

OSW – Semestrální práce

Josef Štěřovský

15. října 2023 - 6. ledna 2024

Checkpoint 1

Téma a otázka

V rámci projektu bych se rád zabýval legalitou konzumování alkoholu na veřejnosti na území Hlavního města Prahy. V Praze je obecně povoleno požívat alkohol na veřejných prostranstvích, za posledních 10 let však stále přibývá výjimek z tohoto pravidla a v legislativě je poměrně obtížné se vyznat. Hlavní město Praha totiž v současnosti neposkytuje mapu, kde by byly vyznačeny oblasti platnosti zákazu požívání alkoholu, a v těchto oblastech z velké části není umístěno ani žádné varování, aby bylo zřejmé, že v nich zákaz platí. Otázka by tedy mohla znít:

Na kterých veřejných prostranstvích na území Hlavního města Prahy je zákonem povoleno požívat alkohol?

Zdroje a legislativa

V současnosti platným právním předpisem věnujícím se tomuto tématu je [Obecně závazná vyhláška č. 10/2022 Sb.](#) Hlavního města Prahy. Do přílohy k vyhlášce je zanesen seznam tříděný podle městských částí, který vyjmenovává veřejná prostranství ([§ 14b zákona č. 131/2000 Sb.](#)), kde je zakázáno požívat alkohol. Kromě těchto jmenovaných výjimek zákaz platí také v okolí

- dětských hřišť a pískovišť,
- školských zařízení a škol ([§ 7 zákona č. 561/2004 Sb.](#)),
- zdravotnických zařízení ([§ 4 odst. 1 zákona č. 372/2011 Sb.](#)),
- stanic metra,

a to ve všech případech v okruhu 100 m. Dále platí zákaz také na zastávkách veřejné linkové dopravy a v jejich bezprostředním okolí.

Výběr zdrojů dat

Vzhledem k tomu, že veřejná prostranství, na nichž je požívání alkoholu zakázáno městskou vyhláškou, jsou uvedena pouze v této vyhlášce v podobě seznamu jejich názvů, bude stěžejní získat informace o geografické poloze veřejných prostranství a platnost zákazu správně namapovat na polohu. K tomu by bylo ideální získat přístup k datům o [veřejně přístupných plochách](#) na území Prahy, která jsou v současnosti nedostupná. Prozatím budu předpokládat, že tento přístup nezískám, ale možná budou také k určení polohy prostranství stačit další zdroje uvedené níže.

Veřejně přístupné datové sady:

- OpenStreetMap má na území Prahy dobré pokrytí, co se týče vyznačení ulic, parků atd. Pomocí nástrojů Overpass nebo Nominatim by šlo snadno hromadně získat všechny relevantní lokality
 - OpenStreetMap jako zdroj dat uvádí [RÚIAN](#), možná by informace o přístupnosti pozemků šlo dostat přímo odsud
- [Dětská hřiště Praha](#) — polohu hřišť lze jednoduše získat z JS skriptu
- [Atlas školství](#) — obsahuje přehledné a aktuální informace o školách včetně GPS souřadnic, které by šlo vyscrapovat, zahrnuje základní, střední i vysoké školy, VOŠ atd.

Checkpoint 2

Extrakce RDF a tvorba ontologií

Pro přípravu RDF souborů jsem použil různé nástroje v Pythonu, konkrétně byly velmi užitečné knihovny `rdflib` pro tvorbu grafů a export Turtle souborů a `geopandas` pro zpracování geografických dat. Pro všechny datasety, které přímo obsahují umístění/geometrii objektů, jsem převedl souřadnice do systému WGS84 a do WKT formátu.

Pro všechny datasety byly vytvořeny následující soubory:

- skript v Pythonu pro vytvoření RDF souboru ve složce `pipeline` (např. `pipeline/dětské-hřiště.py`),
- OWL ontologie v RDF Turtle formátu ve složce `dataset_ontologies` (např. `dataset_ontologies/dětské-hřiště-ontologie.ttl`),
- dataset v RDF Turtle formátu ve složce `rdf_data` (např. `rdf_data/dětské-hřiště.ttl`).

