

# Parallel File Analyzer

**Autor:** Jakub Davídek

**Škola:** SPŠE Ječná

**Datum:** 24. 11. 2025

**Kontakt:** neuvedeno

## 1. Úvod

Tento dokument popisuje první verzi projektu *Parallel File Analyzer*.

Jde o jednoduchý nástroj pro paralelní analýzu souborového systému, který využívá více vláken (threading) a architekturu producer–consumer.

Projekt umožňuje rychlou analýzu adresáře, získání informací o velikostech souborů a tvorbu souhrnných statistik.

## 2. Požadavky uživatele

- možnost analyzovat libovolnou složku,
- rychlá analýza pomocí více vláken,
- přehledné statistiky: největší soubory, nejmenší soubory, přehled přípon,
- snadná konfigurace v samostatném konfiguračním souboru.

### 3. Architektura aplikace

Aplikace je rozdělena do několika modulů:

- **producer.py** – vyhledává soubory a vkládá je do fronty,
- **worker.py** – zpracovává soubory ve více vláknech,
- **analyzer.py** – vyhodnocuje výsledky a generuje statistiky,
- **config.py** – obsahuje volby konfigurace,
- **main.py** – hlavní spouštěcí logika programu.

### 4. Běh aplikace

Po spuštění aplikace se vytvoří pracovní vlákna, která odebírají soubory z fronty a zpracovávají je.

Po dokončení analýzy se vypíše výsledky.

### 5. Konfigurace

Konfigurace se provádí v souboru **conf/config.py**.

Lze upravit:

- cílovou složku pro analýzu,
- počet worker vláken.

### 6. Instalace a spuštění

Aplikace nevyžaduje žádné externí knihovny.

Spouští se příkazem:

```
python -m src.main
```

## 7. Chybové stavy

- neexistující složka – program vypíše chybu a ukončí se,
- nedostupná oprávnění – některé soubory mohou být přeskočeny.

## 8. Testování

Projekt obsahuje jednoduché unit testy ověřující funkčnost základních částí:

- producer – správné vkládání cest do fronty,
- worker – správné zpracování souboru a uložení výsledků,
- analyzer – výpočet statistik.

Testování:

```
python -m unittest discover tests
```

## 9. Známé problémy a budoucí rozšíření

### Známé problémy

- pomalé čtení na síťových discích,
- ignorování některých systémových souborů.

### Možnosti rozšíření

- detekce duplicitních souborů (hashování),
- export do HTML nebo CSV,
- progress bar během analýzy.