

Checkpoint 3

- Marian Krotil - krotima1

Úprava checkpoint 2

cesta: conceptual_model/ pipeline/ rdfized_data/ dataset_schema/

Pro dosažení třetího kontrolního bodu jsem musel provést několik klíčových úprav v procesu zpracování dat. Prvním krokem bylo přepracování souborů obsahujících datová schémata tak, aby byla zajištěna kompletní integrace prostorových vlastností. Tato modifikace vyžadovala pečlivou revizi transformačních skriptů, které byly následně aktualizovány a přizpůsobeny nově přidaným prvkům.

Při finálním dotahování projektu do konce jsem se potýkal s výzvou spojování dat. Bylo nezbytné znovu upravit datová schémata, tentokrát s důrazem na propojení dat podle druhu odpadu. Tímto krokem jsem zajistil, že jednotlivé části projektu vzájemně efektivně komunikují a vytvářejí komplexní celek.

Jedním z klíčových prvků mé práce bylo extrahování pojmového obsahu z legislativních dokumentů a jeho začlenění do samostatných schémat. Tento postup vytvořil robustní mezivrstvu, která sloužila jako spojovací článek mezi konkrétními daty a konceptuálním modelem ontologie.

Během tohoto procesu jsem opakovaně upravoval transformační skripty, abych zahrnul nově přidané vlastnosti a zajistil, že celý systém funguje koherentně a efektivně. Tato iterativní a pečlivá práce na zdokonalování projektu byla klíčovým prvkem v dosažení zamýšleného třetího kontrolního bodu.

Mapping

cesta: mapping/mapping_{dataset}.ttl

K vytvoření mapování jsem využíval nástroj Protege, kde jsem každé datové schéma propojil s odpovídající ontologií konceptuálního modelu. Tento proces zahrnoval vytvoření mapování i pro taxonomii odpadků. Vytvoření těchto mapování bylo usnadněno použitím schématu pojmů z legislativy, které efektivně propojovaly jednotlivá data s konceptuálním modelem. Datová schémata tak byla schopna přebírat klíčové vlastnosti těchto pojmů, což přispělo k celkové koherentnosti a konzistenci projektu.

Důkaz integrace

cesta: integration_tests/query.rq

Během jednotlivých fází mapování jsem systematicky testoval funkčnost a propojení komponent jednotlivých schémat s odpovídajícími daty. Tento proces probíhal v nástroji Protege, kde jsem využíval reasoner pro ověření logiky a také prováděl krátké dotazy ve SparQL v prostředí programu GraphDB. Po dokončení integračních testů jsem provedl

komplexní testování celé ontologie s namapovanými daty, opět s využitím reasoneru a dotazů v programu GraphDB. Sparql dotazy testující integraci jsou uloženy v souborec.

Otázka

“Kde je nejbližší kontejner na daný odpad?”

cesta: *queries/sparql.rq*

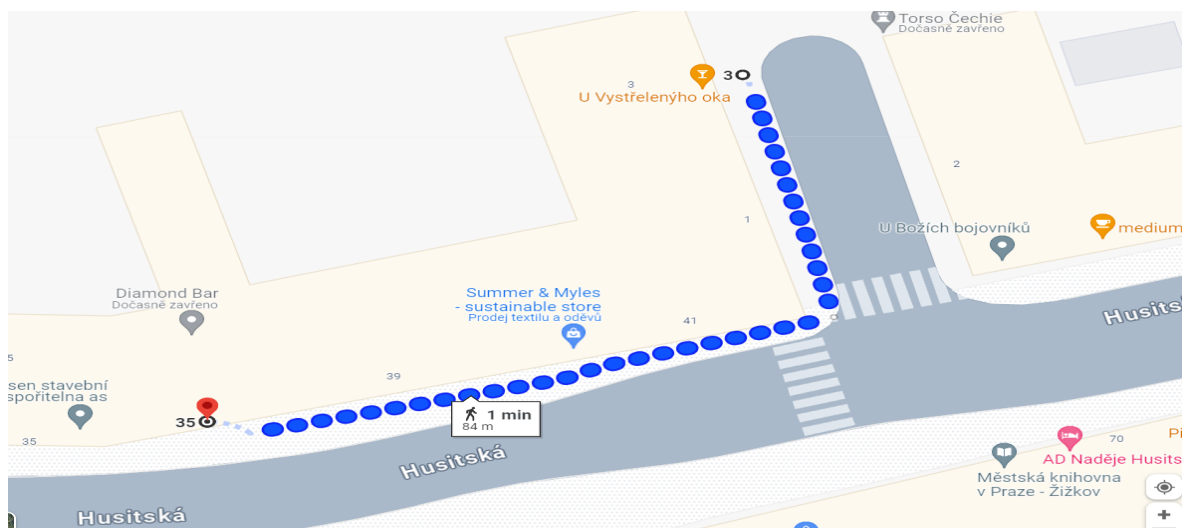
Otázku jsem byl schopen odpovědět pomocí konceptuálního modelu, integrací datových sad a sestrojení SparQL dotazu. Příkladám následující dotaz: Kam mohu vyhodit lahvič od Plzně, když se nacházím u Vystřeleného Oka?

