

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta Informačních Technologií

COVID-19

UPA projekt - 1. část

1. listopadu 2021

Vojtěch Jahoda (xjahod06)
Hana Křížová (xkrizo03)
Aleš Kašpárek (xkaspa48)

Analýza dílčí datové sady

Pro účely druhé části projektu, která se zabývá zpracováním dotazů nad daty, jsme ze stránky <http://onemocneni-aktualne.mzcr.cz/api/v2/covid-19> vybrali následující datové sady:

- **Základní přehled**

Základní přehled obsahuje stručný náhled na základní epidemiologická data o pandemii COVID-19 v ČR.

- **Přehled osob s prokázanou nákazou dle hlášení krajských hygienických stanic (v2)**

- **Přehled vyléčených dle hlášení krajských hygienických stanic**

- **Přehled úmrtí dle hlášení krajských hygienických stanic**

Datová sada obsahuje záznamy o úmrtí osob, které byly pozitivně testovány (metodou PCR) bez ohledu na to, jaké byly příčiny jejich úmrtí.

- **Přehled hospitalizací**

Přehled hospitalizací zaznamenává aktuální a celkový počet hospitalizovaných pacientů, spolu s průběh nemoci a příznaky.

- **Celkový (kumulativní) počet osob s prokázanou nákazou dle krajských hygienických stanic včetně laboratoří, počet vyléčených, počet úmrtí a provedených testů (v2)**

- **Přehled epidemiologické situace dle hlášení krajských hygienických stanic podle okresu**

- **Přehled epidemiologické situace dle hlášení krajských hygienických stanic podle ORP**

Přehled zahrnuje základní epidemiologické parametry na úrovni obcí s rozšířenou působností.

- **Epidemiologická charakteristika obcí**

Epidemiologická charakteristika obcí obsahuje základní epidemiologické parametry na úrovni obcí v ČR.

- **Epidemiologická charakteristika městských částí hlavního města Prahy**

- **Přehled osob s prokázanou nákazou dle krajských hygienických stanic včetně laboratoří za 7 a 14 dní za ČR**

- **Přehled osob s prokázanou nákazou dle krajských hygienických stanic včetně laboratoří za 7 a 14 dní podle krajů**

- **Přehled osob s prokázanou nákazou dle krajských hygienických stanic včetně laboratoří za 7 a 14 dní podle okresů**

- **Přehled provedených testů podle typu a indikace**

Přehled provedených testů zahrnuje denní počty provedených testů (PCR a antigenní testy) na onemocnění COVID-19 dle hlášení laboratoří.

- **Celkový (kumulativní) počet provedených testů podle krajů a okresů ČR**

Datová sada se zaměřuje na provedené PCR testy s korekcí na opakovaně pozitivní (kontrolní) testy.

- **Přehled vykázaných očkování podle krajů ČR**

Přehled zahrnuje základní data o vykázaných očkováních na úrovni krajů ČR.

- **Demografický přehled vykázaných očkování v čase**

Demografický přehled zaznamenává očkování dle pohlaví a věkové skupiny.

Způsob získání a uložení datových sad

Získání datových sad je zpracováno v jazyce Python. Vytvoříme HTTP dotaz na webovou stránku `http://onemocneni-aktualne.mzcr.cz/api/v2/covid-19`, kterou následně procházíme a hledáme všechny relevantní soubory.

Pro ukládání dat jsme zvolili NoSQL databázi MongoDB, kde jsou data ukládána ve formátu JSON. Vytvořili jsme vlastní databázi s názvem `upa`. Při procházení výše uvedené webové stránky pro každý nalezený soubor vytvoříme kolekci, kam vyexportujeme data ze souboru. MongoDB jsme si zvolili z důvodu širokého využití a snadného zacházení s daty.

Požadavky na spuštění

- Docker 3.7
- VirtualBox

Spuštění programu

Celý projekt je kontejnerizován za pomoci nástroje Docker. Jak databáze MongoDB, tak i samotné skripty plnící databázi běží ve svém samostatném kontejnu. Celý projekt je spuštěn přes příkazovou řádku v kořeni projektu příkazem `sudo docker-compose up --build`.