

Práca so súbormi

Poznámky k téme

Všetko, čo program uloží do premenných, zmizne, keď program skončí. Ak chceme údaje zachovať – výsledky, nastavenia, záznamy – alebo spracovať veľké množstvo údajov pripravených vopred, použijeme **súbory**. V tejto téme sa naučíš čítať a zapisovať **textové súbory**: otvoriť súbor, prečítať ho riadok po riadku, rozobrať riadky na údaje a výsledky zapísať do iného súboru.

Čo sa naučíš

- otvoriť súbor v správnom režime ("r" , "w" , "a") a s kódovaním UTF-8,
- čítať celý súbor aj po riadkoch a odstrániť znak konca riadku,
- zapisovať do súboru a pridávať na jeho koniec,
- používať blok `with`, ktorý súbor spoľahlivo zatvorí,
- spracovať súbor s údajmi oddelenými bodkočiarkou.

1 Textový súbor

Textový súbor je postupnosť riadkov textu. Každý riadok (možno okrem posledného) končí neviditeľným znakom **konca riadku** `\n`. Program hľadá súbor podľa mena v **pracovnom priečinku** – zvyčajne v tom istom priečinku, v ktorom je uložený program. Súbor môže byť aj inde, potom treba uviesť cestu ("data/body.txt").

V ukážkach budeme používať tento súbor:

Súbor `teploty.txt`:

```
12.5
14.0
9.5
17.0
```

2 Otvorenie súboru

Súbor otvoríme funkciou `open`. Dostane meno súboru, **režim** a **kódovanie**:

```
f = open("teploty.txt", "r", encoding="utf-8")
```

Režim	Význam	Ak súbor neexistuje	Ak súbor existuje
"r"	čítanie (predvolený)	chyba <code>FileNotFoundError</code>	číta sa od začiatku
"w"	zápis	vytvorí sa nový	pôvodný obsah sa zmaže
"a"	pridávanie na koniec	vytvorí sa nový	píše sa za pôvodný obsah

Pozor: Uvádzať vždy `encoding="utf-8"`. Bez neho použije Python kódovanie operačného systému a na Windows sa potom slovenské znaky (č, ě, ľ...) môžu načítať alebo zapísať pokazené.

Ak súbor na čítanie neexistuje (preklep v mene, iný priečinok), program skončí chybou:

```
f = open("tepoty.txt", "r", encoding="utf-8")
```

Výstup (program skončí chybou):

```
Traceback (most recent call last):
  File "program.py", line 1, in <module>
    f = open("tepoty.txt", "r", encoding="utf-8")
    ~~~~~^
FileNotFoundError: [Errno 2] No such file or directory: 'tepoty.txt'
```

Ako takúto chybu zachytiť, sa dozvieš v téme Výnimky.

3 Blok with

Otvorený súbor treba po práci **zatvoriť** – až vtedy sa zapisované údaje naozaj uložia na disk a súbor sa uvoľní pre iné programy. Najbezpečnejšie je použiť blok `with`: súbor sa zatvorí automaticky na konci bloku, aj keby v ňom nastala chyba.

```
with open("teploty.txt", "r", encoding="utf-8") as f:
    obsah = f.read()
print(obsah)
```

Výstup:

```
12.5
14.0
9.5
17.0
```

`as f` pomenuje otvorený súbor – cez `f` s ním v bloku pracujeme. Za blokom (bez odsadenia) je súbor už zatvorený. Starší spôsob bez `with` vyzerá takto: `f = open(...)`, práca so súborom a nakoniec `f.close()` – zatvorenie sa však ľahko zabudne.

4 Čítanie súboru

4.1 Celý súbor naraz: read()

`f.read()` vráti celý obsah ako jeden reťazec (aj so znakmi `\n`). Hodí sa pre menšie súbory, napríklad keď chceme spočítať znaky alebo slová.

4.2 Riadok po riadku: for

Najčastejšie súbor čítame cyklom `for` – v každom opakovaní dostaneme jeden riadok:

```
with open("teploty.txt", "r", encoding="utf-8") as f:
    for riadok in f:
        print(repr(riadok))
```

Výstup:

```
'12.5\n'
'14.0\n'
'9.5\n'
'17.0\n'
```

Funkcia `repr` ukázala to, čo je inak neviditeľné: **každý riadok končí znakom `\n`**. Preto pred ďalším spracovaním riadok takmer vždy „očistíme“ metódou `strip()`, ktorá odstráni koniec riadku a medzery na okrajoch:

```
sucet = 0
pocet = 0
with open("teploty.txt", "r", encoding="utf-8") as f:
    for riadok in f:
        sucet += float(riadok.strip())
        pocet += 1
print(f"Priemerná teplota: {sucet / pocet:.2f} °C")
```

Výstup:

```
Priemerná teplota: 13.25 °C
```

(`float` a `int` si s koncom riadku poradia aj sami, ale pri porovnávaní textu – `riadok == "Eva"` – by bez `strip()` podmienka nikdy neplatila.)