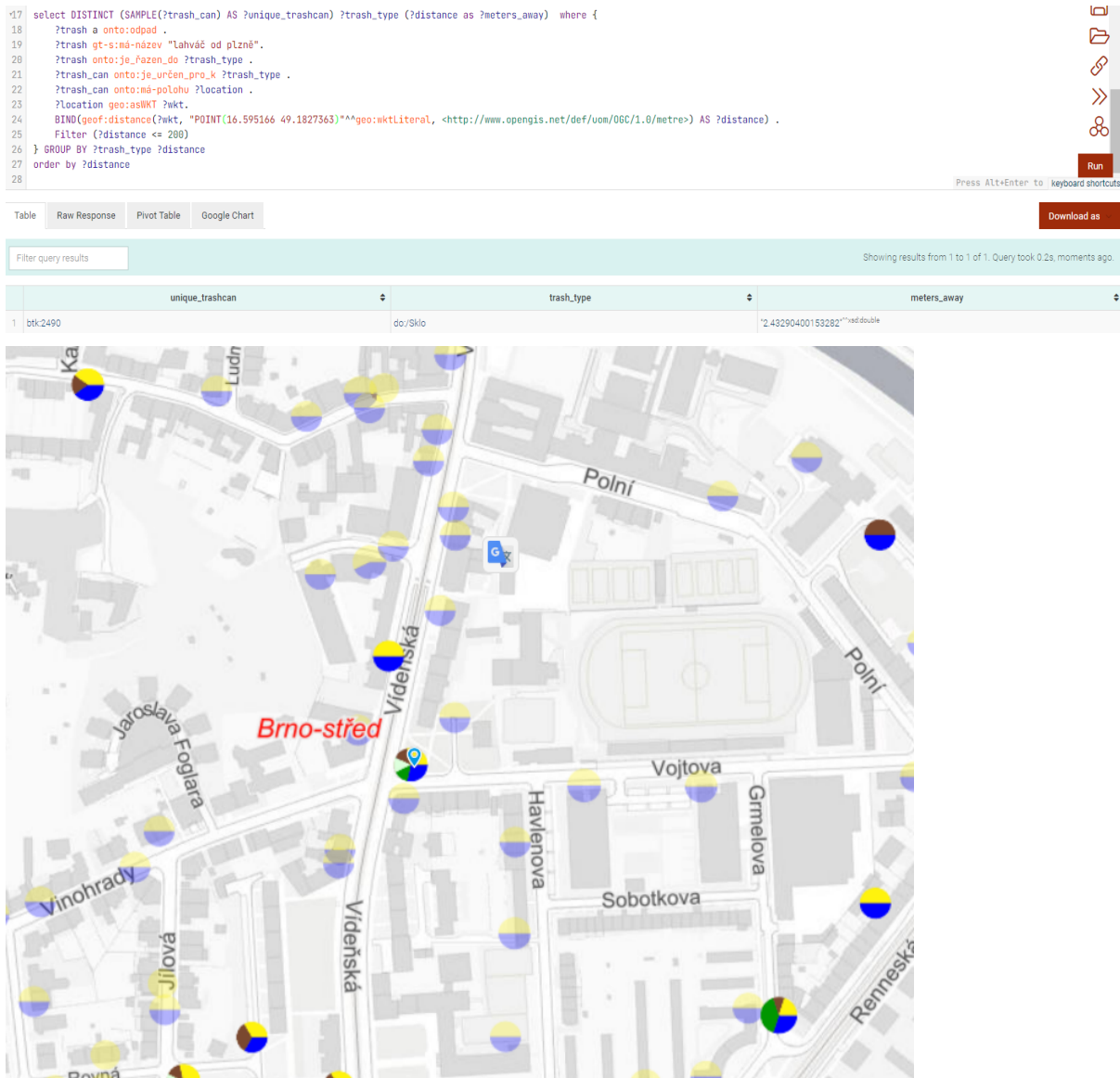
Pro konvertování adres na longitude a latitude jsem používal stránku s [odkazem](#).

