

Cvičení 11

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu
studentů

Uložení seznamu
studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném
semestru

Počty studentů v
oborech

Počty studentů v
oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

IB111: cvičení 11

Zpracování textu, regulární výrazy

29. 11. 2013

Demo - soubory

Cvičení 11

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu
studentů

Uložení seznamu
studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném
semestru

Počty studentů v
oborech

Počty studentů v
oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

```
# otevreni souboru pro cteni
f = open("vstupni_soubor.txt", "r")
# nacteni celeho obsahu souboru do retezce
obsah = f.read()
# nacteni souboru souboru jako seznam radku
radky = f.readlines()
# otevreni souboru pro zapis
f = open("vystupni_soubor.txt", "w")
# zapis retezce do souboru
f.write(co)
# uzavreni souboru
f.close()
```

Demo - řetězce

Cvičení 11

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu studentů

Uložení seznamu studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném semestru

Počty studentů v oborech

Počty studentů v oborech seřazené

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

```
>>> radek = "jablko, hruska, pomeranc \n"
>>> radek = radek.rstrip()
#-> radek = 'jablko, hruska, pomeranc'
>>> radek.split(", ")
['jablko', 'hruska', 'pomeranc']
>>> radek.replace(", ", " -")
'jablko - hruska - pomeranc'
>>> radek.find("hruska")
8

>>> x = "\n"
>>> print x,

>>> x = r"\n"
>>> print x,

\n
```

Výpis seznamu studentů

Cvičení 11

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu studentů

Uložení seznamu studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném semestru

Počty studentů v oborech

Počty studentů v oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

Stáhněte si ze studijních materiálů zmutovaný seznam_studentu.csv.

Napište funkci, která vypíše seznam všech studentů v abecedním pořadí.

```
>>> studentiAbecedne()
Ambrozkova, Marcel
Ambrozkova, Pavla
Beranova, David
Butulova, Karel
Butulova, Martina
Butulova, Tomas
Chlubnova, Vladimir
Chudik, David
Cukr, Filip
...
```

Uložení seznamu studentů

Cvičení 11

Napište funkci, která vypíše uloží seznam všech studentů v abecedním pořadí do textového souboru.

```
>>> ulozStudentyAbecedne ("studenti_serazene.txt")
```

↓ *studenti_serazene.txt*

```
Ambrozkova, Marcel  
Ambrozkova, Pavla  
Beranova, David  
Butulova, Karel  
Butulova, Martina  
Butulova, Tomas  
Chlubnova, Vladimir  
Chudik, David  
...
```

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu studentů

Uložení seznamu studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném semestru

Počty studentů v oborech

Počty studentů v oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

Demo - regulární výrazy

Cvičení 11

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu
studentů

Uložení seznamu
studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném
semestru

Počty studentů v
oborech

Počty studentů v
oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel
Imitace textu

```
>>> import re
>>> m = re.search(r"(\w+) (\w+)", \
                  "Isaac Newton, fyzik")
>>> m.group(0)
"Isaac Newton"
>>> m.group(1)
"Isaac"
>>> m.group(2)
"Newton"
```

- libovolný znak `.`, opakování `*` `+` `?`
- výběr ze skupiny `[abc]`, `[a-z]`
- "cokoliv jiného než" `[^abc]`
- číslice `\d`, alfanumerické znaky `\w`, bílé znaky `\s`

Studenti v daném semestru

Cvičení 11

Napište funkci, která vypíše pouze ty studenty, kteří studují v zadaném semestru.

```
>>> studentiVSemestru(3)
Beranova, David
Ejem, Klara
Fojt, Roman
Hajkova, Tomas
Jurka, Klara
Kristanova, Barbora
Kucirek, Jakub
Lounek, Veronika
Malecek, Kamila
Nemcikova, Martin
Sarman, Lukas
...
```

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu studentů

Uložení seznamu studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném semestru

Počty studentů v oborech

Počty studentů v oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

Počty studentů v oborech

Cvičení 11

Napište funkci, která vypíše pouze počty studentů v jednotlivých oborech.

```
>>> pocetyStudentuVOborech ()
PSK 3
PTS 1
BIO 1
PDS 3
MI 2
INVS 8
SOC1 18
GRA 5
PSZD 4
UMI 13
PLIN 26
BcAP 26
```

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu studentů

Uložení seznamu studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném semestru

