



CZECH
TECHNICAL
UNIVERSITY
IN
PRAGUE

**FACULTY OF
ELECTRICAL ENGINEERING**

OSW - Road accidents within the weather context in Prague - CP 1

Version 1.3, 17-11-2017

Table of Contents

Motivation	1
Introduction	1
Data source overview	1
Integrated Traffic Information System for the Czech Republic (DS1)	2
Data set description	2
Data pipeline description	2
CHMI (DS2 and DS3)	3
Data set description	3
Data pipeline description	4
JDVM CZ (DS4)	4
Data set description	4
Data pipeline description	5
UML diagrams of the data set schemas	6
Extension Discussion - Openweathermap	6
Possible Ontology	6

Motivation



A lots of traffic accidents are caused by weather (poor visibility, strong wind or rainfall, humidity on the road etc.). An integration of a road info (accidents) and meteorological sensors/stations could help to see connections between accidents and weather. This information could be used in an road safety application that will be able to notificate drivers against imminent danger due to weather.

Introduction

This report describes semi-automated data pipeline, based on the data sources from checkpoint 0. For this purpose was created C#. NET solution (OSW.DataParser). The solution creates data-sources like a input to the OntoRefine application. The OntoRefine helps to transform data-sources into RDF (.ttl). The datasets are limited to the period from the beginning of December 2016 to its end. The UML diagrams of the data set schemas are provide.

Data source overview

Table 1. Data source overview

Data source/set #	Provider	Web sides	Format	Description
DS1	Ředitelství silnic a dálnic ČR	http://www.dopravniinfo.cz/	xml	Integrated Traffic Information System for the Czech Republic
DS2	Český hydrometeorologický ústav	http://portal.chmi.cz	xls	Prague Ruzyně station. A dataset of daily hydrometeorological data from 1961 to 2016. Updates are done once a year.
DS3	Český hydrometeorologický ústav	http://portal.chmi.cz	xls	Prague Libuš station. A dataset of daily hydrometeorological data from 1961 to 2016. Updates are done once a year.
DS4	JDVM CZ	http://pcr.jdvm.cz	api	Statistical view of road accidents

Integrated Traffic Information System for the Czech Republic (DS1)

The machine readable XML dataset (appx. 300MB) about the road traffic from the B4M33OSW website [1: https://cw.fel.cvut.cz/wiki/courses/osw/semestral_work] with events about road accidents in the Czech Republic in several months.

Data set description

The data are freely accessible (<https://portal.gov.cz/portal/ostatni/volny-pristup-k-ds.html>). Only a subset of data (a month) will be selected due to the data size. The data schema contents the detailed description of the accident including the dates and values of the geographical location.

- The dataset structure description:
 - Id = original Id of record
 - MTIME_TGEN = čas gen
 - MTIME_TSTA = čas start
 - MTIME_TSTO = čas stop
 - MTIME_Format = formát času
 - MTXT_Language = typ jazyka popisu
 - MTXT_Text = popis situace
 - MEVT_OTXT = dodatek k situaci
 - MLOC_TXPL = typ situace
 - Latitude = zeměpisná šířka nehody
 - Longitude = zeměpisná délka nehody
 - CountryName = název země
 - RegionCode = kód regionu
 - RegionName = název regionu
 - TownShipCode = town ship code
 - TownShip = town ship
 - TownCode = town code
 - TownName = town name
 - TownDistrictCode = town district code
 - TownDistrictName = town district name

Data pipeline description

The Data pipeline steps:

1. Download **dopravni-info.zip** about the road traffic from the B4M33OSW website.
2. Run OSW.DataParser solution, select the date (yyyy/MM/dd) and Dopravni info data pipeline
3. In .\SemestralWork\Checkpoint1\Data\DopravniInfo\Final will be created a DopravniInfoNehodyPraha_YYYY_MM.json
4. Create new OntoRefine project
5. Upload the json file
6. Click to RDF button and fill out the RDF settings
7. Use SPARQL CONSTRUCT to create triples
8. Open the query in main SPARQL editor
9. Download data set as in the Turtle format



The Weather datasets has data only until the end of the December 2016. Due to this obstacle I can not collect actual data from DopravniInfo web page.

