

Lekce 6

Práce se soubory

Práce se soubory. Zpracování binárních a textových souborů.

C2184 Úvod do programování v Pythonu
podzim 2015



C2184
Úvod do programování
v Pythonu

Soubory

Práce se soubory v
Pythonu

Práce s textovými
soubory

Práce s binárními
soubory

Cvičení

Mgr. 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 , t^4 , b^5)
- 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

⁴textový formát souboru, defaultní volba

⁵byte = 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 (řádciích)



C2184
Úvod do programování
v Pythonu

- funkce seek nastaví pozici ukazatele v souboru
- funkce tell nám řekne aktuální pozici v souboru

Soubory

Práce se soubory v
Pythonu

Práce s textovými
soubory

Práce s binárními
soubory

Cvičení



C2184
Úvod do programování
v Pythonu

Cvičení 1: Napište vlastní grep pro hledání klíčového slova
v souboru.

Soubory

Práce se soubory v
Pythonu

Práce s textovými
soubory

Práce s binárními
soubory

Cvičení



C2184
Úvod do programování
v Pythonu

Cvičení 2: Napište skript pro čtení molekuly ve formátu mol.
Jedná se o tzv. formát s fixní šířkou (Fixed Width Text File)
podobně jako formát PDB.

[Soubory](#)[Práce se soubory v
Pythonu](#)[Práce s textovými
soubory](#)[Práce s binárními
soubory](#)[Cvičení](#)



C2184
Úvod do programování
v Pythonu

Cvičení 3: Napište skript pro čtení souboru ve formátu CSV.

Soubory

Práce se soubory v
Pythonu

Práce s textovými
soubory

Práce s binárními
soubory

Cvičení



C2184
Úvod do programování
v Pythonu

Cvičení 4: Napište skript pro extrakci atomových nábojů z výstupu Gaussianu.

Soubory

Práce se soubory v Pythonu

Práce s textovými soubory

Práce s binárními soubory

Cvičení