



NI-SWE: Pandemie Covid-19 a Life Satisfaction

Petr Pondělík
pondepe1@fit.cvut.cz

01

DATASETY



POPIS DATASETŮ

1) CovidWorld

Denní statistiky k onemocnění Covid-19 ve světě (pozitivní, zemřelí, vyléčení, aktuálně pozitivní, hospitalizovaní, hospitalizovaní na jednotce intenzivní péče).

2) LifeSatisfaction

Hodnoty spokojenosti se životem pro jednotlivé země dle odpovědí obyvatel v jednotlivých letech.

3) CovidCZE

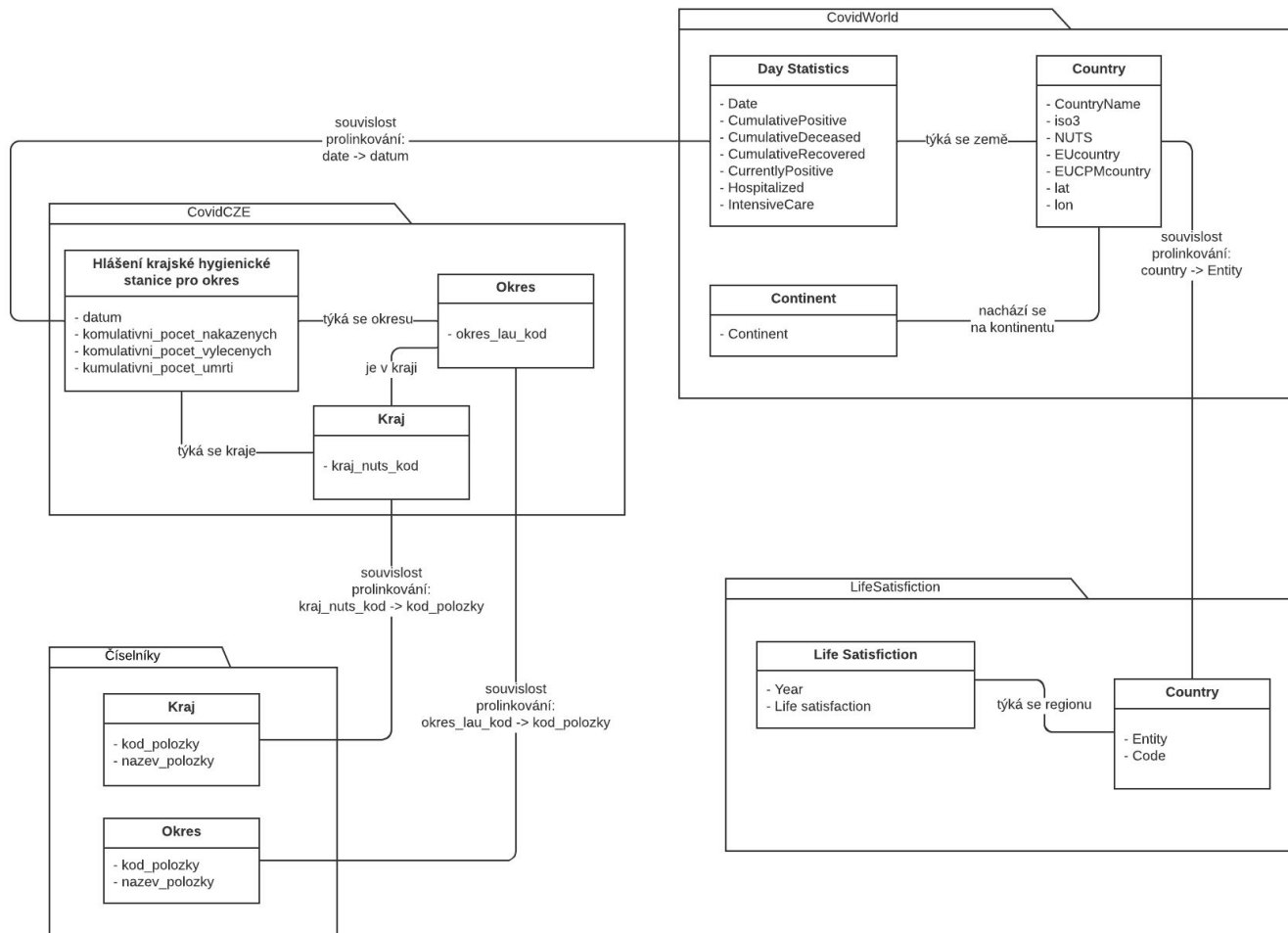
Denní statistiky pandemie COVID-19 pro ČR dle okresů.