4.3 Všetky riadky do zoznamu: `readlines()` a `read().splitlines()`

```
with open("teploty.txt", "r", encoding="utf-8") as f:
    riadky = f.read().splitlines()
print(riadky)
print("Počet riadkov:", len(riadky))
print("Najvyššia:", max(float(r) for r in riadky))
```

Výstup:

```
['12.5', '14.0', '9.5', '17.0']
Počet riadkov: 4
Najvyššia: 17.0
```

`f.readlines()` vráti zoznam riadkov **aj so** znakmi `\n`; `f.read().splitlines()` ich vráti bez nich. Jeden riadok prečíta `f.readline()`.

Pozor: Súbor sa číta od miesta, kde sa čítanie naposledy skončilo. Po prečítaní celého súboru by druhé `f.read()` vrátilo prázdny reťazec. Ak potrebuješ údaje viackrát, ulož si ich do zoznamu.

5 Zápis do súboru

Na zápis otvoríme súbor v režime "w" a použijeme metódu `write`. Tá na rozdiel od `print` **nepridáva koniec riadku** – musíme ho napísať sami (`"\n"`). Do `write` patrí iba text, čísla treba previesť (`str` , f-reťazec).

```
with open("nasobky.txt", "w", encoding="utf-8") as f:
    for i in range(1, 6):
        f.write(f"{i} x 7 = {i * 7}\n")

with open("nasobky.txt", "r", encoding="utf-8") as f:
    print(f.read())
```

Výstup:

```
1 x 7 = 7
2 x 7 = 14
3 x 7 = 21
4 x 7 = 28
5 x 7 = 35
```

Do súboru sa dá zapisovať aj príkazom `print` s parametrom `file` – potom sa koniec riadku pridá sám:

```
with open("pozdrav.txt", "w", encoding="utf-8") as f:
    print("Ahoj,", "svet!", file=f)
    print("Druhý riadok", file=f)

with open("pozdrav.txt", encoding="utf-8") as f:
    print(f.read())
```

Výstup:

```
Ahoj, svet!
Druhý riadok
```

(Režim "r" je predvolený, preto ho pri čítaní môžeme vynechať.)

5.1 Pridávanie na koniec: režim "a"

Režim "w" pôvodný obsah **zmaže**. Ak chceme pridávať – napríklad do denníka alebo záznamu výsledkov –, použijeme "a" (*append*):

```
with open("dennik.txt", "w", encoding="utf-8") as f:
    f.write("pondelok: 12 °C\n")
with open("dennik.txt", "a", encoding="utf-8") as f:
    f.write("utorok: 15 °C\n")
with open("dennik.txt", "a", encoding="utf-8") as f:
    f.write("streda: 9 °C\n")

with open("dennik.txt", encoding="utf-8") as f:
    print(f.read())
```

Výstup:

```
pondelok: 12 °C
utorok: 15 °C
streda: 9 °C
```

6 Spracovanie riadkov s údajmi

Údaje v súboroch často bývajú v **stĺpcoch** oddelených znakom – bodkočiarkou, čiarkou alebo tabulátorom (formát CSV). Každý riadok rozdelíme metódou `split` a jednotlivé časti prevedieme na správny typ.

Súbor `vysledky.txt`:

```
meno;trieda;body
Ema;3.A;47
Ján;3.B;38
Sofia;3.A;50

Tomáš;3.B;44
```

Súbor má **hlavičku** (prvý riadok s názvami stĺpcov) a jeden **prázdny riadok** – oboje treba pri spracovaní preskočiť:

```
vysledky = []
with open("vysledky.txt", encoding="utf-8") as f:
    f.readline()          # preskočíme hlavičku
    for riadok in f:
        riadok = riadok.strip()
        if riadok == "":
            continue      # prázdny riadok
        meno, trieda, body = riadok.split(";")
        vysledky.append((meno, trieda, int(body)))

for meno, trieda, body in vysledky:
    print(f"{meno:<6} {trieda} {body:>3}")
priemer = sum(b for _, _, b in vysledky) / len(vysledky)
print(f"Priemer: {priemer:.2f}")
```

Výstup:

```
Ema    3.A  47
Ján    3.B  38
Sofia  3.A  50
Tomáš  3.B  44
Priemer: 44.75
```

Postup spracovania riadku si zapamätaj: **strip** → (**preskoč prázdny**) → **split** → **prevod na čísla**.

7 Riešené úlohy

7.1 Výsledky podľa tried do nového súboru

Zadanie: Zo súboru `vysledky.txt` vypočítaj priemer bodov každej triedy a výsledok zapíš do súboru `triedy.txt` v tvare `3.A: 48.5`.

