppwörter sind, aus. (siehe Wikipedia: Stoppwörter nennt man im Information Retrieval Wörter, die bei einer Volltextindexierung nicht beachtet werden, da sie sehr häufig auftreten und gewöhnlich keine Relevanz für die Erfassung des Dokumentinhalts besitzen.) Eine Liste von Standard-Stoppwörtern für Deutsch und Englisch finden Sie auf <http://www.ranks.nl/stopwords>.