

# [PB071] Cvičný zápočtový príklad

## Data mining

V tejto úlohe si vyskúšate vo veľmi zjednodušenej podobe prácu s dátami – **data mining**. Vašou hlavnou úlohou bude dáta spracovať a následne vypísať požadovanú informáciu z daných dát, v tomto prípade konkrétne zo zjednodušenej tabuľky cien akcií firiem patriacich do prestížnej skupiny **S&P 500**<sup>1</sup>.

### Zadanie:

Napíšte program, ktorý načíta dáta zo súboru vo formáte **CSV**<sup>2</sup> (Comma-separated values) do vhodnej štruktúry, vylúči všetky firmy, ktorých **MarketCap** je menší ako požadovaný, tieto dáta následne zoradí a vypíše požadovanú informáciu z týchto spracovaných dát podľa prepínača, s ktorým bude tento program spustený.

### Volanie programu:

Program bude volaný nasledovne:

```
./mining [OPTION] [OPTION_PARAM] [MARKET_CAP] [FILE]
```

pričom **OPTION** je jeden prepínač, **OPTION\_PARAM** je nepovinná súčasť a obsahuje prípadný parameter viažuci sa k **OPTION**, **MARKET\_CAP** je **číslo**, ktoré udáva filtračnú hranicu a **FILE** je cesta k súboru, ktorý sa má spracovať. Prepínače môžu byť **3 rôzne**:

1. **-e [PARAM]** -> vypíše najvyššiu **Price** (cena jednej akcie) zo zadaného sektoru. Túto informáciu zmysluplne poskytne užívateľovi, prípadne ho informuje, že v danom sektore sa nenachádza žiadna firma.
2. **-c [PARAM]** -> vypíše počet firiem zastúpených v danom sektore.
3. **-s [PARAM]** -> vypíše súčet hodnoty všetkých firiem v danom sektore podľa MarketCap.

### Doplňujúce informácie

- O dĺžke jednotlivých riadkov ani o počte riadkov v danom súbore nemáte dopredu žiadne informácie.
- Dáta v súbore budú v tvare:  
**<Symbol>,<Name>,<Sector>,<Price>,<MarketCap>**
- Dáta budú zoradené podľa **Sector**, do ktorého patria. Ak budú dve firmy spadať do rovnakého sektoru, radte ďalej podľa **Price**.
- Môžete očakávať, že **Sector** bude mať maximálne dĺžku 32 bytov. Zároveň sa môžete spoliehať, že v **Price** a **MarketCap** budú validné číselné hodnoty. Zamyslite sa, či je potrebné ukladať aj ostatné stĺpce tabuľky.

<sup>1</sup> <https://datahub.io/core/s-and-p-500-companies-financials>

<sup>2</sup> <https://cs.wikipedia.org/wiki/CSV>

## Požiadavky na implementáciu:

- Dáta musíte z jedného riadku súboru ukladať do **Vami definovanej užívateľskej štruktúry**.
- Minimálne na jednom mieste musíte použiť **dynamickú alokáciu**.
- Na zoradenie dát môžete použiť knižničnú funkciu **qsort()**.
- Nezapomnite po sebe uvoľňovať všetky zdroje, ktoré získate.
- Pokiaľ nastane **nečakaná situácia** (súbor sa nedá otvoriť, zlyhala dynamická alokácia, zlyhanie iných štandardných funkcií, atď.) zariadte sa spôsobom známym z domácich úloh.
- Počet argumentov príkazovej riadky musí byť presne **4**, pričom MARKET\_CAP musí byť validne číslo. V inom prípade program ukončíte a vypíšete na vhodnú chybovú hlášku na **stderr**.
- Ukážka hodnôt v CSV súbore:  
**ADBE,Adobe Systems Inc,Information Technology,185.16,94550214268**
- Na uloženie hodnoty **MarketCap** využite typ **správny typ**. Ukážka hodnôt:  
138721000000; 181386000000; 52518668144; ... (viac v priloženom CSV súbore).

Veľa šťastia do práce :-)