Ostatní dotazy a jejich výsledky jsou přiloženy v adresáři *queries/*.

```
select DISTINCT (SAMPLE(?trash_can) AS ?unique_trashcan) ?trash_type (?distance as ?meters_away) where {
  ?trash a onto:odpad .
  ?trash gt-s:má-název "lahvič od plzně".
  ?trash onto:je_fazen_do ?trash_type .
  ?trash_can onto:je_určen_pro_k ?trash_type .
  ?trash_can onto:má-polohu ?location .
  ?location geo:asWKT ?wkt.
  BIND(geof:distance(?wkt, "POINT(14.452275 50.086634)"^^geo:wktLiteral, <http://www.opengis.net/def/uom/OGC/1.0/metre>) AS ?distance) .
  Filter (?distance <= 200)
} GROUP BY ?trash_type ?distance
order by ?distance
```

	unique_trashcan	trash_type	distance
1	ptk15443	do/Sklo	"54.994845124825666"^^xsd:double
2	ptk9244	do/Sklo	"60.31325798879625"^^xsd:double
3	ptk10880	do/Sklo	"88.23699095342158"^^xsd:double
4	ptk1369	do/Sklo	"89.0259050795423"^^xsd:double
5	ptk11402	do/Sklo	"115.30015531134393"^^xsd:double





Můj názor k semestrální práci

Semestrální práce mi poskytla cenné zkušenosti a dovednosti v oblasti sémantického modelování a propojování dat. Práce s nástroji jako Protege a GraphDB mi umožnila lépe porozumět konceptům OWL (Web Ontology Language) a RDF (Resource Description Framework). Během vytváření mapování mezi datovými schématy a ontologií jsem se naučil nejen pracovat s konkrétními nástroji, ale také lépe chápat význam propojení a odkazů mezi jednotlivými entitami.

Při testování funkčnosti a propojení jednotlivých částí datových schémat a odpovídajících dat jsem využíval nástroje jako Protege s testováním pomocí reasoneru a krátkými dotazy v jazyce SPARQL v programu GraphDB. Integrační testy následně zahrnovaly celou ontologii s namapovanými daty a testování reasonerem spolu s dotazy v programu GraphDB.

Tato práce mě naučila nejen technickým dovednostem spojeným s konkrétními nástroji, ale také obecnějším konceptům sémantického zpracování dat a významu ontologických struktur.

Získal jsem hlubší vhled do problematiky mapování a propojení dat, což považuji za klíčové pro efektivní práci v oblasti sémantického webu a ontologií.

Bohužel, ale někdy bylo delší pochopit a správně namodelovat daný checkpoint z důvodu nedostatku informací.

Checkpoint 2

- Marian Krotil - krotima1

Ontologie konceptuálního modelu

cesta: conceptual_model_ontology/ontologie.ttl

Pro vytvoření ontologie v rámci mé semestrální práce jsem využil software Protege. S pomocí tohoto nástroje jsem modeloval třídy, jejich závislosti a vztahy podle lehce upraveného konceptuálního modelu, který je poskytnut na konci checkpointu 2 a také ve složce conceptual_model_ontology. Protege se ukázal jako efektivní prostředek pro detailní navrhování ontologií, umožňující definici komplexních struktur a vzájemných vztahů v souladu s požadavky mé práce. K popisu pojmů jsem použil slovník SKOS a pro stereotypizaci a propojení pojmů slovníky RDF(S) a OWL.

