

Potřebujouroušku.cz report

Tento projekt vznikl s cílem maximálně zjednodušit nařízení vydaná kvůli zamezení šíření COVID-19 v dobách, kdy žádný obdobný informační zdroj, kde by bylo všechno, nebyl. Vývoj započal v srpnu 2020, kdy byly tehdy "enormní přírůstky" nových nakažených, třeba 500 lidí za den. První verze webu byla spuštěna 1.9.2020.

Jaký problém/úkol práce řeší, proč bylo potrebujourousku.cz třeba?

Protože více jak půl roku od začátku pandemie COVID-19 nebyl v České republice k dispozici kvalitní oficiální zdroj informací. Nařízení se měnila třeba i několikrát denně a udržet přehled o tom, co se zrovna teď může a co ne, byl nadlidský úkol (i na existujících webech se informace objevovaly se značným zpožděním). Komunikace ze strany ministerstva zdravotnictví byla žalostná. Problém nebyl až tak úplně v tom, že by informace nebyly, ale spíš v tom, že byly pohřbené v surovém (pro ne-právnicka ne příliš dobře čitelném) PDF nařízení někde na webu.

Existovala stránka [onemocnění aktuálně](#), kde se časem začaly objevovat i opatření (ne příliš detailně popsaná), ale jen v rámci regionů. Celostátní opatření byla vyvěšována v rámci již neexistující podstránky na webu [mzcr](#), s dosti výrazným zpožděním. Najít tam něco byl pomalu nadlidský úkol, navigace webu s neustále měnila.

Co člověk nepochytil z rádia, nebo z televize, či médií, prostě nevěděl. Tehdy to ale bylo hlavně o tom, kde že se má nosit rouška, či jiná ochrana dýchacích cest a kde ne. Česká republika se nacházela na počátku druhé vlny, byť se tomu tak ještě nesmělo říkat.

Bylo tedy třeba nějakého místa, kde by byly k dispozici důležité informace týkající se nás všech, jednoduše a přehledně podané. Tak vzniklo potrebujourousku.cz.

Během listopadu 2020 situace hodně změnila a vznikl portál covid.gov.cz, který vznikl za pomoci dobrovolníků z [Česko.digital](#) pod správou [nakit.cz](#).

Co vlastně je potrebujourousku.cz?

Webová stránka s informacemi o COVID nařízeních, které se zobrazují v závislosti na zvolené poloze (uživatel tedy nemusí sledovat na jednom místě celostátní opatření a na druhém místě lokální opatření a nějak si to skládat dohromady). Může si zobrazit opatření, které platí v místě, kde se nachází a vůbec jej nemusí zajímat, jaká opatření platí na druhém konci republiky.

Uživateli se zobrazí opatření platná na zvoleném místě (např. Ostrava), tedy jak všechna opatření platná na úrovni města, okresu, kraje i ČR. Alternativně lze zobrazit pouze celostátně platná opatření.


Lokální opatření

Hlavní město Praha

[Nařízení](#) [Doporučení](#) [Informace](#) [Nadcházející](#) [Nařízení za nouzového stavu](#)


Pro zobrazení detailu, odkazu, platnosti, či výjimek klikněte/tapněte na název opatření.
Zobrazují i nařízení, které nabydou platnosti během následujících 7 dní.

Lockdown




Omezení kontaktů a pobyту na veřejnosti

Celoplošné




Zákaz vycházení v noci

Celoplošné



Zákaz zbytečného pohybu přes den


Celoplošné




Datové schánky budou bezplatné po dobu nouzového stavu

Celoplošné


Hromadné akce / Omezení



Omezení demonstrací



Zákaz pití alkoholu na



Roušky všude na

Lokální opatření (Hlavní město Praha)

Opatření jsou zobrazena v barevných kartičkách (určuje význam) s možností náhledů, seřazena podle jednotlivých kategorií pro lepší přehlednost.