4) Číselníky

Číselníky krajů a okresů ČR.

NI-SWE-Datasets-Class-Diagram

Petr Pondělík | October 19, 2021





02

POUŽITÉ NÁSTROJE

POUŽITÉ NÁSTROJE

1) TRIPLIFIKACE DO RDF: SPARQL CONSTRUCT + TARQL

```
$TARQL -e utf-8 --dedup 10000 "./life-satisfaction.sparql" > "${OUTPUT_DIR}/life-satisfaction.ttl"
```

- --dedup 10000: odstranění duplicitních trojic

```
CONSTRUCT {  
  ...  
  ?lifeSatisfactionObs a qb:Observation ;  
    qb:dataSet          ls-voc:dataset ;  
    voc:refCountry      ?country ;  
    voc:refYear          ?obsYear ;  
    ls-voc:happiness-cantril-ladder ?obsValueDec .  
  ...  
}  
FROM <file:../data/life-satisfaction.csv>  
WHERE {...}  
OFFSET 1
```

POUŽITÉ NÁSTROJE

2) SPARQL ENDPOINT: APACHE JENA FUSEKI

- Využití upravené docker image **stain/jena-fuseki**
- Navíc instalace utility procps
 - při ukončení původního kontejneru se nesprávně ukončovaly procesy
 - konflikty v přístupu k volume

