

Lekce 7

Práce se soubory

Práce se soubory. Zpracování textových souborů.

C2184 Úvod do programování v Pythonu
podzim 2016



C2184
Úvod do programování
v Pythonu

Soubory

Práce se soubory v
Pythonu

Práce s textovými
soubory

Příklady

Stanislav Geidl
Národní centrum pro výzkum biomolekul
Masarykova univerzita



- soubor je sada informací uložených pod konkrétním jménem na datovém médiu, např. soubor lecture07.pdf uložený na vašem pevném disku
- rozdělení souborů na
 - 1 **textové** soubory obsahují informace uložené čitelné textové podobě, kterou můžeme jednoduše uložit do typu string, příkladem může být formát molekul mol, mol2 nebo pdb, některé výstupy z programů, ...
 - 2 **binární** soubory obsahují informace uložené pro nás v nečitelné podobě sekvenci bytu, jedná se o zkompilevané programy, obrázky, videa, ...

Open, read, write, close

- otevření souboru `open(name, mode)`, módy: (r^1 , $r+^2$, w^3 , a^4 , t^5 , b^6)
- čtení dat v souboru `read`
- zápis dat do souboru `write`
- uzavření souboru `close`

```
filehandler = open("test.txt", "r")  
data = filehandler.read()  
filehandler.close()
```

¹read = pro čtení, pokud soubor neexistuje nastane výjimka `IOError`

²pro čtení a zápis, pokud soubor neexistuje nastane výjimka `IOError`

³write = pro zápis, stávající soubor bude přepsán, v opačném případě se soubor vytvoří

⁴append = pro zápis na konec souboru, pokud soubor neexistuje nastane výjimka `IOError`

⁵textový formát souboru, defaultní volba

⁶byte = binární formát souboru





- konstrukce `with` provede automatické uzavření souboru

```
with open("test.txt", "r") as filehandler:  
    data = filehandler.read()
```

Na co si dát pozor!

- kódování souboru (UTF-8, ISO-8895-2, Windows 1250 neboli CP1250)
 - funkce `unicode` a `encode`
- zakončení konce řádků (Windows `\r\n` vs. UNIX `\n`)
 - funkce `strip` a `rstrip`



- funkce `readlines` přečte celý soubor a řádky uloží do listu
- funkce `readline` přečte jeden řádek
- čtení celého souboru po řádcích pomocí cyklu `for`:

```
with open("test.txt", "r") as filehandler:  
    for line in filehandler:  
        print(line)
```

- tento způsob je pro zpracování rozsáhlých souborů nejlepší, do paměti nenhrájeme v jeden okamžik celý soubor, ale pracuje s ním po částech (řádcích)





- funkce `seek(position)` nastaví pozici (`position`) ukazatele v souboru
- funkce `tell()` nám řekne aktuální pozici v souboru



V souboru ciska.txt nahraďte středníky (;) za tabulátory (\t) a přidejte slopeček jež bude výsledkem:

sloupec1+sloupec2-sloupec3

Požadovaný výstup:

```
17 4.2 30 -8.8
21 4.0 32 -7.0
23 3.6 57 -30.4
20 3.8 48 -24.2
18 4.7 70 -47.3
```




```
numberfile2 = open("cisla2.txt", "w")
with open("cisla.txt") as numberfile:
    for line in numberfile:
        x1, x2, x3 = line.strip().split(";")
        x4 = float(x1) + float(x2) - float(x3)
        line = line.rstrip().replace(";", "\t") + "\t"
        numberfile2.write(line + str(x4) + "\n")
numberfile2.close()
```

Příklad I - řešení (obecněji)

```
numberfile2 = open("cisla2.txt", "w")
with open("cisla.txt") as numberfile:
    for line in numberfile:
        line = line.strip()
        numbers = line.split(';')
        """ numbers = [float(x) for x in numbers] """
        for i in range(len(numbers)):
            numbers[i] = float(numbers[i])
        numbers.append(numbers[0]+numbers[1]-numbers[2])
        """ numbers = [str(x) for x in numbers] """
        for i in range(len(numbers)):
            numbers[i] = str(numbers[i])
        numberfile2.write("\t".join(numbers)+"\n")
numberfile2.close()
```





Vytvořte souboru `cisla3.txt`, který bude na bude navíc obsahovat součty sloupečků.

Požadovaný výstup:

17; 4.2; 30

21; 4.0; 32

23; 3.6; 57

20; 3.8; 48

18; 4.7; 70

99.0; 20.3; 237.0

```
numberfile3 = open("cisla3.txt", "w")
sums = [0,0,0]
with open("cisla.txt") as numberfile:
    for line in numberfile:
        line = line.strip()
        numbers = line.split(';')
        for i in range(3):
            sums[i] += float(numbers[i])
        numberfile3.write(line+'\n')
numberfile3.write('{};{};{}\n'.format(
    sums[0], sums[1], sums[2]))
numberfile3.close()
```