Ontologie datových sad

cesta: dataset_schema_ontologies/

V semestrální práci jsem celkem použil 5 datových sad z veřejně publikovaných otevřených dat a 2 vlastně vytvořené privátní datové sady webscrapnuté z internetu.

1. Brno sběrné dvory
2. Brno smíšené kontejnery
3. Brno tříděné kontejnery
4. Praha sběrné dvoru
5. Praha tříděné kontejnery
6. Kam patří - taxonomie odpadků (<https://www.samosebou.cz/kampatri/>)
7. Jak správně třídit - taxonomie odpadků (https://portalzp.praha.eu/jnp/cz/odpady/pro_obcany/jak_spravne_tridit#X-20140415111471)

Datová schémata jednotlivých datových sad jsou poskytována v **adresáři** "*dataset_schema_ontologies/*".

1. Brno sběrné dvory

Ontologie datové sady Brno sběrné dvory byla vytvořena v textovém editoru. Používá slovník RDF(S), XSD a SKOS. Používají se pouze pojmy pro dataset.

2. Brno smíšené kontejnery

Ontologie datové sady Brno smíšené kontejnery byla vytvořena v textovém editoru. Používá slovník RDF(S), XSD a SKOS. Používají se pojmy pro tento dataset a rovněž je zde definovány nadřazené pojmy legislativy ze Zákon č. 541/2020 Sb., Zákon o odpadech.

3. Brno tříděné kontejnery

Ontologie datové sady Brno tříděné kontejnery byla vytvořena v textovém editoru. Používá slovník RDF(S), XSD a SKOS. Používají se pojmy pro tento dataset a rovněž je zde definovány nadřazené pojmy legislativy ze Zákon č. 541/2020 Sb., Zákon o odpadech. Dále je zde definovaná třída druhu tříděného odpadu, pro řazení kontejnerů.

4. Praha sběrné dvory

Ontologie datové sady Praha sběrné dvory byla vytvořena v textovém editoru. Používá slovník RDF(S), XSD a SKOS. Používají se pojmy pro tento dataset a rovněž je zde definovány nadřazené pojmy legislativy ze Zákon č. 541/2020 Sb., Zákon o odpadech. Dále je zde definovaná třída provozovatel, pro řazení provozovatelů sběrných dvorů.

5. Praha tříděné odpady

Ontologie datové sady Praha tříděné odpady byla vytvořena v textovém editoru. Používá slovník RDF(S), XSD a SKOS. Používají se pojmy pro tento dataset a rovněž je zde definovány nadřazené pojmy legislativy ze Zákon č. 541/2020 Sb., Zákon o odpadech. Dále je zde definovaná třída druhu tříděného odpadu, pro řazení kontejnerů.

6. Kam patří a Jak správně třídit

Ontologie datových sad Kam patří a Jak správně třídit reprezentující taxonomii odpadů byla vytvořena pouze jedna v textovém editoru. Používá slovník RDF(S), XSD a SKOS. Dále jsou využity pouze pojmy pro tento dataset a definice odpadu z legislativy ze Zákonu č. 541/2020 Sb., Zákon o odpadech.

Pipeline pro RDFizaci datových sad

cesta: pipeline/

Pro pipeline rdfizaci je použit programovací jazyk Python.

Datové sady 1-5 obsahují prostorová data (longitude a latitude) pojednávající o umístění. Prostorové souřadnice ve formátu SJTSK datových sad 4-5 byly transformovány na WGS reprezentaci pomocí knihovny "pyproj". Následně všechna prostorová data jsou reprezentována jako body geoSPARQL například: `"POINT (16.565526400000067 49.168732)"^^geo:wktLiteral`; Další hodnoty dat byly transformovány do rdf pomocí RDFlib a geopandas.

Program lze spustit pomocí příkazu: `python3 pipeline/transform_*.py`

Datové sady 6-7 byly webscrapnuty pomocí "BeautifulSoup" knihovny, následně pomocí knihovny csv uloženy do souboru. Tento script lze najít ve složce "scripts/". Pro rdfizační pipeline byl použit software "OntoText Refine" pro tyto datové sady, jelikož neobsahují prostorové souřadnice.

