Personalised real estate search application using semantic web technologies

Ing. Tomáš Dvořák, Supervisor: Ing. Milan Dojčinovski, Ph.D.

CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE

MOTIVATION 🤔

Most real estate portals **lack the ability to filter their properties based on civic amenities or transport availability** and thus force users to find the right property by their own research leading to user dissatisfaction.

GOALS 💬

Create a real estate portal that aggregates data from leading real estate portals and enriches them with **open-linked semantic data** to provide its own advanced search capabilities, such as filtering by the nearest surroundings (parks, civic amenities, public transport) and building the internal system knowledge graph.

INFORMATION EXTRACTION 💡

The platform scrapes data from the leading real estate portals, validates, and converts them into a **single semantic data structure (RDF) stored in the Virtuoso Triple Store**. Those data are then periodically revalidated and related changes are persisted to create the **history snapshots**. The scraping/validation/persist flow pipeline works in a plug-and-play fashion and internally uses **Redis queues**.

APPLICATION 💡

The *Open Street Map* project is used for displaying the map on the client side, whereas the data about map objects are converted to **RDF** and used as the core geographic data source. Those objects are then linked or inferred from the **RUIÁN** *and* **Wikidata knowledge graph**. The resulting data are combined to give the user as much information as possible, such as the opening hours of the nearest restaurant or the number of flats in the given building. Advanced users can access our SPARQL endpoint and run more powerful SPARQL queries across the shared interlinked knowledge graph.

PREVIEW 🏡
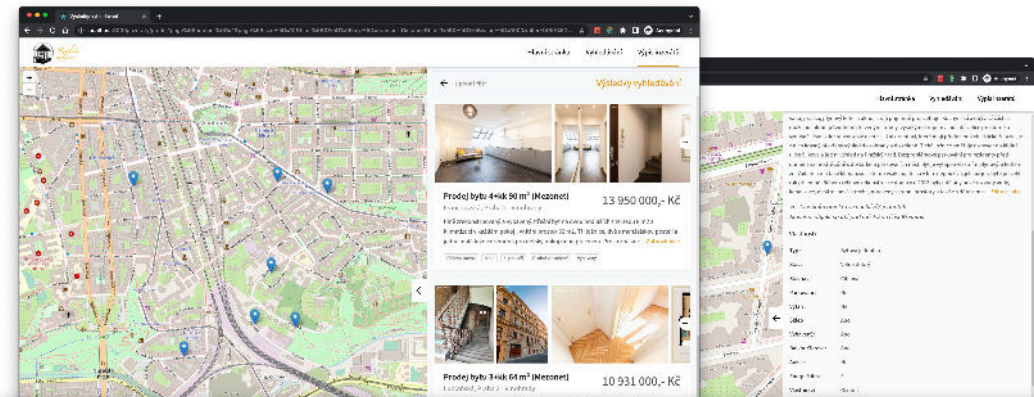
OPEN SOURCE CONTRIBUTIONS 🎗️

Published Open Source library for creating direct connection pool to the Virtuoso Triple Store (**https://bit.ly/3B6R7Mz**).

Reported eight issues to **Virtuoso Triple Store** database (**https://bit.ly/3d4qTCr**).