Pouze v případě datasetu `dětské-hřiště` bylo nejjednodušší získat zdrojová data web scrapingem, viz soubor ve složce `scraping`. Postupy pro získání zdrojových souborů pro ostatní datasety:

- `zdravotnické-zařízení` – na stránce <https://nrpzs.uzis.cz/index.php?pg=vyhledavani-poskytovatele--pro-verejnost> v rozšířeném vyhledávání vybrat kraj Hlavní město Praha, výsledek vyexportovat jako CSV soubor
- `školské-zařízení` – ve formuláři <http://stistko.uiv.cz/registr/vybskolrn.asp> zvolit území „CZ01 Praha“, vyexportovat výsledky (export má příponu `.xls`, ale jedná se o HTML tabulku)
- `veřejné-prostranství` – získat soubor Výměnného formátu RÚIAN <https://www.cuzk.cz/vfr> s názvem `20230930_0B_554782_UKSH.xml` nebo ekvivalentní (kompletní datová sada pro hl.m. Prahu)
- `adresní-bod` – tentýž soubor jako u `veřejné-prostranství`
- `zastávka-veřejné-dopravy` – na webu Geoportálu Praha [https://geoportalpraha.cz/vyhledavani?topic=data&type=\[opendata\]](https://geoportalpraha.cz/vyhledavani?topic=data&type=[opendata]) vybrat dataset „Pražská integrovaná doprava - zastávky (aktuální stav)“ a stáhnout jako Shapefile
- `vstup-do-metra` – na webu Geoportálu Praha [https://geoportalpraha.cz/vyhledavani?topic=data&type=\[opendata\]](https://geoportalpraha.cz/vyhledavani?topic=data&type=[opendata]) vybrat dataset „Pražská integrovaná doprava - vstup do metra“ a stáhnout jako Shapefile

Ve všech skriptech je proměnná `SRC_FILE_PATH`, kterou je nutné nahradit cestou k danému souboru.

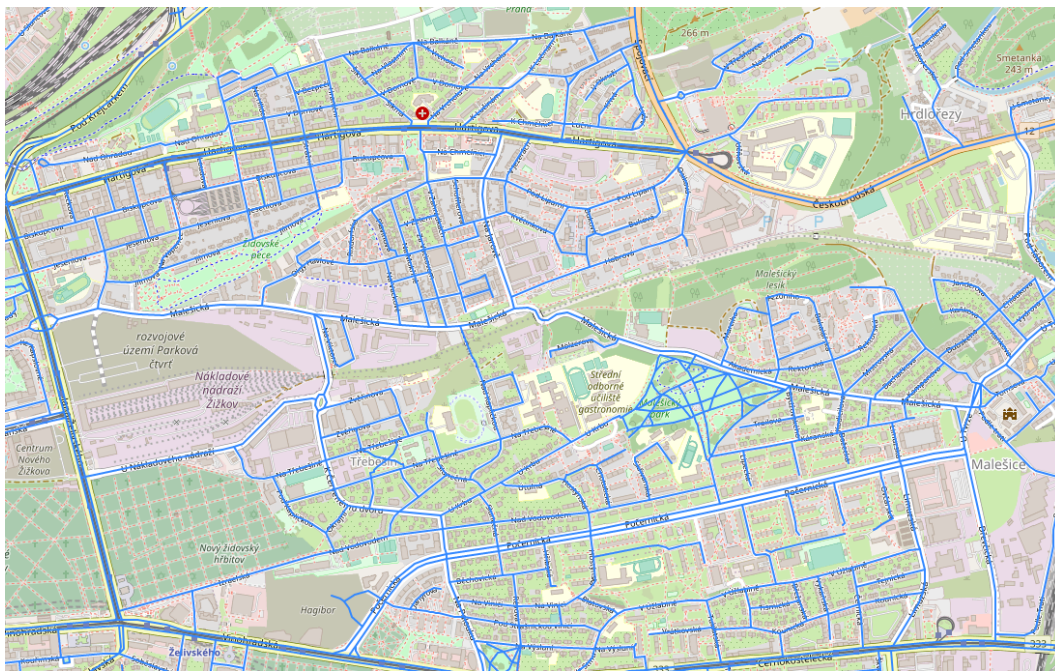
Ontologie jsem vytvořil v Protégé. Ontologie obsahující geometrie by měly být kompatibilní s GeoSPARQL, správnost jsem ověřil SPARQL dotazy s filtrováním pomocí `geof:sfWithin` v GraphDB. Dataset `školské-zařízení` bohužel nemá vlastní geometrii ani RÚIAN kódy, nicméně mělo by být možné ho poměrně dobře namapovat na adresní body RÚIAN podle ulice, PSČ a čísla domu – viz SPARQL dotaz v souboru `mapovani_adres.sparql`.