Testování a karanténa



Seznam zemí s mírným rizikem nákazy

od 9. listopadu 2020 0:00

Celoplošné



Testování při příjezdu ze zahraničí

od 9. listopadu 2020 0:00

Celoplošné



Algoritmus pro izolaci a karanténu

Celoplošné



Seznam zemí s nízkým rizikem nákazy

končí 9. listopadu 2020 0:00


Celoplošné



Testování při příjezdu ze zahraničí

končí 9. listopadu 2020 0:00

Celoplošné

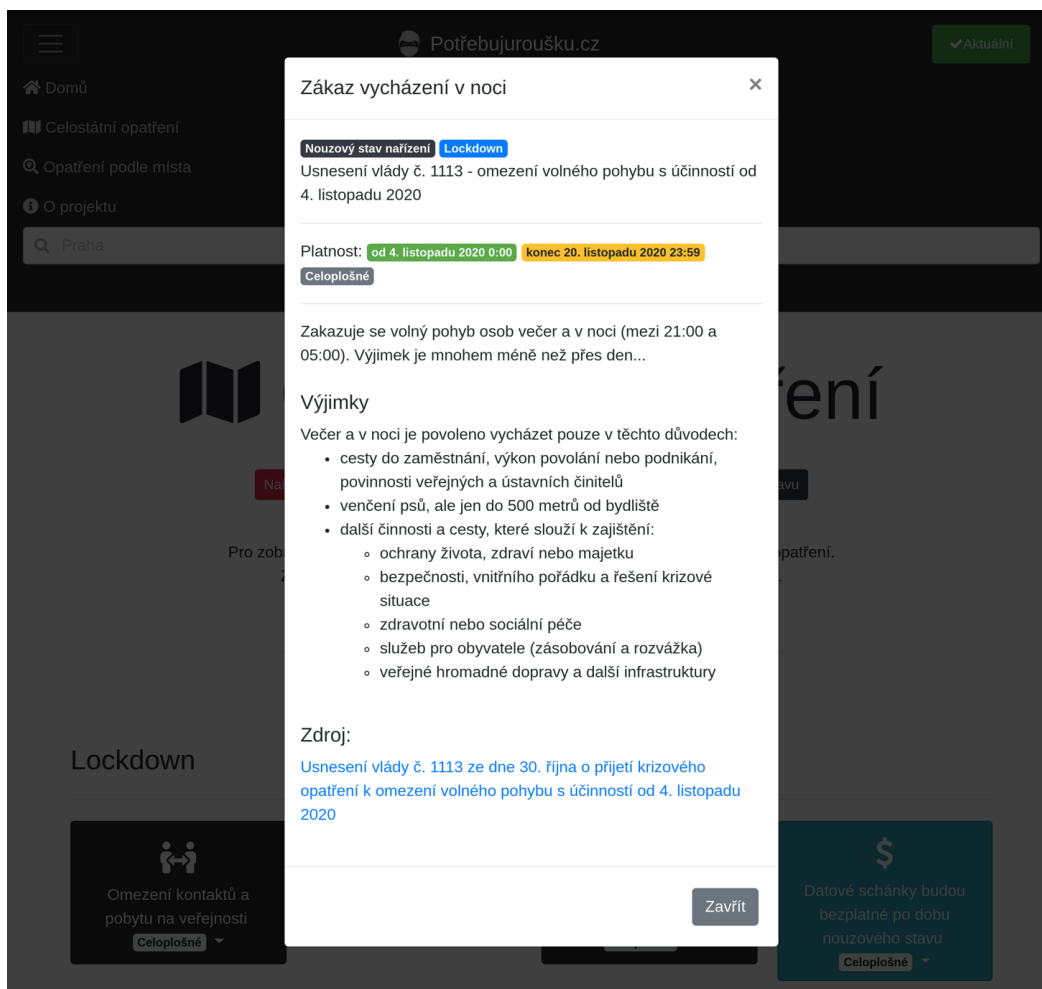


Výsledky testů na COVID zasílány elektronicky

Celoplošné

Opatření z kategorie "Testování a karanténa"

Po kliknutí/tapnutí na kartičku se uživateli zobrazí detail opatření, obsahující dodatečné informace, upřesnění, výjimky, externí odkazy, zdroje a další.



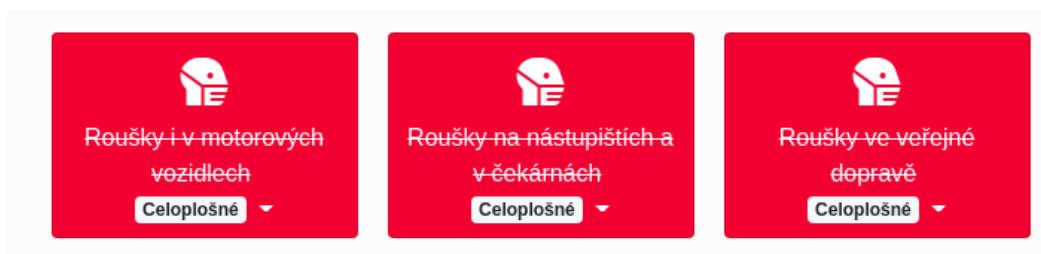
Sunrise by the mountains

Použité metody/postupy/algoritmy

Projekt se zabývá v podstatě zobrazením dat z databáze, není pod tím nic moc hlubšího, tudíž asi není moc o čem psát. Donedávna vše řešilo pár (relativně komplikovaných) SQL dotazů a proběhlo pár snah o migraci do Django.Models.

