# Dokumentace k projektu UPA

## Zvolené téma

COVID-19 (varianta 04) – dr. Rychlý

## Řešitelé

* Filip Jeřábek (xjerab24)
* Daniel Konečný (xkonec75)
* Tomáš Ryšavý (xrysav27)

## Zvolené dotazy a formulace vlastního dotazu

### Dotaz varianty A

V grafech zobrazte tempo změny počtů aktuálně nemocných (absolutní i procentuální přírůstek pozitivních případů a klouzavý průměr různých délek v různých časech).

### Dotaz varianty B

Určete vliv počtu nemocných a jeho změny v čase na sousední okresy (aneb zjistěte jak se šíří nákaza přes hranice okresů).

### Vlastní dotaz

Lokalizujte ohniska přírůstku nakažených nemocí COVID-19. Dotaz u každého okresu porovná jeho denní přírůstek s denními přírůstky všech sousedních okresů. Pokud je jeho přírůstek ze všech nejvyšší, je daný okres označený jako ohnisko.

## Stručná charakteristika zvolené datové sady

Zdrojem dat jsou otevřené datové sady publikované Ministerstvem zdravotnictví České republiky týkající se onemocnění COVID-19 dostupné na adrese: <https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19>. Používaná data jsou z datové sady s názvem COVID-19: Přehled epidemiologické situace dle hlášení krajských hygienických stanic podle okresu a zpracovaná jsou ve formátu JSON, který je dostupný na této adrese: <https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19/kraj-okres-nakazeni-vyleceni-umrti.json>. Datová sada obsahuje denní informace o přírůstku nakažených, vyléčených a zemřelých osob v jednotlivých krajích a okresech. Je aktualizovaná s týdenním zpožděním z důvodu validace informací krajskými hygienickými stanicemi.

Data ve formátu JSON jsou stažena z internetové adresy uvedené výše. Kořenový JSON objekt obsahuje informaci o čase vytvoření této datové sady, dále zdroj, ze kterého data pocházejí, a poté pole samotných dat o přírůstcích. Jedna položka tohoto pole odpovídá informacím z jednoho dne v jednom okrese a navíc je uveden i kraj. Dále obsahuje kumulativní počet nakažených, vyléčených a zemřelých.

{

"modified": "2020-10-21T01:01:32+02:00",

"source": "https:\/\/onemocneni-aktualne.mzcr.cz\/",

"data": [

{

"datum": "2020-03-01",

"kraj\_nuts\_kod": "CZ010",

"okres\_lau\_kod": "CZ0100",

"kumulativni\_pocet\_nakazenych": 2,

"kumulativni\_pocet\_vylecenych": 0,

"kumulativni\_pocet\_umrti": 0

},

{

"datum": "2020-03-01",

"kraj\_nuts\_kod": "CZ020",

"okres\_lau\_kod": "CZ020A",

"kumulativni\_pocet\_nakazenych": 0,

"kumulativni\_pocet\_vylecenych": 0,

"kumulativni\_pocet\_umrti": 0

},

...

]

}

Dále se využívá námi vytvořená datová sada, která obsahuje informace o tom, který okres sousedí se kterým. A datová sada přiřazující názvu okresu LAU kód, která pochází z webu Ministerstva zemědělství České republiky. Datová sada se nachází zde: <https://eagri.cz/ssl/nosso-app/DataKeStazeni/Okresy?page=1&sortBy=NAZEV~NAZEV&pageSize=1000&collapsed=False>

## Zvolený způsob uložení surových dat

Data budou uložena v NoSQL databázi Neo4j. Každý denní záznam v jednom okrese reprezentuje jeden uzel. Sousedící okresy jsou propojeny patřičnou vazbou. Dále jsou postupně propojeny vazbou jednotlivé denní záznamy náležející jednomu okresu. V každém uzlu je uveden přírůstek nakažených, vyléčených a zemřelých jako vlastnost tohoto uzlu.