

Zobrazování dat v mapě

Jakub Černý (cernyj86)

Semestrální práce

Cíl práce:

Cílem této práce bylo nasbírat data (konkrétně data o počasí) pro všechny obce České republiky a zobrazit je v mapě. Nešlo o žádná konkrétní data, ale spíše o technické provedení grafu, tj. načtení polygonové struktury a zobrazení hodnot. Dále bylo třeba data převést graf do nějaké aplikace a vytvořit interaktivní prvky.

Těžení dat:

Z webu <https://www.in-pocasi.cz> je možné získat archivní data o počasí, konkrétně záznamy o teplotě, srážkách, tlaku a dalších parametrech. V semestrální práci se podařilo napsat kód, který dokáže stáhnout záznamy z téměř 250 meteorologických stanic, zpracovat je a přiřadit k obcím ČR. Stahování dat proběhlo poměrně bez problému, protože stránka má jasnou strukturu umožňující snadné scrapování.

Zpracování dat:

Tato část byla poměrně problematická, neboť názvy meteorologických stanic se občas neshodují s názvy obcí a neobsahují ani žádný kód, který by je k obci přiřazoval. Část stanic byla označena díky (alespoň částečné) shodě názvů a pár desítek z nich bylo označeno ručně. Data byla propojena s kódy z číselníku obcí, který je k dispozici na <https://www.czso.cz>. Pomocí dat z webu <https://www.rozpocetobce.cz>, která obsahují souřadnice všech obcí, bylo možné spočítat vzájemné vzdálenosti všech obcí a následně tato data použít k přiřazení nejbližší met. stanice ke každé obci (nebylo nutné počítat všechny vzdálenosti, ale pro potřeby autora byl proveden kompletní výpočet). Data z uvedeného webu obsahují i další informace o obcích, které měly být použity jako nadstavba do finální mapy. Vyskytovalo se v nich ovšem několik chybějících hodnot, takže pro úplnost datasetu bylo třeba vyscrapovat vše potřebné z Wikipedie. Další komplikací byl fakt, že met.stanice se dělí na klimatické a soukromé a každá měří trochu jiné hodnoty. Proto bylo nutné paralelně pracovat s dvěma datasety a poté vhodně vybrat z které stanice bude pocházet přesnější hodnota, protože je blíže. Data o počasí obsahovala poměrně dost chybějících hodnot, které byly doplněny průměrnou hodnotou (za účelem vylepšení finálního grafu).

Zobrazení dat:

Tato část byla problematická především do doby, než se podařilo najít polygonovou strukturu ČR a zobrazit ji pomocí některého z vhodných packageů. Vzhledem k tomu, že jsem chtěl finální graf zobrazit pomocí streamlitu, rozhodl jsem pro plotly, které umožňuje hezké popisky s dodatečnými informacemi. Inspiroval jsem se kódem ze cvičení a vytvořil jednoduchou aplikaci (bohužel pouze „offline“ verzi) ve streamlitu, která umožňuje vybrat proměnnou a zobrazit ji v mapě. Dále je v ní možné stáhnout další data o počasí pro konkrétní den. Nepodařilo se mi ovšem správně „cache-ovat“, takže se data načítají znovu po každém kliknutí v aplikaci.

Finální výsledek:

Co se týče zobrazování dat, práce se z větší části povedla, neboť jsem se naučil hezky zobrazit jakákoliv data na obecní úrovni. V tomto směru se nepovedla pouze jedna věc, a sice že polygony nejsou úplně aktuální, ale několik let staré a pár obcí v nich neodpovídá skutečnosti (viz například bílé místo v grafu- vojenský újezd Brdy).