Součástí je kontrola aktualizací, kde je použita knihovna [BeautifulSoup](#), pomocí které dochází ke kontrole Wordpressového webu [mzcr](#), kde byla nařízení/opatření publikována zpravidla jako příspěvek s jedním, nebo více odkazy na PDF soubor.

Scrapper tedy v podstatě projde všechny příspěvky z kategorie [Mimořádná opatření](#) a sesbírá odkazy na PDF nařízení. Odkazy pak kontroluje vůči databázi a zobrazí report. Takováto kontrola probíhala automaticky každých 5 minut a v případě že bylo něco odhaleno (například odebrání nařízení), uživatel byl obeznámen, že informace nemusí být aktuální. Všechny obsah pak kontrolovali správci. Uživatel si mohl dokonce zobrazit výstup této kontroly na potrebujuroušku.cz/aktualnost, kde bylo sepsáno co vše chybí, či přebývá.



Takto se projevilo odebrání opatření z webu MZCR

Výsledky a co by mohlo být lepší

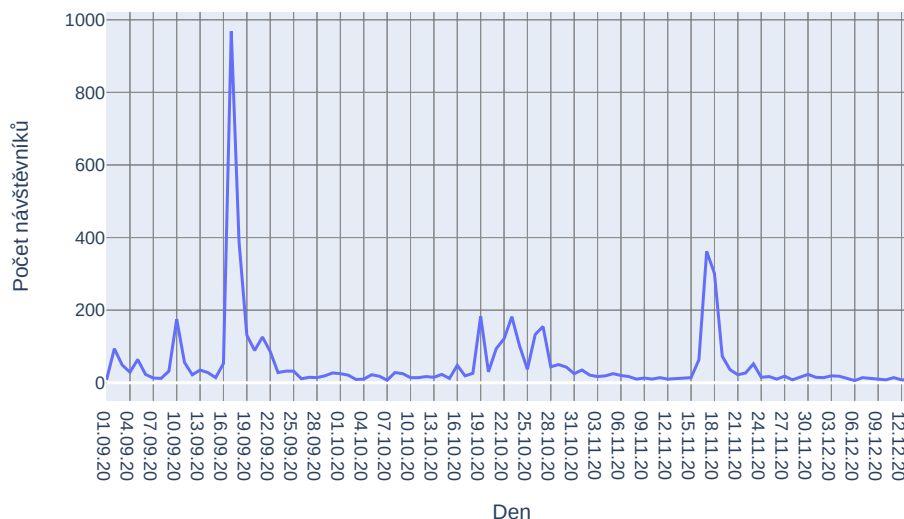
Návštěvnost

Během čtyř měsíců od spuštění web dosáhl cca 15K návštěv (podle Google Analytics) a navštívilo jej necelých 4.5K návštěvníků, což nejsou zanedbatelná čísla, ale mohlo by to být mnohem lepší.

O projektu také vyšlo pár článků, například

- [respekt](#)
- [Deník N](#) (paywall)