Vymezení pojmu „veřejné prostranství“

Stěžejním problémem pro zodpovězení otázky vytyčené v checkpointu 1 je správně určit, co považujeme za veřejné prostranství. Na území Prahy se vysloveně nevztahuje [zákon č. 128/2000 Sb.](#) o obcích (viz §150 téhož), proto je relevantní opravdu pouze definice ze [§ 14b zákona č. 131/2000 Sb.](#), která říká, že se jedná obecně o veřejně přístupné prostory sloužící obecnému užívání.

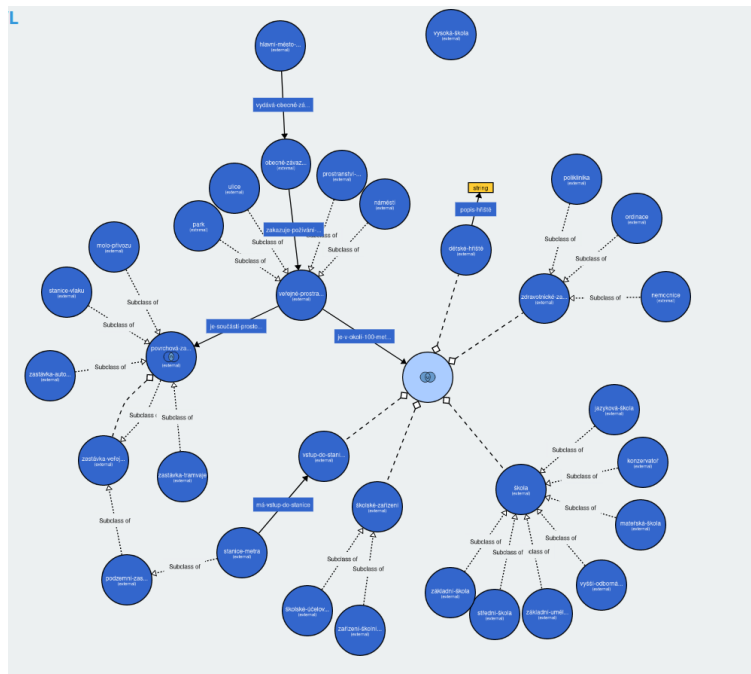
Bohužel se zdá, že neexistuje dataset, který by zcela uspokojivě pokrýval oblasti, které odpovídají takovému popisu. Na OpenStreetMap jsou mnohé parky určeny poměrně dobře, problém je, že uživatelé používají různé tagy a bylo by asi poměrně obtížné získat z OpenStreetMap kompletní reprezentaci takových prostor v Praze.

Poměrně dobré pokrytí má již zmíněný RÚIAN (Registr územní identifikace, adres a nemovitostí) v kolekci „Ulice“. Nevýhodou je, že jak již napovídá název kolekce, RÚIAN chápe všechna veřejná prostranství jako čáry, nikoliv polygony. Tudíž např. park není prostorová oblast, nýbrž množina cest vedoucích přes tento park. Na obrázku níže jsou modrými čarami vizualizovány ulice RÚIAN v části Prahy 3 a 10.



Vzhledem k tomu, že k aspoň částečně úspěšnému řešení otázky bude už tak potřeba integrace většího počtu dalších datasetů, nebudu kompilovat další datasety veřejných prostranství a omezím se pouze na RÚIAN.

Konceptuální model (ve slozce `conceptual_model`) byl prozatím přepracován do podoby vyobrazené na dalším obrázku pomocí WebVOWL. Otázka je, do jaké míry bude nutné integrovat všechny podtřídy, protože část z nich nejspíš nebude pro zodpovězení otázky relevantní. Vhodné by mohlo být podrobnější propojení datasetu zastávek veřejné dopravy s konceptuálním modelem, aby bylo možné rozlišit např. nádraží od zastávky autobusu (pro odhad rozlohy daného prvku).



Zdroje dat

Pro část dat jsem našel důvěryhodnější zdroje, než byly uvedeny v checkpointu 1, a to konkrétně:

- **Národní registr poskytovatelů zdravotních služeb (NRPZS)** každý poskytovatel má jednu nebo více adres, kde působí, seznam poskytovatelů zahrnuje jak nemocnice, tak ordinace zubních lékařů apod.

