

# Posudek oponenta diplomové práce

Autor/autorka práce: **František Kaiser**

Název práce: **SPARQL Endpoint a interaktivní vizualizace dat**

Obsah práce

Autor v textu práce představuje koncept propojených dat a SPARQL dotazování. Následně se zabývá možnostmi dashboardových vizualizací dat a představuje tři současné nástroje – Power BI, Looker Studio, Tableau. Jelikož žádný z nich nemá přirozenou podporu pro SPARQL dotazování, vytváří autor v praktické části vlastní řešení. Řešení je rozděleno na návrh získávání dat, zpracování a vizualizace, implementaci klient-server aplikace, testování a validaci. Na závěr je srovnáváno vlastní řešení s ostatními nástroji a jsou vymezeny možná další rozšíření aplikace.

Kvalita řešení a dosažených výsledků

Vytvořená aplikace je funkční a lze použít k vytvoření základních dashboardů vykreslujících data získána z libovolného veřejného SPARQL endpointu. Součástí aplikace je logika pro detekci datových typů zpracovávaných dat a obohacení dat na základě dostupných ontologií. Tím se aplikace liší od ostatních dashboardovacích aplikací bez podpory RDF/SPARQL.

Je otázkou, jak efektivně bude aplikace pracovat s dotazy vracejícími velké množství záznamů. Dle výsledků testů se pohybuje doba výpočtu statistik a metadat kolem 10 sekund již při 5K vrácených záznamů. V extrémním případě trvá výpočet statistik 198 sekund pro jeden sloupec při 129K záznamech.

Uživatelské rozhraní je sestaveno smysluplně a zjevně je inspirováno prostředím Power BI. Provedení grafů by bylo možné zlepšit – např. odstraněním os u výšečových grafů, či umožnění horizontálního vyskládání sloupcového grafu pro lepší vykreslení popisků hlavní osy.

Kladně hodnotím důkladný soupis specifikace požadavků (příloha B), z které bude možné vycházet při případném rozšiřování aplikace.

Formální úroveň

Úroveň je celkově dobrá. V kapitole 5 bylo v několika případech obtížné zorientovat se, o kterém nástroji autor mluví.

Práce s literaturou

Použitá literatura je aktuální a vhodná pro řešené téma.

Splnění zadání

Zadání bylo splněno v plném rozsahu.

#### Dotazy k práci

- Je pro uživatele omezující, pokud trvá výpočet statistik dlouhou dobu, resp. je možné s daty v omezené míře pracovat i bez spočtených statistik?
- Jak obtížné by bylo v řešení přidat podporu D3.js knihovny nebo Python knihoven (např. Seaborn) pro tvorbu vizualizací, které lze využívat v konkurenčním Power BI?

Navrhuji hodnocení známkou **výborně** a práci doporučuji k obhajobě.

V Plzni 3.6.2024

Ing. Martin Kryl