Návštěvnost 1.9.2020 - 14.12.2020



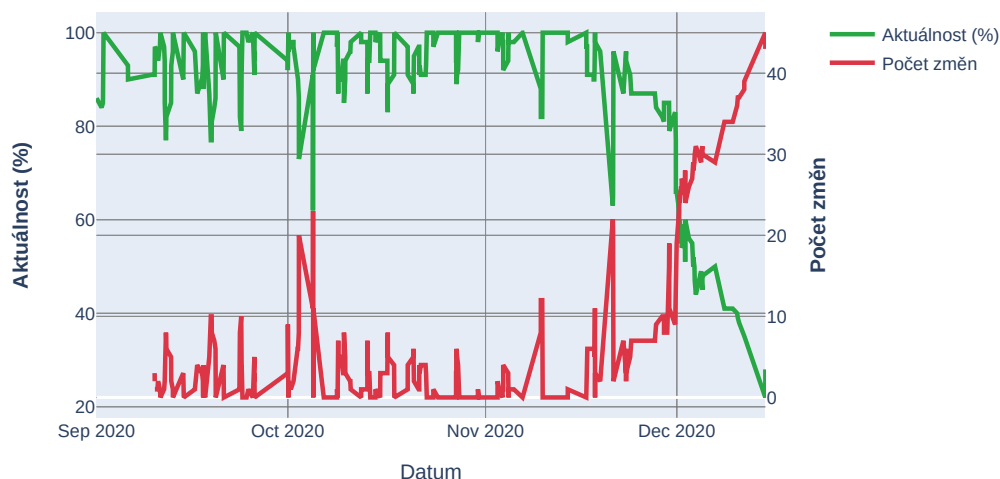
Zdroj informací a zpracování

Na začátku se zdálo jako nejrozumější brát data z [Ministerstva Zdrvotnictví](#). Nejvíce, co šlo "vytěžit" byly PDF s nařízeními, které se tam objevovaly do několika hodin (místy až dní) po tiskových konferencích. Vesměs se tam ale informace náhodně objevovaly a mizely. Struktura stránek se za dobu chodu třikrát výrazně změnila.

I tak se ale díky týmu, který kolem [potrebujurousku.cz](#) časem vznikl, dařilo zpracovat většinu nařízení do 24 hodin od zveřejnění nařízení na webu [mzcr](#) tyto PDF zpracovat do lidsky čitelného textu, rozdělit je do logických celků a vložit do databáze.

Jistě by pomohlo mít lidí více a mít nějaký spolehlivější informační zdroj.

Aktuálnost a počet detekovaných změn



Technické řešení

Projekt využívá frameworku Django, je zde použit Bootstrap. Vyhledávání na webu je řešené pomocí javascriptu (zde je určitý prostor pro zlepšení). Na webu jsou jak statické, tak dynamicky generované prvky.

Při návrhu potřebujouroušku nebyl až tak kladen důraz na výkon, jako spíše na minimalizaci nároků na čas správců. Při větší zátěži pak docházelo k odezvě. Toto by se dalo vykompenzovat pokročilejším cachováním obsahu - naprosto dostačující je vygenerovat statické stránky např. jednou za 30 minut, nebo více.

Srovnání s covid.gov

Jak již bylo řečeno, během listopadu 2020 začal vznikat za výrazné podpory skupiny Česko.Digital portál covid.gov.

Covid.gov je staticky generovaná webová stránka, kde jsou informace nejenom o protiepidemiologických opatřeních, ale také hlavně o životních situacích, na které klade mnohem větší důraz. Nabízí i vyhledávání a obsah je z části přeložen do anličtiny. Od začátku byl zde kladen důraz na výkon a zvládne tedy obsloužit mnohem více uživatelů. Možná díky vyšší návštěvnosti, každopádně na [PageSpeed Insights](#) vychází potřebujouroušku o něco lépe.



MOBILNÍ ZAŘÍZENÍ



POČÍTAČ



<https://potrebujurousku.cz/>

0-49 50-89 90-100 ⓘ



<https://potrebujurousku.cz/>

0-49 50-89 90-100 ⓘ



<https://covid.gov.cz/>

0-49 50-89 90-100 ⓘ



<https://covid.gov.cz/>

0-49 50-89 90-100 ⓘ

Srovnání PageSpeed Insights z 1.1.2020

Dle mého subjektivního pohledu má covid.gov

Výhody:

- Rychlé načítání (souvisí s technickým řešením), ustojí mnohem větší zátěž
- Podporované řadou organizací, "oficiální řešení"
- Rozsáhlý tým lidí, kteří mají informace přímo u zdroje
- Financované
- Možnost vyhledávání

Nevýhody:

- Obsah tvoří hlavně text, žádné obrázky, videa (třeba na dovysvětlení proč, mapa "zelených" zemí atd.)
- Řeší pouze celoplošná opatření (což může být časem komplikace)

Možné vylepšení do budoucna

Závěrem bych se chtěl změřit na "nice to have" možné pokračování do budoucna. Nedostal jsem se k nim, protože se zdá, že nyní (01/2021) již není potřeba tento projekt nadále vyvíjet. Portál [covid.gov](https://covid.gov.cz/) již umí většinu z toho, co umělo [potrebujurousku](https://potrebujurousku.cz/) a ve většině ohledů je to lepší řešení. Také nemá příliš smysl pokračovat v přidávání nových funkcí, když není aktuální obsah. Přinejmenším to ale byla cenná zkušenost a posunulo mne to o dost dál.