3) LINKOVÁNÍ: SILK WORKBENCH

- Využití docker image **silkframework/silk-workbench**

4) VALIDACE DCAT-AP: itb.ec.europa.eu/shacl/dcat-ap/upload

5) VALIDACE JSON-LD: Google's Rich Results Test

6) VALIDACE SHACL: ZAZUKO SHACL PLAYGROUND

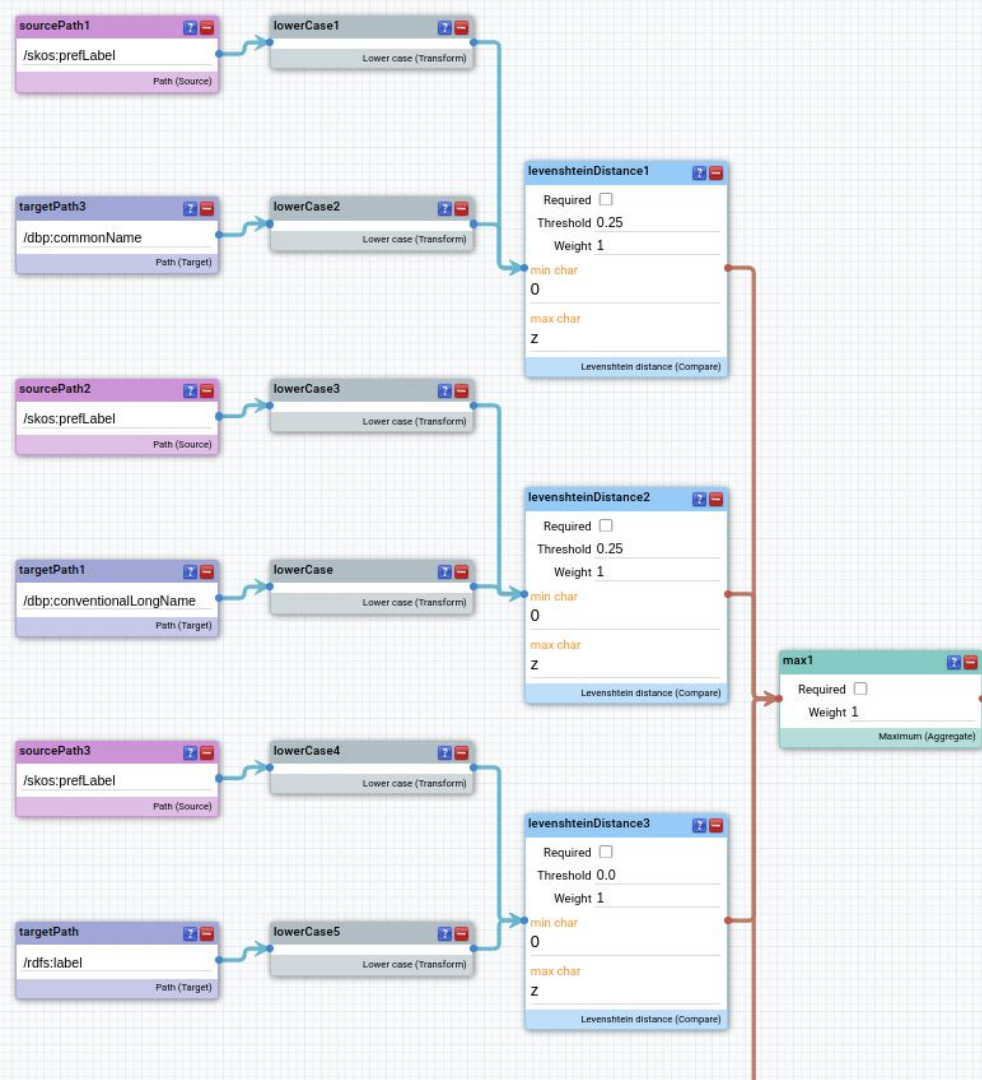


03

LINKOVÁNÍ

SILK WORKBENCH

- Vytvoření příslušných linking tasků mezi datasety
- Výsledky linkování ve výstupním datasetu
 - formát n-triples



04

SPARQL DOTAZY



SEPARATED CONSTRUCT QUERY

Sestav průměr (zaokrouhlený na 5 desetinných míst) hodnoty life satisfaction pro každou zemi.

```
CONSTRUCT {  
    ?country ls-voc:happiness-cantril-ladder-avg ?lsAvg  
}  
WHERE {  
    {  
        SELECT ?country (ROUND(AVG(?lsVal)*100000000)/100000 AS ?lsAvg)  
        WHERE {  
            ?obs rdf:type qb:Observation ;  
                voc:refCountry ?country ;  
                ls-voc:happiness-cantril-ladder ?lsVal .  
        }  
    }  
    GROUP BY ?country  
}
```

COMBINED QUERY

(Subquery: Výběr států s nejvyšším LS)

Výběr států EU, které měly 1.1.2020 méně pozitivních případů, než země s nejvyšším life satisfaction pro rok 2020.

```
...  
{  
  SELECT ?IsCountry  
  WHERE {  
    ?obs rdf:type qb:Observation ;  
          voc:refCountry          ?IsCountry ;  
          voc:refYear             ?IsYear ;  
          ls-voc:happiness-cantril-ladder ?IsVal .  
  
    FILTER ( year(?IsYear) = 2020 )  
  }  
  ORDER BY DESC(?IsVal)  
  LIMIT 1  
}  
...
```

COMBINED QUERY

(Subquery: Napojení na dataset CovidWorld)

Výběr států EU, které měly 1.1.2020 méně pozitivních případů, než země s nejvyšším life satisfaction pro rok 2020.

```
{  
  SELECT ?cwCountry ?cwCurrPos  
  WHERE {  
    ?cwCountry <http://www.w3.org/2002/07/owl#sameAs> ?lsCountry .  
  
    ?cwObs rdf:type qb:Observation ;  
    voc:refCountry ?cwCountry ;  
    voc:refDate "2021-01-01"^^xsd:date ;  
    cw-voc:currently-positive ?cwCurrPos .  
  
    {  
      # Subquery: Výběr státu s nejvyšším LS  
    }  
  }  
}
```

COMBINED QUERY

(Finální dotaz)

Výběr států EU, které měly 1.1.2020 méně pozitivních případů, než země s nejvyšším life satisfaction pro rok 2020.

```
SELECT ?country (?currPos AS ?Positive_2021_01_01)
WHERE {
    ?obs rdf:type qb:Observation ;
    voc:refCountry      ?country ;
    voc:refDate         "2021-01-01"^^xsd:date ;
    Cw-voc:currently-positive  ?currPos .

    ?country rdf:type skos:Concept ;
    cw-voc:eu-country ?isEuCountry .
    {
        # Subquery: Napojení na dataset CovidWorld
    }
}
FILTER ( ?currPos < ?cwCurrPos )
FILTER ( ?isEuCountry = true )
}
```

```

SELECT ?country (?currPos AS ?Positive_2021_01_01)
WHERE {
    ?obs rdf:type qb:Observation ;
        voc:refCountry      ?country ;
        voc:refDate         "2021-01-01"^^xsd:date ;
        cw-voc:currently-positive ?currPos .
    ?country rdf:type skos:Concept ;
        cw-voc:eu-country      ?isEuCountry .
    {
        SELECT ?cwCountry ?cwCurrPos
        WHERE {
            ?cwCountry <http://www.w3.org/2002/07/owl#sameAs> ?IsCountry .
            ?cwObs rdf:type qb:Observation ;
                voc:refCountry      ?cwCountry ;
                voc:refDate         "2021-01-01"^^xsd:date ;
                cw-voc:currently-positive ?cwCurrPos .
            {
                SELECT ?IsCountry
                WHERE {
                    ?obs rdf:type qb:Observation ;
                        voc:refCountry      ?IsCountry ;
                        voc:refYear         ?IsYear ;
                        ls-voc:happiness-cantril-ladder ?IsVal .
                    FILTER ( year(?IsYear) = 2020 )
                }
                ORDER BY DESC(?IsVal)
                LIMIT 1
            }
        }
    }
    FILTER ( ?currPos < ?cwCurrPos )
    FILTER ( ?isEuCountry = true )
}