Program lze spustit pomocí nahrání mapování do OntoText Refine.

Pro demonstraci všech programů přikládám zdrojová data v csv a v geojson formátech.

Rdfizované datové sady

Pomocí pipeline byly transformovány všechny datové sady.

V rámci průběhu práce na checkpointu 2 jsem vytvořil schéma datových sad, následně ontologii konceptuálního modelu a RDFizační pipeline v python programovacím jazyce. Tato pipeline sloužila k transformaci dat do formátu RDF, což umožnilo efektivní zpracování a reprezentaci dat v ontologickém kontextu.

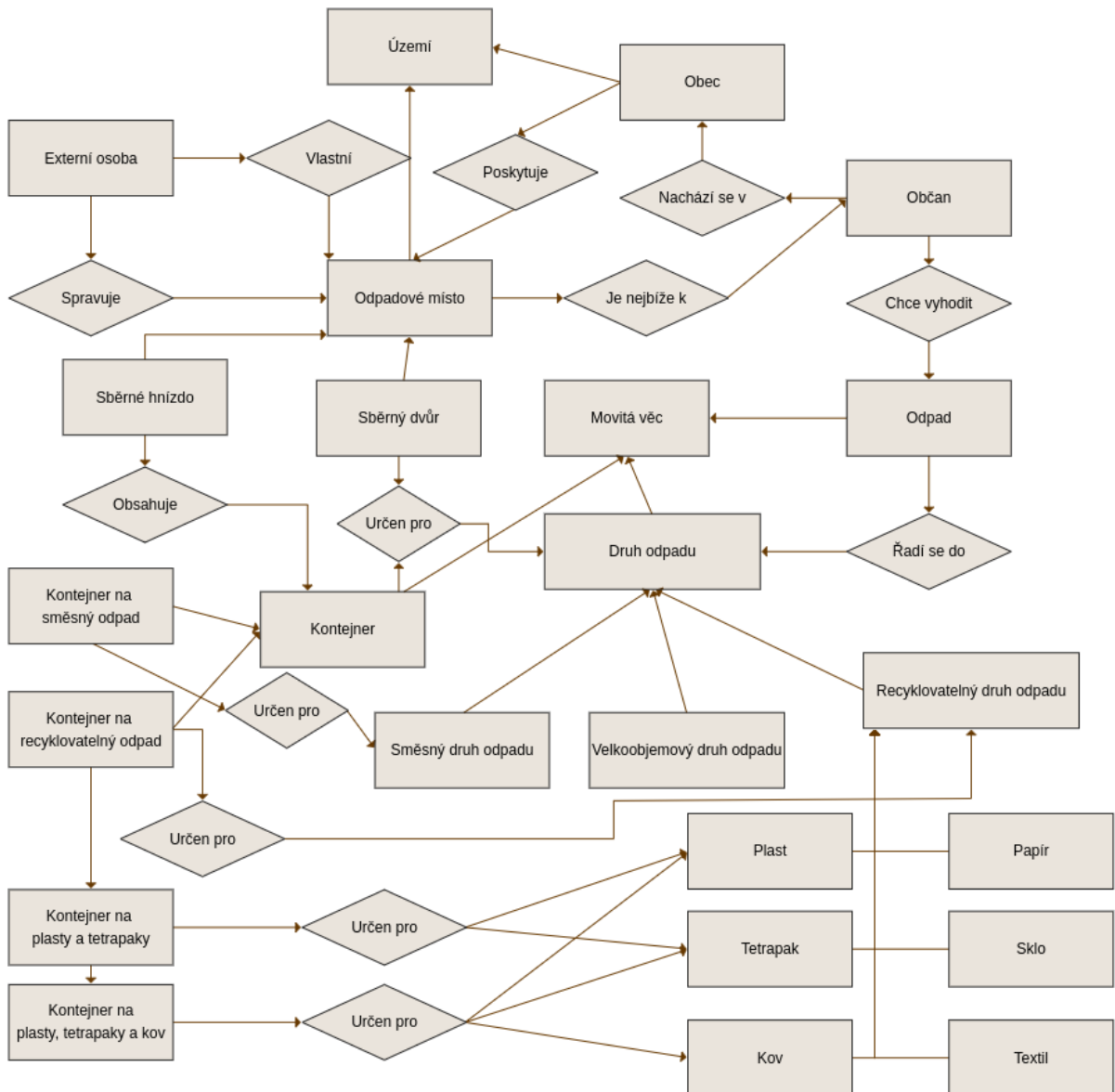
V průběhu vytváření ontologie v mé semestrální práci jsem čelil několika výzvám. Jednou z největších bylo hledání vhodné reprezentace prostorových dat. Po několika pokusech a zkoumání různých možností jsem zvolil určitý formát, který lépe vyhovoval specifikacím mého konceptuálního modelu.

