

Projektová dokumentace

Varianta 3: COVID-19

Ukládání a příprava dat

Šimon Galba, Bc. (xgalba03)
Gladiš Damián, Bc. (xgladi00)
Jeřábek František, Bc. (xjerab25)

16. prosince 2021

Obsah

1	Úvod	1
2	Spustenie	1
3	Využitie knižnice	1
4	Využitie dáta	1
5	Známe nedostatky	2

1 Úvod

Cieľom projektu je odpovedať na príslušné otázky ohľadom pandemickej situácie Covid-19 v Českej Republike efektívne a rýchlo pomocou dát uložených v NON-SQL databáze MongoDB v prvej časti projektu.

2 Spustenie

Keďže sa jedná o programovací jazyk Python, jednotlivé časti projektu sa bez prekladu spustia príkazom

- **database_extraction.py** pre vytvorenie súboru typu csv z databáze
- **data_processing.py** pre odpovedanie na dotazy pomocou dát z csv súborou z predošlého kroku
- **data_manipulation.py** pre vytvorenie a následnú úpravu csv súboru slúžiaceho ako vstupné dáta dolo-
vacieho algoritmu

za predpokladu naplnenej databázy MongoDB. Pre stiahnutie potrebných dát a ich uloženie do databáze je potrebné zavolať príkaz **python3 download.py**.

3 Využité knižnice

- matplotlib - vytvorenie grafov ako grafické odpovede na otázky
- numpy - spracúvanie dát požitých v grafoch
- pymongo.collection - prístup do databázy MongoDB
- MongoClient - prístup do databázy MongoDB
- Pandas - spracovávanie dát v súbore typu csv

4 Využité dáta

Pre účely vytvorenia prehľadov a grafov v druhej časti tohto projektu sú stiahnuté a spracované nasledujúce datasey (kliknutie na kategóriu odkazuje na daný súbor), všetky dáta sú čerpané zo zdroja **mzcr**:

- Celkový (kumulativní) počet osob s prokázanou nákazou dle krajských hygienických stanic včetně labo-
ratoří, počet vyléčených, počet úmrtí a provedených testů
- Přehled osob s prokázanou nákazou dle hlášení krajských hygienických stanic
- Přehled vykázaných očkování podle krajů ČR
- Přehled vykázaných očkování podle profesí (očkovací místo, bydliště očkovaného)
- Přehľad celkovej populácie
- Přehľad hospitalizovaných osôb s covidom-19
- Číselník intervalu vekového rozptylu
- Rozloženie veku obyvateľstva ČR

5 Známe nedostatky

Dáta o očkovaní jednotlivých osôb sú z dôvodu zrušenia online súboru poskytujúceho tieto dáta vložené do programu ručne pomocou priloženého csv súboru. Táto úprava je nevyhnutná, keďže pri návrhu projektu sme na tieto dáta spoliehali a sú nevyhnutnou súčasťou jeho riešenia.