CHMI (DS2 and DS3)

Following weatherstation were selected Prague Ruzyně station and Libuš station. Datasets of daily hydrometeorological data from 1961 to 2016. Updates are done once a year. Datasets DS2 and DS3 have the same data schema and also other data aspects.

Data set description

Simple data table format (xls). The datasets content xls sheets. The first one describes meteorological station (e.g. latitude and longitude) and the next sheets contains daily hydrometeorological data about rainfall, the total amount of snow, humidity, air pressure, wind speed, sunshine and max/min/average temperature from 1961 to 2016, thus the datasets contain a large amount of data.

- The dataset structure description:
 - Id = StationIndicator_Year_Month_Day
 - Date = MM/dd/yyyy
 - StationIndicator = indikativ stanice
 - WMOIndicator = WMO indikativ
 - StationName = název stanice
 - StartObservation = začátek pozorování
 - EndObservation = konec pozorování
 - Latitude = zeměpisná šířka stanice
 - Longitude = zeměpisná délka stanice
 - Altitude = nadmořská výška stanice

- Basin = povodí
- StationType = typ stanice
- AvgAirPressure = průměrná tlak vzduchu
- AvgHumidity = průměrná vlhkost vzduchu
- AvgTemp = průměrná teplota
- AvgWindSpeed = průměrná rychlost větru
- DailyRainfall = denní srážky
- MaxTemp = maximální teplota
- MinTemp = minimální teplota
- Sunshine = slunečnost
- TotalAmountSnow = celkové množství sněhu

Data pipeline description

The Data pipeline steps:

1. Download .xls file with daily hydrometeorological data from <http://portal.chmi.cz> to .\SemestralWork\Checkpoint1\Data\CHMI\Input.
2. Run OSW.DataParser solution, select the date (yyyy/MM/dd) and CHMI data pipeline
3. In .\SemestralWork\Checkpoint1\Data\CHMI\Final will be crated a WeatherInfo_YYYY_MM.tsv
4. Create new OntoRefine project
5. Upload the tsv file
6. Click to RDF button and fill out the RDF settings
7. Use SPARQL CONSTRUCT to create triples
8. Open the query in main SPARQL editor
9. Download data set as in the Turtle format

JDVM CZ (DS4)

Statistical view of road accidents.

Data set description

- The dataset structure description:
 - AccidenID = ID nehody
 - IdNumber = Identifikační číslo
 - Date = datum nehody
 - AccidenType = typ nehody

- CollisionType = Typ srážky
- BarrierType = Typ bariéry
- AccidentCharacter = charakter nehody
- Culpable = zavinění
- Alcohol = Alkohol
- MainCause = Hlavní důvod
- PersonsDead = počet usmrcených osob
- HeavyInjured = počet těžce zraněných osob
- SlightlyInjured = počet lehce zraněných osob
- StavVozovky = stav vozovky
- WeatherConditions = podmínky počasí
- Visibility = viditelnost
- SpecificLocations = specifikace
- DirectionalConditions = směrové poměry
- VehicleType = Typ vozidla
- VehicleCharacteristics = charakteristika vozidla
- FluidLeak = Únik tekutiny
- City = název města
- XLocation = Latitude
- YLocation = Longitude

Data pipeline description

The Data pipeline steps:

1. Run OSW.DataParser solution, select the date (yyyy/MM/dd) and CHMI data pipeline
2. In SQL DB will be created a AccidentItems table
3. Select all rows
4. Export csv file with Encoding UTF-8
5. Use <http://martin.hinner.info/geo/> to transform geolocations
6. Create new OntoRefine project
7. Upload the csv file
8. Click to RDF button and fill out the RDF settings
9. Use SPARQL CONSTRUCT to create triples
10. Open the query in main SPARQL editor
11. Download data set as in the Turtle format