```

COMBINED FEDERATED QUERY

(Subquery: Stát s nejvyšším HDI)

Výběr celkového počtu zemřelých pro stát EU s nejnižším a nejvyšším HDI.

Výběr pouze ze států s počtem obyvatel nad 5 milionů.

```
SELECT ?dbpCountry ?dbpHDevInx ?totalDec
WHERE {
  ?cwCountry <http://www.w3.org/2002/07/owl#sameAs> ?dbpCountry .
  ?obs rdf:type qb:Observation ;
    voc:refCountry          ?cwCountry ;
    voc:refDate              "2021-01-01"^^xsd:date ;
    cw-voc:cumulative-deceased ?totalDec .
  SERVICE <http://dbpedia.org/sparql> {
    ?dbpCountry rdf:type dbo:Country, yago:WikicatCountriesInEurope ;
      dbo:humanDevelopmentIndex ?dbpHDevInx ;
      dbo:populationTotal       ?dbpPopulation .
  }
  FILTER ( ?dbpPopulation > 5000000 )
}
ORDER BY DESC(?dbpHDevInx)
LIMIT 1
```


COMBINED FEDERATED QUERY

(Finální dotaz)

Výběr celkového počtu zemřelých pro stát EU s nejnižším a nejvyšším HDI.

Výběr pouze ze států s počtem obyvatel nad 5 milionů.

```
SELECT *  
WHERE {  
  {  
    # Subquery: Stát s nejvyšším HDI  
  }  
  UNION  
  {  
    # Subquery: Stát s nejnižším HDI  
  }  
}
```

```
SELECT *  
WHERE {
```

```
{
```

```
SELECT ?dbpCountry ?dbpHDevInx ?totalDec  
WHERE {  
    ?cwCountry <http://www.w3.org/2002/07/owl#sameAs> ?dbpCountry .  
    ?obs rdf:type qb:Observation ;  
        voc:refCountry          ?cwCountry ;  
        voc:refDate              "2021-01-01"^^xsd:date ;  
        cw-voc:cumulative-deceased ?totalDec .  
    SERVICE <http://dbpedia.org/sparql> {  
        ?dbpCountry rdf:type dbo:Country, yago:WikicatCountriesInEurope ;  
            dbo:humanDevelopmentIndex ?dbpHDevInx ;  
            dbo:populationTotal       ?dbpPopulation .  
    }  
    FILTER ( ?dbpPopulation > 5000000 )  
}  
ORDER BY DESC(?dbpHDevInx)  
LIMIT 1
```

```
}
```

```
UNION
```

```
{
```

```
SELECT ?dbpCountry ?dbpHDevInx ?totalDec  
WHERE {  
    ?cwCountry <http://www.w3.org/2002/07/owl#sameAs> ?dbpCountry .  
    ?obs rdf:type qb:Observation ;  
        voc:refCountry          ?cwCountry ;  
        voc:refDate              "2021-01-01"^^xsd:date ;  
        Cw-voc:cumulative-deceased ?totalDec .  
    SERVICE <http://dbpedia.org/sparql> {  
        ?dbpCountry rdf:type dbo:Country, yago:WikicatCountriesInEurope ;  
            dbo:humanDevelopmentIndex ?dbpHDevInx ;  
            dbo:populationTotal       ?dbpPopulation .  
    }  
    FILTER ( ?dbpPopulation > 5000000 )  
}  
ORDER BY ASC(?dbpHDevInx)  
LIMIT 1
```

```
}
```

```
}
```

The background features a complex network of thin grey lines connecting various-sized dark grey circular nodes. These nodes are scattered across the frame, with some forming dense clusters and others standing alone. The overall effect is a technical, digital aesthetic. A thin vertical line is positioned to the left of the main text.