Riešenie: Body sčítavame v slovníku *trieda* → *zoznam bodov* (téma Slovníky). Potom priemery zapíšeme do nového súboru.

```
body_tried = {}
with open("vysledky.txt", encoding="utf-8") as f:
    f.readline()
    for riadok in f:
        riadok = riadok.strip()
        if not riadok:
            continue
        meno, trieda, body = riadok.split(";")
        body_tried.setdefault(trieda, []).append(int(body))

with open("triedy.txt", "w", encoding="utf-8") as f:
    for trieda in sorted(body_tried):
        b = body_tried[trieda]
        f.write(f"{trieda}: {sum(b) / len(b):.1f}\n")

with open("triedy.txt", encoding="utf-8") as f:
    print(f.read())
```

Výstup:

```
3.A: 48.5
3.B: 41.0
```

`setdefault(trieda, [])` vráti zoznam pre danú triedu, a ak ešte neexistuje, najprv ho vytvorí prázdny – praktická skratka.

7.2 Štatistika textového súboru

Zadanie: Spočítaj, koľko riadkov, slov a znakov má súbor.

Súbor `basen.txt`:

```
Prší, prší, len sa leje,
kde ste, vtáčky, kde ste?
```

```
with open("basen.txt", encoding="utf-8") as f:
    text = f.read()
print("Riadkov:", len(text.splitlines()))
print("Slov:", len(text.split()))
print("Znakov (bez koncov riadkov):", len(text.replace("\n", "")))
```

Výstup:

```
Riadkov: 2
Slov: 10
Znakov (bez koncov riadkov): 49
```

7.3 Očíslovanie riadkov

Zadanie: Skopíruj súbor `basen.txt` do súboru `cislovana.txt` a pred každý riadok pridaj jeho číslo.

```
with open("basen.txt", encoding="utf-8") as vstup, \
    open("cislovana.txt", "w", encoding="utf-8") as vystup:
    for cislo, riadok in enumerate(vstup, start=1):
        vystup.write(f"{cislo}: {riadok}")

with open("cislovana.txt", encoding="utf-8") as f:
    print(f.read())
```

Výstup:

```
1: Prší, prší, len sa leje,
2: kde ste, vtáčky, kde ste?
```

Riadok už obsahuje `\n`, preto ho za riadok nepridávame. V jednom `with` môžeme otvoriť aj dva súbory naraz.

8 Časté chyby

- **Zabudnutý `strip()`**: riadok obsahuje `\n`, takže `riadok == "Eva"` neplatí ani pri riadku s menom Eva.
- **Režim `"w"` namiesto `"a"`**: zápis zmaže všetko, čo v súbore bolo.
- **Chýbajúci `\n` pri `write`**: všetko sa zapíše do jedného dlhého riadku.
- **Zápis čísla**: `f.write(42)` je chyba – `write` prijíma iba text (`f.write(str(42))`).
- **Súbor sa nenašiel**: skontroluj meno, príponu (`.txt`) a priečinok, z ktorého sa program spúšťa.
- **Chýbajúce `encoding="utf-8"`**: pokazená diakritika.
- **Opakované čítanie**: po `read()` je súbor „dočítaný“; ďalšie čítanie vráti prázdny text.

Zhrnutie

- `with open(meno, režim, encoding="utf-8") as f:` – súbor sa na konci bloku sám zatvorí.
- Režimy: `"r"` čítanie, `"w"` zápis (zmaže obsah), `"a"` pridávanie na koniec.
- Čítanie: `f.read()` (celý text), `for riadok in f` (po riadkoch), `f.read().splitlines()` (zoznam riadkov bez `\n`), `f.readline()` (jeden riadok).
- Každý riadok končí `\n` – odstráň ho `strip()`.
- Zápis: `f.write(text)` (bez automatického `\n`) alebo `print(..., file=f)`.
- Spracovanie údajov: `strip` → preskoč prázdne → `split(";")` → prevod typov.

Úlohy na precvičenie

1. Zapiš do súboru čísla 1 až 100, každé na samostatný riadok, a potom zo súboru vypočítaj ich súčet.
2. Zo súboru s menami (jedno na riadok) vypíš mená zoradené abecedne a zapiš ich do nového súboru.
3. Zo súboru `vysledky.txt` vypíš meno s najvyšším počtom bodov.
4. Napíš program, ktorý pri každom spustení pridá do súboru `log.txt` riadok s textom zadaným z klávesnice.
5. Spočítaj, koľkokrát sa v textovom súbore vyskytuje každé slovo, a 5 najčastejších zapiš do súboru.

Úlohy k tejto téme nájdeš na learnwithmente.com v časti Informatika → Python.