1) Zobrazení platných nařízení podle zadané polohy

Většinu této funkcionality má na starost poměrně rozsáhlý SQL dotaz (nyní přepsaný do ORM). Ten podle ID obce/města, ID obce s rozšířenou působností, ID okresu, ID kraje, nebo podle celé ČR zobrazí nařízení vztahující se na ono konkrétní zvolené místo (například obec Holčovice) a zároveň

opatření vztahující se na vyšší územní celky (tedy např. na okres, kraj, stát). Pro html výstup je zde důležitá hromada podmínek a cyklů psaných v built-in tags Django, která generují výslednou stránku podle toho, co přijde z databáze.

Možnost vylepšení:

- Možnost filtrování (podle kategorie)
- Možnost specifikace časového okna, pro které se položky opatření zobrazují (nyní se pevně zobrazují na 7 dní dopředu, pokud je teda stihneme do databáze zadat. Poslední dobou se opatření mění v jednotkách dnů, takže pevně stanovených 7 dní je možná až příliš)
- Možnost něčeho jako diff - tedy "co se změnilo od poslední návštěvy"
- Možnost notifikací

2) Automatická kontrola aktualizací

Součástí stránek je co cca 5 minut spouštěná instance třídy `kontrola`, která scrappuje obsah wordpressového webu ministerstva zdravotnictví (konkrétně příspěvky vytvořené v [kat. mimořádné opatření](#)) pomocí knihovny BeautifulSoup4 a kontroluje je oproti databázi, se kterou na projektu pracuji. Za dobu, co je web na světě, jsem ji předělával asi třikrát a to proto, že byl zdrojový web pozměněn. y

Možnosti vylepšení:

- Paralelizace při scrappování
- vytěžování dat přímo z PDF nařízení
- Oddělení kontroly aktuálnosti a veškerých částí, které nutně musí pracovat s DML operacemi od zbytku řešení (ten by pracoval jen s readonly přístupem)
- Propojení s existujícím covid.gov.cz a brát informace odsud. Tady jsou totiž vcelku dobře zpracované a úplně bez práce.

3) Vyhledávání polohy

Kvůli rozsáhlému množství míst, které máme obsažené v databzi (14 krajů, 70+ okresů, 200+ správních celků, 10000+ měst a obcí) a na základě kterých pak zobrazujeme lokálně platná nařízení, není vhodné dělat select list o několika desítek tisíc položkách. Používáme tedy řešení podobné, jako mají třeba České dráhy, nebo IDOS. Na frontendu se javascript "ptá" na nějaké rozumné množství položek, které začínají, či obsahují nějakou část vstupu uživatele. Na základě obdržené JSON odpovědi se vytvoří několik návrhů, na které může uživatel kliknout a tím zvolit místo.

Možnost vylepšení:

- Ošetření onSubmit na úrovni django (doposud toto řeší kus jquery kódu)
- Řádná sanitace vstupů (z nějakého důvodu při psaní kódu nešly použít bind variables ve spojení s LIKE '%:variable%'. Povoleno jsou v tuto chvíli jen alfanumerické znaky a mezery, cokoli jiného vůbec do databáze nepřejde dát)
- Lepší vyřešení vytváření "tlačítek"/karet pro výběr na straně javascriptu
- Vyhledávání podle polohy zařízení

4) Administrátorské rozhraní

Postačilo by použít výchozí django admin stránku a trochu vyladit, aby vyhovovala potřebám správců

5) Pipeliny, automatické nasazení

V současnosti vše běží na apache serveru a když na [githubu projektu](#) dojde k nějaké změně v masteru, je třeba ručně stáhnout novou verzi. Bylo by pohodlné nové verze automaticky nasadit na server.

6) Migrace databáze a začlenění do struktury použité ve framewokru django

Z hlediska života tohoto projektu by s trochou dobré vůle nemuselo být nutně třeba s databází cokoli dělat, protože ten půl rok, rok to takhle vydržet půjde. I tak jsem se ale pustil do převodu z Oracle databáze na FITu na PostgreSQL a přepsání raw sql dotazů `Django.models`

7) Cachování a optimalizace výkonu

Není nezbytně nutné pro každého návštěvníka provádět dotaz na databázi znovu a znovu - hlavně pro jednu a tu samou stránku. V současnosti probíhá cachování jen na úrovni DNS na Cloudlfaru, Django ale cachovat také (nějak) umí.

Také by mohlo být užitečné mít k dispozici lokální kopii dokumentů a obrázků. S tím by byla spojena možnost automaticky vytvořit náhledové obrázky, které by se posílaly místo obrázků v plném rozlišení, nebo také archivovat externí dokumenty.