- **Adresář škol a školských zařízení** – přístup do adresáře poskytován MŠMT, všechny školy a školská zařízení mají adresu a RÚIAN kód

Žádný z těchto zdrojů sice neposkytuje geodata, formátování adres je však dostatečně dobré na to, aby bylo možné je lokalizovat pomocí adresních bodů z RÚIAN. V případě dat z MŠMT lze dokonce přímo použít RÚIAN kód pro propojení datasetů.

V případě dat z Geoportálu Praha týkajících se zastávek PID jsem narazil na nedostatečnou dokumentaci a zřejmě chybné souřadnice v CSV souboru, ale v odpovídajícím Shapefile byly správné S-JTSK souřadnice, proto jsem použil tento Shapefile jako zdroj dat.

Checkpoint 3

Poznámky k mapování a propojování datasetů

Ideálně by měla být co největší část propojování datasetů provedena na úrovni mapování ontologií. V mém případě např. mám dva datasety, které obsahují adresy míst, ale nikoliv jejich geometrii, a tak je nutné tyto datasety propojit s RÚIAN adresními body, aby je bylo možné použít k výpočtu vzdáleností těchto míst od veřejných prostranství. Toto propojení by mělo být teoreticky možné realizovat pomocí `owl:hasKey` a RÚIAN kódu, případně kombinace čísla domu a ulice. Bohužel se mi však nepodařilo toto propojení implementovat na úrovni OWL tak, aby bylo automaticky vyinferováno v GraphDB, a musel jsem ho nakonec realizovat SPARQL dotazy.

Kvůli různým chybám v zápisu adres se mi nepodařilo umístit na mapu malý počet zdravotnických zařízení (cca 0,25 %) a větší počet školských zařízení (cca 6 %). U školských zařízení se však může také stát, že se jedná o zařízení provozované pražskou školou, které ovšem není na území Hlavního města Prahy. Školská zařízení jsou identifikována kódem adresního bodu RÚIAN. Namátkovým vyhledáváním na [webu VDP ČÚZK](#) bylo ověřeno, že se skutečně vesměs jedná o mimopražské adresy, ale značnou část tvoří bohužel také kódy, které jsou buď chybné, nebo již nejsou platné.

Velký vliv na rychlost dotazů má pořadí operací (především při vytváření nových proměnných pomocí `bind`, které by mělo být použito tak, aby bylo nutno vytvořit co nejméně hodnot nové proměnné).

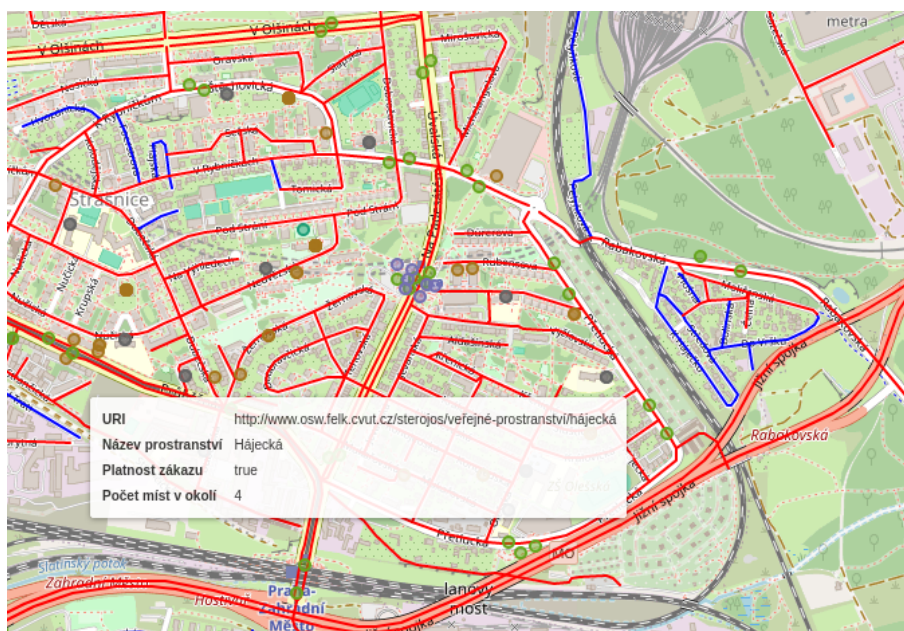
Vzhledem k tomu, že u datasetu `školské-zařízení` nebylo zjišťováno, zda konkrétní objekty jsou školy nebo školská zařízení, a podmínky vyhlášky na tomto nezávisí, byla v rámci mapování vytvořena pomocná třída `škola-nebo-školské-zařízení`, pod níž spadají oba typy objektů. Bohužel ale GraphDB z definice této třídy (`škola or školské-zařízení`) zřejmě nedokáže vyvodit, že pokud je veřejné prostranství v okolí objektu z tohoto datasetu, pak musí být v okolí školy nebo školského zařízení. To je důležité, protože definice veřejného prostranství se zákazem pití alkoholu v konceptuálním modelu používá třídy `škola` a `školské-zařízení`. Proto bylo nutné ještě doplnit hrany do grafu k veřejným prostranstvím, které ze všech objektů zákazu měly v okolí pouze školu nebo školské zařízení – viz složka `sparql_fix`.