UML diagrams of the data set schemas

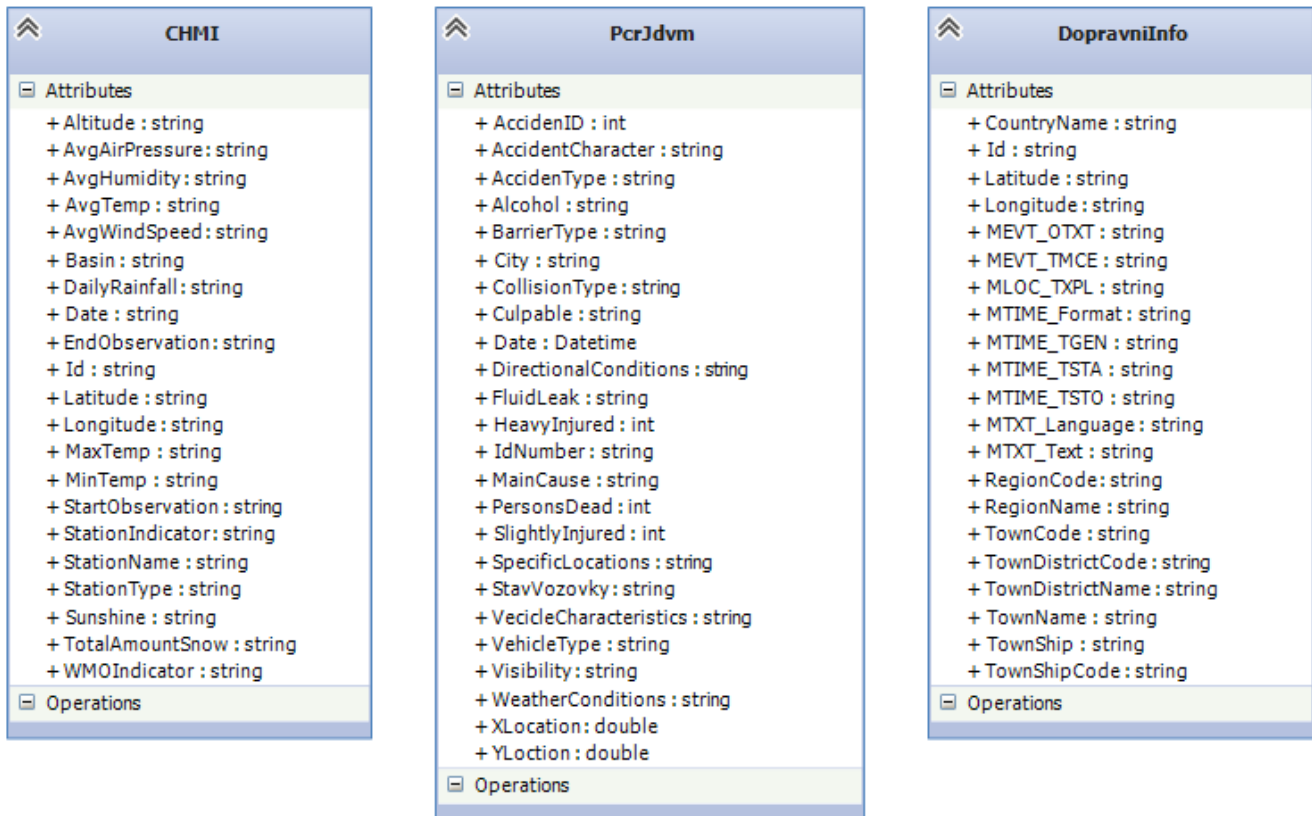


Figure 1. UML diagrams

Extension Discussion - Openweathermap



Very usefull API <http://openweathermap.org/history#geo> , but is not free.

Possible Ontology

<http://ontology.fiesta-iot.eu/ontologyDocs/m3-lite.owl#Weather>