Počty studentů v oborech

Počty studentů v oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

Počty studentů v oborech seřazeně

Cvičení 11

Vylepšete předchozí funkci tak, aby vypisovala obory sestupně podle počtu studentů v daném oboru.

```
>>> pocytyStudentuVOborechSerazene ()  
PLIN 26  
BcAP 26  
SOCI 18  
UMI 13  
INVS 8  
GRA 5  
PSZD 4  
PSK 3  
PDS 3  
MI 2  
PTS 1  
BIO 1
```

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu
studentů

Uložení seznamu
studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném
semestru

Počty studentů v
oborech

**Počty studentů v
oborech seřazeně**

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

Demo - HTML

Cvičení 11

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu
studentů

Uložení seznamu
studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném
semestru

Počty studentů v
oborech

Počty studentů v
oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

```
import urllib2
URL_FI = "http://www.fi.muni.cz"
def stahniHtml(url):
    f = urllib2.urlopen(url)
    obsah = f.read()
    f.close()
    return obsah

def najdiNadpisy(url):
    text = stahniHtml(url)
    nadpisy = re.findall(r'<h(\d)>(.*?)</h\d>', \
                        text)

    print nadpisy

# [('1', 'Fakulta informatiky Masarykovy
# univerzity'), ('2', 'Studenti a zamestnanci'),
# ...
```

Odkazy

Cvičení 11

Vypište všechny odkazy na zadané stránce.

```
>>> odkazy("http://www.fi.muni.cz")
https://is.muni.cz/prihlaska/info.pl?op=n;utyp...
http://www.cerit.cz/cs/cerit-sp/
http://www.video.muni.cz/prednasky.cgi?fakulta...
http://is.muni.cz/do/1433/uredni_deska/
http://is.muni.cz/
http://is.muni.cz/obchod/fakulta/fi/
http://is.muni.cz/do/fi/studas/index.html
http://www.skm.muni.cz
http://library.muni.cz
http://inet.muni.cz
```

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu
studentů

Uložení seznamu
studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném
semestru

Počty studentů v
oborech

Počty studentů v
oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

Státy abecedně

Cvičení 11

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu studentů

Uložení seznamu studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném semestru

Počty studentů v oborech

Počty studentů v oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

Podívejte se tabulku států podle populace na adrese:

http://en.wikipedia.org/wiki/List_of_countries_by_population

Napište funkci, která vypíše všechny státy z této tabulky v abecedním pořadí.

```
>>> statyPodleAbecedy()
```

```
Afghanistan
```

```
Albania
```

```
Algeria
```

```
American Samoa
```

```
Andorra
```

```
Angola
```

```
Anguilla
```

```
Antigua and Barbuda
```

```
...
```

Počty obyvatel

Cvičení 11

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu
studentů

Uložení seznamu
studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném
semestru

Počty studentů v
oborech

Počty studentů v
oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

Napište funkci

`statyPodlePopulace(minPopulace, maxPopulace)`
která podle tabulky na adrese

`http://en.wikipedia.org/wiki/List_of_`
`countries_by_population`

nalezne a vypíše ty státy, které mají počet obyvatel
v intervalu [*minPopulace*, *maxPopulace*].

```
>>> statyPodlePopulace(2e8, 1e9)
317121000 United States
237641326 Indonesia
201032714 Brazil
```

Imitace textu

Cvičení 11

Soubory

Demo - soubory

Demo - řetězce

Výpis seznamu
studentů

Uložení seznamu
studentů do souboru

Regulární výrazy

Demo - regex

Studenti v daném
semestru

Počty studentů v
oborech

Počty studentů v
oborech seřazeně

Práce s HTML

Demo - HTML

Odkazy

Státy abecedně

Domácí úkol č.5

Počty obyvatel

Imitace textu

Napište funkci `imitaceTextu(soubor, pocetSlov)`, která provede analýzu četnosti slov a dvojic slov (tj. která slova následují po kterých slovech a jak často) a využije této analýzy k vygenerování náhodného textu o dané délce.

Tento text jednak vypíše na standardní výstup, jednak zapíše do souboru. Název výstupního souboru určete modifikací jména vstupního souboru, např.

krakatit.txt → *krakatit_imitace.txt*

```
>>> imitaceTextu("krakatit.txt", 30)
```

```
S hrůzou prsty. Potom vyslechl domovnici; zvěděl  
sice na tom, co budeš sebou výsměšná a rudé,  
jako vždy. A nám ztratil. Ovšem něco vyžvanil,  
uvědomoval si doktoři stopy jejího nitra
```