

# Informatik I: Einführung in die Programmierung

Prof. Dr. Bernhard Nebel  
Dr. Stefan Wöfl, Thorsten Engesser  
Wintersemester 2015/2016

Universität Freiburg  
Institut für Informatik

## Übungsblatt 11

**Abgabe: Freitag, 22. Januar 2016, 20:00 Uhr**

**WICHTIGE HINWEISE:** Zur Bearbeitung der Übungsaufgaben legen Sie bitte ein neues Unterverzeichnis `sheet11` im Wurzelverzeichnis Ihrer Arbeitskopie des SVN-Repositories an. Ihre Lösungen werden dann in Dateien in diesem Unterverzeichnis erwartet.

Beachten Sie bitte bei allen Aufgaben die *Hinweise zur Bearbeitung der Übungsaufgaben* unter der folgenden URL:

<http://gki.informatik.uni-freiburg.de/teaching/ws1516/info1/wiki/hinweise.html>

Bewertet wird bei allen Aufgaben die letzte Version, die zur Deadline des Übungsblattes auf dem SVN-Server eingereicht ist.

### Aufgabe 11.1 (Reguläre Ausdrücke; Punkte: 3+3; Datei: `regex.txt`)

Beschreiben Sie umgangssprachlich, welche Mengen von Strings von den folgenden regulären Ausdrücken spezifiziert werden. Geben Sie dann für jeden regulären Ausdruck jeweils zwei Strings an: die Anwendung des regulären Ausdrucks mittels `findall` soll dabei für den einen String eine leere Liste und für den anderen eine nicht-leere Liste zurückgeben.

- (i) `r'[^\\']+\\d[X-Z|?]`
- (ii) `r'(( [A-Za-z]+? ) (\\d) \\2-\\3)'`

### Aufgabe 11.2 (Reguläre Ausdrücke II; Punkte: 3+1; Datei: `month.py`)

- (a) Schreiben Sie eine Funktion `find_dates`, die aus einem Text alle Datumsangaben der Gestalt wie im Beispiel

`29. Februar 2016, 8:05 Uhr`

findet und als eine Liste von Dictionaries zurückgibt. Dabei soll das Leerzeichen vor der Monatsangabe optional sein. Jedes Dictionary habe dabei die Schlüssel `year`, `month`, `day`, `hour`, und `minute`, denen numerische Werte zugewiesen sind. Das Dictionary im Beispiel würde also wie folgt aussehen:

```
{'minute': 5, 'month': 2, 'hour': 8, 'year': 2016, 'day': 29}.
```

- (b) Schreiben Sie für die neu implementierte Funktion eine Test-Funktion mit 4 Assertions.

**Aufgabe 11.3** (Städte in Deutschland; Datei: `cities.py`; Punkte: 3+3+2)

In dieser Aufgabe geht es darum, aus den Wikipedia-Webseiten Informationen zu deutschen Städten zu extrahieren.

Die Liste der Städte in Deutschland finden Sie unter der folgenden URL:

[https://de.wikipedia.org/wiki/Liste\\_der\\_Städte\\_in\\_Deutschland](https://de.wikipedia.org/wiki/Liste_der_Städte_in_Deutschland).

Auf dieser Webseite finden Sie auch eine Auflistung der Kürzel der Bundesländer in Deutschland.

- (a) Implementieren Sie eine Funktion `cities(stateid)`, die zu einem eingegeben Länderkürzel ("`BW`", "`BY`", etc.) eine Liste der Städte in dem jeweiligen Bundesland zurückgibt. Jeder Eintrag der Liste soll hierbei den Namen der Stadt und ihre Wikipedia-URL enthalten.
- (b) Implementieren Sie eine Funktion `cityinfo(url)`, die aus der Wikipedia-Webseite `url` der Stadt aus der Infobox ("Basisdaten") die Anzahl der Einwohner der Stadt sowie die Partei des (Ober-) Bürgermeisters extrahiert (sofern vorhanden).
- (c) Schreiben Sie das Hauptprogramm, das ein Länderkürzel als Kommandozeilenargument übernimmt und auf der Standardausgabe alle Städte mit ihrer Einwohnerzahl und der Partei des Bürgermeisters auflistet. Am Schluss der Auflistung soll eine kurze Statistik ausgegeben werden, die pro Partei auflistet, wieviele Einwohner von einem Bürgermeister der Partei repräsentiert werden.

*Hinweis:* Benutzen Sie bei der Bearbeitung dieser Aufgabe die in der Vorlesung vorgestellten Techniken zum Scraping von Webseiten sowie reguläre Ausdrücke (keine zusätzlichen Python-Libraries).

Um die Wikipedia-Webseiten nicht zu sehr zu belasten, **beachten Sie unbedingt folgende Hinweise:**

- (a) Verwenden Sie zum Testen des Hauptprogramms zunächst Bundesländer mit wenigen Städten, also z.B. Bremen oder das Saarland.
- (b) Beachten Sie ferner, dass Sie nach jedem Aufruf `cityinfo`-Funktion mindestens 0.5s warten, bis die Funktion erneut aufgerufen wird: Importieren Sie dazu das Modul `time`; verwenden Sie dann den Befehl

```
time.sleep(0.5)
```

vor jedem Aufruf von `cityinfo`.

**Aufgabe 11.4** (Erfahrungen; Datei: `erfahrungen.txt`; Punkte: 2)

Legen Sie im Unterverzeichnis `sheet11` eine Textdatei `erfahrungen.txt` an. Notieren Sie in dieser Datei kurz Ihre Erfahrungen beim Bearbeiten der Übungsaufgaben (Probleme, benötigter Zeitaufwand nach Teilaufgabe, Bezug zur Vorlesung, Interessantes, etc.).