Dalším klíčovým krokem bylo generování všech datových sad, což jsem úspěšně realizoval pomocí nástroje OntoRefine. Během procesu jsem si všiml, že některé datové sady neobsahují prostorová data, což bylo nejprve zjištěno až v pozdější fázi vývoje. Tato situace vyžadovala další úpravy a přizpůsobení mého přístupu k manipulaci s daty.

Celkově tyto výzvy představovaly příležitost k zdokonalení mého porozumění práci s ontologiemi a prostorovými daty, přičemž jsem se mohl lépe přizpůsobit specifickým potřebám projektu.

Upravený konceptuální model

Vyhlášky daných čtvrtí a měst pro dané kontejnery přijímající více tříděného odpadu jsou reprezentovány jako podtřídou Kontejner na recyklovaný odpad.



Checkpoint 1

Marian Krotil - krotima1

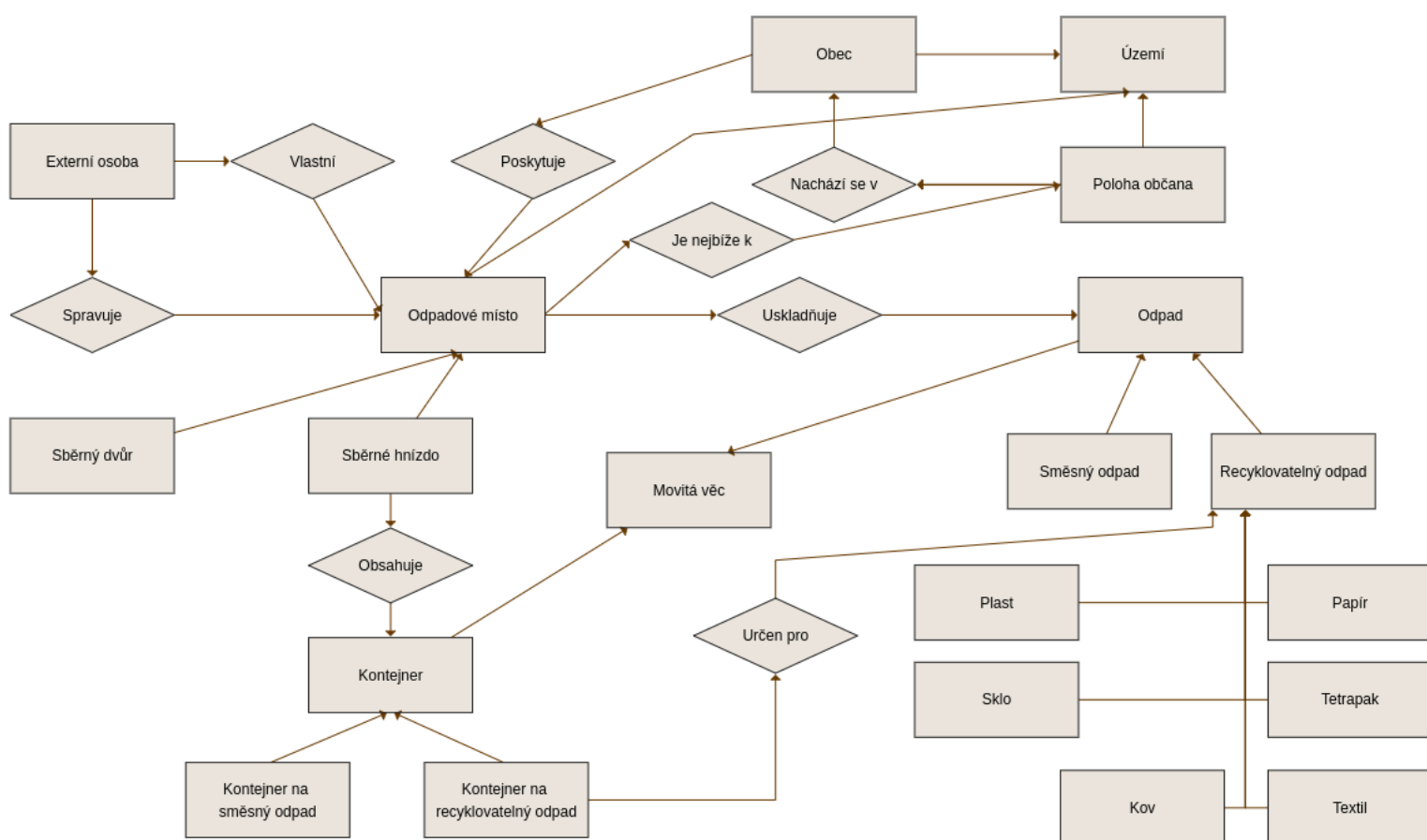
Otázka a motivace

“Kde je nejbližší kontejner na daný odpad?”

Touto otázkou se zabývá nejspíš každý producent odpadu. Ovšem hledání kontejnerů v neznámém prostředí není snadné a vyžaduje brouzdání na internetu. Proto jako řešení navrhuji jednoduchý systém, který spojí data s odpadovými místy publikovaná různými městy a odpoví tímto na položenou otázku.

Konceptuální model

Definice a relační vztahy týkajících se komunálního odpadu a věcí s ním spojených dle zákona jsem čerpal z *Vyhláška č. 445/2022 Sb. o podrobnostech nakládání s odpady*, *Vyhláška č. 8/2021 Sb. o Katalogu odpadů a posuzování