Postup pro naimportování řešení do GraphDB

1. Naimportovat všechny datasety a ontologie z Checkpointu 2 do repozitáře s OWL rulesetem (je potřeba pro odvození členů některých tříd). Mělo by trvat cca 1 minutu.
2. Naimportovat soubor `mapování.ttl` (mapování datasetů). Mělo by zabrat několik sekund.
3. Spustit všechny SPARQL soubory ze složky `sparql_mapovani` v Checkpointu 3. Tímto se dokončí části mapování, které se mi nepodařilo realizovat na úrovni OWL. Tyto updaty proběhnou téměř instantně.
4. Spustit všechny SPARQL soubory ze složky `geosparql`. Doplní se tak vztahy z konceptuálního modelu, jejichž význam závisí na geometrické vzdálenosti mezi veřejnými prostranstvím a jinými objekty. Tyto soubory mohou v GraphDB běžet déle (kolem 10 minut) kvůli složitosti geometrických operací, a možná také proto, že nejsou dostatečně optimalizované.
5. Spustit SPARQL ve složce `sparql_fix`.
6. Spustit soubor `export.rq`. Soubor uloží nové znalosti (platnost zákazu na daném veřejném prostranství, a z jakých důvodů tam zákaz platí) do nového grafu, který by mělo být možné vyexportovat tak, jako v přiloženém souboru `výsledek.ttl`.

Vizualizace

Pro zobrazení výsledků jsem vytvořil jednoduchou vizualizaci, která zobrazuje veřejná prostranství a objekty, v okolí nichž platí zákaz. Data jsem vyexportoval v CSV z GraphDB pomocí SPARQL dotazů a použil jsem Geopandas funkci `explore` k zobrazení těchto dat na mapě. Prostranství, kde platí zákaz, jsou vyznačena červeně, a ostatní jsou vyznačena modře. Funkce `explore` používá JS knihovnu Leaflet, takže lze data bez problémů interaktivně prohlížet i ve webovém prohlížeči.



Závěr

Jak je vidět z vizualizace i z počtu zbývajících veřejných prostranství, platí zákaz požívání alkoholu v Praze téměř všude. Je to částečně způsobeno tím, že k zastávkám veřejné dopravy jsem přistupoval velmi opatrně, a také tím, že jakmile platí zákaz v kterékoliv části veřejného prostranství, byl zákaz aplikován na celé toto prostranství. To ale víceméně odpovídá realitě – když například někomu řekneme, že v ulici Pelyňkové je legální pít alkohol, asi by bylo nevhodné, kdyby to platilo jen v polovině této ulice.

Práce pro mě byla přínosná, protože jsem se dozvěděl o kvalitních veřejných zdrojích informací např. o zdravotnických zařízeních, nicméně původně jsem doufal, že se dostanu i k řešení nějaké složitější otázky. Myslím si, že tato semestrální práce poměrně dobře ilustruje, že Obecně závazná vyhláška č. 10/2022 Sb. Hlavního města Prahy je dost komplikovaná. Je pravděpodobné, že i značná část výjimek, která vyhláška stanovuje, je zbytečná, protože na ně lze aplikovat některé z pravidel stanovených vyhláškou. Především by však zodpovězení otázky, kterou řešila tato práce, mělo být snadné v zájmu veřejnosti i policie, která by zákazy měla vymáhat. Proto by možná bylo vhodnější tuto vyhlášku nahradit plošným zákazem, ze kterého by se pak dělaly výjimky pro různé akce, které ve vyhlášce tak jako tak jsou i v současné podobě.