05

EMBEDDED DATA

HTML+RDFa

```
<main about="http://www.example.org/covid-world/countries/DEU/2021-02-12" typeof="qb:Observation">
  <a href="http://www.example.org/covid-world/countries/DEU">Back to country</a>
  <h1>COVID infection data (DEU, 2021-02-12)</h1>
  <dl>
    <dt>Dataset</dt>
    <dd>
      <a rel="qb:dataSet" href="http://www.example.org/covid-world/vocabulary/dataset">
        Dataset description
      </a>
    </dd>
    <dt>Country</dt>
    <dd>
      <a rel="voc:refCountry" href="http://www.example.org/covid-world/countries/DEU">DEU</a>
    </dd>
    <dt>Date</dt>
    <dd property="voc:refDate" datatype="xsd:date">2021-02-12</dd>
    <dt>Cumulative positive</dt>
    <dd property="cw-voc:cumulative-positive" datatype="xsd:integer">2320093</dd>
    <dt>Cumulative deceased</dt>
    <dd property="cw-voc:cumulative-deceased" datatype="xsd:integer">64191</dd>
    <dt>Cumulative recovered</dt>
    <dd property="cw-voc:cumulative-recovered" datatype="xsd:integer">2112000</dd>
    <dt>Currently positive</dt>
    <dd property="cw-voc:currently-positive" datatype="xsd:integer">143902</dd>
  </dl>
</main>
```

HTML+RDFa

[Back to country](#)

COVID infection data (DEU, 2021-02-12)

Dataset

[Dataset description](#)

Country

[DEU](#)

Date

2021-02-12

Cumulative positive

2320093

Cumulative deceased

64191

Cumulative recovered

2112000

Currently positive

143902

JSON-LD (Schema.org)

```
{
  "@context": "http://schema.org/",
  "@type": "Dataset",
  "name": "Self-reported Life Satisfaction",
  "description": "Average of survey responses to the 'Cantril Ladder' question in the Gallup World Poll. The survey question asks respondents to think of a ladder, with the best possible life for them being a 10, and the worst possible life being a 0.",
  "url": "http://www.example.org/source/dataset/life-satisfaction",
  "keywords": [
    "Survey",
    "Life",
    "Satisfaction"
  ],
  "license": "https://creativecommons.org/publicdomain/zero/1.0/",
  "creator": {
    "@type": "Organization",
    "name": "World Happiness Report"
  },
  "distribution": [
    {
      "@type": "DataDownload",
      "contentUrl": "http://www.example.org/files/life-satisfaction/life-satisfaction.ttl",
      "encodingFormat": "text/turtle"
    }
  ]
}
```



06

SHACL SHAPES

Shape-based constraint

```
:refOkresShape a sh:NodeShape ;
  sh:property [
    sh:path dbp:nuts ;
    sh:datatype xsd:string ;
  ] ;
  sh:property [
    sh:path skos:broader ;
    sh:nodeKind sh:IRI ;
  ] ;
  sh:property [
    sh:path skos:inScheme ;
    sh:nodeKind sh:IRI ;
  ] .

:observationShape a sh:NodeShape ;
  sh:targetClass qb:Observation ;
  sh:property [
    sh:path voc:refOkres ;
    sh:node :refOkresShape ;
    sh:message "Neplatný shape položky voc:refOkres."@cs ;
    sh:description "Validace shapu okresu pozorování."@cs
  ] ;
```


Regex constraint pro iso3166

(v rámci datasetu - OWID = Our World In Data ???)

```
:iso3166codeFormatShape a sh:PropertyShape ;
  sh:targetObjectsOf      dbp:iso3166code ;
  sh:pattern               "^([A-z]{3}|OWID_[A-z]{3})$" ;
  sh:flags                 "i" ;
  sh:severity              sh:Warning ;
  sh:message               "Invalid format of iso3166code subject."@en ;
  sh:description           "Format validation for subjects of dbp:iso3166code. iso3166code should have XXX or OWID_XXX
format, where X is english alphabet symbol."@en .
```

The background of the slide features a complex, abstract geometric pattern. It consists of numerous thin, light gray lines that intersect to form a network of triangles and other polygons. At the vertices of these shapes, there are small, dark gray circular dots. The overall effect is a modern, technical, and interconnected visual theme.

Děkuji za pozornost

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.