

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)

Úvod

Jako téma projektu do předmětu UPA jsem si vybral s ohledem na aktuální dění COVID-19. Všechna prezentovaná data budeme v rámci zpracování tohoto projektu čerpat z webových stránek <http://onemocneni-aktualne.mzcr.cz/api/v2/covid-19>. Zdroj těchto dat pochází od Národního zdravotnického informačního systému, krajských hygienických stanic a Ministerstva zdravotnictví ČR.

Analýza dílčí datové sady

Pro účely druhé části projektu, která se zabývá zpracováním dotazů nad daty, jsme z výše uvedené webové stránky vybrali následující datové sady:

- **Základní přehled**

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

Položky datové sady jsou: *datum, provedene_testy_celkem, potvrzene_pripady_celkem, aktivni_pripady, vyleceni, umrti, aktualne_hospitalizovani, provedene_testy_vcerejsi_den, potvrzene_pripady_vcerejsi_den, potvrzene_pripady_dnesni_den, provedene_testy_vcerejsi_den_datum, potvrzene_pripady_vcerejsi_den_datum, potvrzene_pripady_dnesni_den_datum, provedene_antigenni_testy_celkem, provedene_antigenni_testy_vcerejsi_den, provedene_antigenni_testy_vcerejsi_den_datum, vykazana_ockovani_celkem, vykazana_ockovani_vcerejsi_den, vykazana_ockovani_vcerejsi_den_datum, potvrzene_pripady_65_celkem, potvrzene_pripady_65_vcerejsi_den, potvrzene_pripady_65_vcerejsi_den_datum, ockovane_osoby_celkem, ockovane_osoby_vcerejsi_den, ockovane_osoby_vcerejsi_den_datum*

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

Datová sada zahrnující denní incidenční přehled osob s prokázanou nákazou COVID-19 dle hlášení krajských hygienických stanic.

Položky datové sady jsou: *datum, vek, pohlavi, kraj_nuts_kod, okres_lau_kod, nakaza_v_zahranici, nakaza_zeme_csu_kod*

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

Datová sada zahrnující záznamy o vyléčených po onemocnění COVID-19 dle hlášení krajských hygienických stanic.

Položky datové sady jsou: *datum, vek, pohlavi, kraj_nuts_kod, okres_lau_kod*

- **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.

Položky datové sady jsou: *datum, vek, pohlavi, kraj_nuts_kod, okres_lau_kod*

- **Přehled hospitalizací**

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

Položky datové sady jsou: *datum, pacient_prvni_zaznam, kum_pacient_prvni_zaznam, pocet_hosp, stav_bez_priznaku, stav_lehky, stav_stredni, stav_tezky, jip, kyslik, hfno, upv, ecmo, tezky_upv_ecmo, umrti, kum_umrti*

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

Datová sada obsahuje kumulativní denní počty nakažených, vyléčených, úmrtí a počet testů dle krajských hygienických stanic.

Položky datové sady jsou: *datum, kumulativni_pocet_nakazenych, kumulativni_pocet_vylecenych, kumulativni_pocet_umrti, kumulativni_pocet_testu, kumulativni_pocet_ag_testu, prirustkovy_pocet_nakazenych, prirustkovy_pocet_vylecenych, prirustkovy_pocet_umrti, prirustkovy_pocet_provedenych_testu, prirustkovy_pocet_provedenych_ag_testu*

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

Přehled zahrnuje základní epidemiologické parametry na úrovni okresů.

Položky datové sady jsou: *datum, kraj_nuts_kod, okres_lau_kod, kumulativni_pocet_nakazenych, kumulativni_pocet_vylecenych, kumulativni_pocet_umrti*

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

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

Položky datové sady jsou: *den, datum, orp_kod, orp_nazev, incidence_7, incidence_65_7, incidence_75_7, prevalence, prevalence_65, prevalence_75, aktualni_pocet_hospitalizovanych_osob, nove_hosp_7, testy_7*

- **Epidemiologická charakteristika obcí**

Datová sada obsahuje základní epidemiologické parametry na úrovni obcí v ČR.

Položky datové sady jsou: *den, datum, kraj_nuts_kod, kraj_nazev, okres_lau_kod, okres_nazev, orp_kod, orp_nazev, obec_kod, obec_nazev, nove_pripady, aktivni_pripady, nove_pripady_65, nove_pripady_7_dni, nove_pripady_14_dni*

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

Přehled zahrnuje základní epidemiologické parametry pro Prahu.

Položky datové sady jsou: *den, datum, okres_nuts_kod, orp_kod, orp_nazev, mc_kod, nove_pripady, aktivni_pripady, nove_pripady_65, nove_pripady_7_dni, nove_pripady_14_dni, zemreli, vyleceni*

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

Datová sada obsahující počty potvrzených případů za posledních 7 a 14 dní za celou ČR.

Položky datové sady jsou: *datum, incidence_7, incidence_14, incidence_7_100000, incidence_14_100000*

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

Datová sada obsahující počty potvrzených případů za posledních 7 a 14 dní podle krajů.

Položky datové sady jsou: *datum, kraj_nuts_kod, kraj_nazev, incidence_7, incidence_14, incidence_7_100000, incidence_14_100000*

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

Datová sada obsahující počty potvrzených případů za posledních 7 a 14 dní podle okresů.

Položky datové sady jsou: *datum, okres_lau_kod, okres_nazev, incidence_7, incidence_14, incidence_7_100000, incidence_14_100000*

- **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.

Položky datové sady jsou: *datum, pocet_PCR_testy, pocet_AG_testy, typologie_test_indik_diagnosticka, typologie_test_indik_epidemiologicka, typologie_test_indik_preventivni, typologie_test_indik_ostatni, incidence_pozitivni, pozit_typologie_test_indik_diagnosticka, pozit_typologie_test_indik_epidemiologicka, pozit_typologie_test_indik_preventivni, pozit_typologie_test_indik_ostatni, PCR_pozit_symp, PCR_pozit_asymp, AG_pozit_symp, AG_pozit_asymp_PCR_conf*

- **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.

Položky datové sady jsou: *datum, kraj_nuts_kod, okres_lau_kod, prirustkovy_pocet_testu_okres, kumulativni_pocet_testu_okres, prirustkovy_pocet_testu_kraj, kumulativni_pocet_testu_kraj, prirustkovy_pocet_prvnich_testu_okres, kumulativni_pocet_prvnich_testu_okres, prirustkovy_pocet_prvnich_testu_kraj, kumulativni_pocet_prvnich_testu_kraj*

- **Přehled vykázaných očkování podle krajů ČR**

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

Položky datové sady jsou: *datum, vakcina, kraj_nuts_kod, kraj_nazev, vekova_skupina, prv-nich_davek, druhych_davek, celkem_davek*

- **Demografický přehled vykázaných očkování v čase**

Demografický přehled zaznamenává očkování dle pohlaví a věkové skupiny.

Položky datové sady jsou: *datum, vakcina, vakcina_kod, poradi_davky, vekova_skupina, po-hlavi, pocet_davek*

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
- 64-bit OS
- Mongo

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`.

Mongo databáze běží v kontejneru na portu 10022 a lze se na ni připojit pomocí příkazu `mongo mongodb://user:passwd@localhost:10022/upa`.