vlastností odpadů*, *Zákon č. 541/2020 Sb. o odpadech* a *Vyhláška č. 381/2001 Sb. o stanovení katalogu odpadů*. Dále jsem čerpal z internetových zdrojů různých měst a stránek zabývajících se odpady. Z důvodů širokého členění odpadů dle zákona jsem si dovolil zjednodušit konceptuální model na očekávané využití systému a rovněž vynechat některé pojmy.



Datové modely

Rozhodl jsem se použít datasey na komunální kontejnery pro Prahu a Brno jako demonstraci. Neumím si nyní představit obtížnost práce, a proto popřípadě mám v plánu přidat či ubrat vhodné datasey.

1. Praha - Stanoviště tříděného odpadu
<https://opendata.praha.eu/datasets/https%3A%2F%2Fapi.opendata.praha.eu%2Fflod%2Fcatalog%2Fca964839-d448-4970-ba6e-92c64410c34a>
2. Praha - Stanoviště sběrných dvorů
<https://opendata.praha.eu/datasets/https%3A%2F%2Fapi.opendata.praha.eu%2Fflod%2Fcatalog%2F2a206f2d-16be-4be1-a2bf-a41fb3387076>
3. Brno - Kontejnery na separovaný odpad
<https://data.gov.cz/datov%C3%A1-sada?iri=https%3A%2F%2Fdata.gov.cz%2Fzdroj%2Fdatov%C3%A9-sady%2F44992785%2Fb3e9f4c735e76be7d336f05dd8b6ad47>
4. Brno - Odpadkové koše
<https://data.gov.cz/datov%C3%A1-sada?iri=https%3A%2F%2Fdata.gov.cz%2Fzdroj%2Fdatov%C3%A9-sady%2F44992785%2F8af9cdd4fdacbbbc328c9f4b325e97eb>
5. Brno - Sběrná střediska odpadů
https://data.brno.cz/datasets/6f68e2da78e14d99bcc3acde3c0ecb96_0/explorate?location=49.227542%2C16.542503%2C11.79

