

Python – Regular Expressions

Albert Weichselbraun

Inhalt

- Wiederholung: String Funktionen
- Regular Expressions
 - Einfache Patterns
 - Meta-Zeichen
 - Substitutionen
 - Back-References

String Manipulationen

- `.strip()`, `.split()`
- `.find()`, `.replace()`
- `.upper()`, `.lower()`
- Liste aller unterstützten Methoden:
`dir(str)`
- Beispiel:

```
'Vienna is the capital of austria'.split()
['Vienna', 'is', 'the', 'capital', 'of',
'austria']
```

Regular Expressions in Python

- Vorgangsweise
 - Regular Expression Modul importieren
 - Den regulären Ausdruck compilieren
=> Regular Expression Objekt
 - Regular Expression Objekt verwenden
 - Suche/Matching
 - Substitutionen

Regular Expression Beispiel

```
# gibt alle zeilen, die den Begriff
# 'Hybridauto' enthalten aus.
import re,sys

re_hybridAuto=re.compile('hybridauto', re.I)

for line in sys.stdin:
    m = re_hybridAuto.search(line)
    if m:
        print line
```

Erweiterte Regular Expressions

- Optionale/Alternative Elemente

```
re_hybridA=re.compile('hybrid(-| )?auto',re.I)
```

- Metazeichen

\s ... Whitespace

\w ... Alphanumerische Zeichen

\d ... numerische Zeichen

{m,n} minimal 'm' und maximal 'n' Vorkommen

+ {1,x}

* {0,x}

? {0,1}

Gruppierung

- Extrahieren von Teilstrings aus dem Match-Objekt.
- Beispiel: Datum nach ISO 8601

```
import re, sys
re_date = re.compile('(\d{4})-(\d{2})-(\d{2})')

isoString = sys.argv[0]
m=re_date.search( isoString)
if m:
    print 'Das Jahr ist %d' % m.group(1)
else:
    print 'Das Datum ist nicht im ISO-Format'
```

Übungsaufgabe

- Schreiben Sie ein Programm, das eine Temperatur in Grad Celsius oder Fahrenheit entgegennimmt und die dann in die entsprechende andere Skala umrechnet.
- Bei der Eingabe muss der Benutzer die Temperatur inkl. der Einheit angeben (F bzw. Fahrenheit oder C bzw. Celsius). Wird zum Beispiel 37.7 C eingegeben, rechnet der Computer dies in Fahrenheit um.
 - $T[^\circ\text{C}] = 5/9 * (T[^\circ\text{F}] - 32)$
 - $T[^\circ\text{F}] = 9/5 * T[^\circ\text{C}] + 32$
- Zuerst soll nach der Eingabe eventueller Whitespaces am Beginn und am Ende abgeschnitten werden.

Übungsaufgabe

- Aus den folgenden Adressangaben sollen der Staatencode, die Postleitzahl und der Ortsname herausgefiltert werden. Für Wien soll zusätzlich der Bezirk extrahiert werden.
- Erlaubte Angaben sind:
 - A-1090 Wien
 - 9800 Spittal/Drau
 - A 9021 Klagenfurt
 - D - 80331 München

Substitutionen

- Ersetzt den gefundenen Text

```
reObject.sub('neuerSubString', string)
```

- Beispiel:

```
re_wind=re.compile('wind( |-)?energie', re.I)
```

```
for line in open('meinFile.text'):
```

```
    newText = re_wind.sub('solar-energie', line)
```

```
    print newText.strip()
```

Übungsaufgabe*

- **Marktstudie:** Erstellen Sie ein Programm, welches ermittelt, wie oft die folgenden Konzepte auf <http://www.heise.de> vorkommen:
('Linux', 'Windows', 'Internet Explorer', 'Mozilla Firefox')
- **Hinweis:** Formulieren Sie passende Regular Expressions, um auf leichte Modifikationen (zum Beispiel IE6 oder Mozilla zu erkennen).