



---

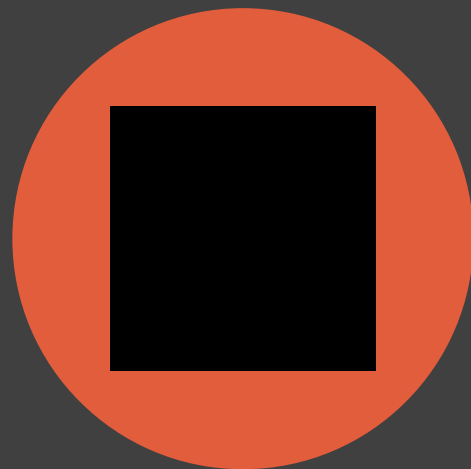
# PLIN021 SÉMANTICKÁ ANALÝZA V PRAXI

ZUZANA NEVĚŘILOVÁ

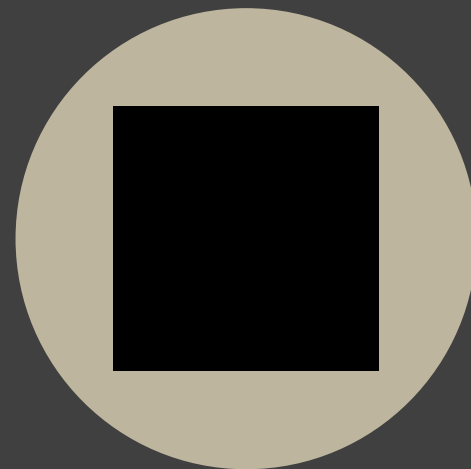
2020/21

---

# ZDROJE ZNALOSTÍ PRO ZPRACOVÁNÍ PŘIROZENÉHO JAZYKA

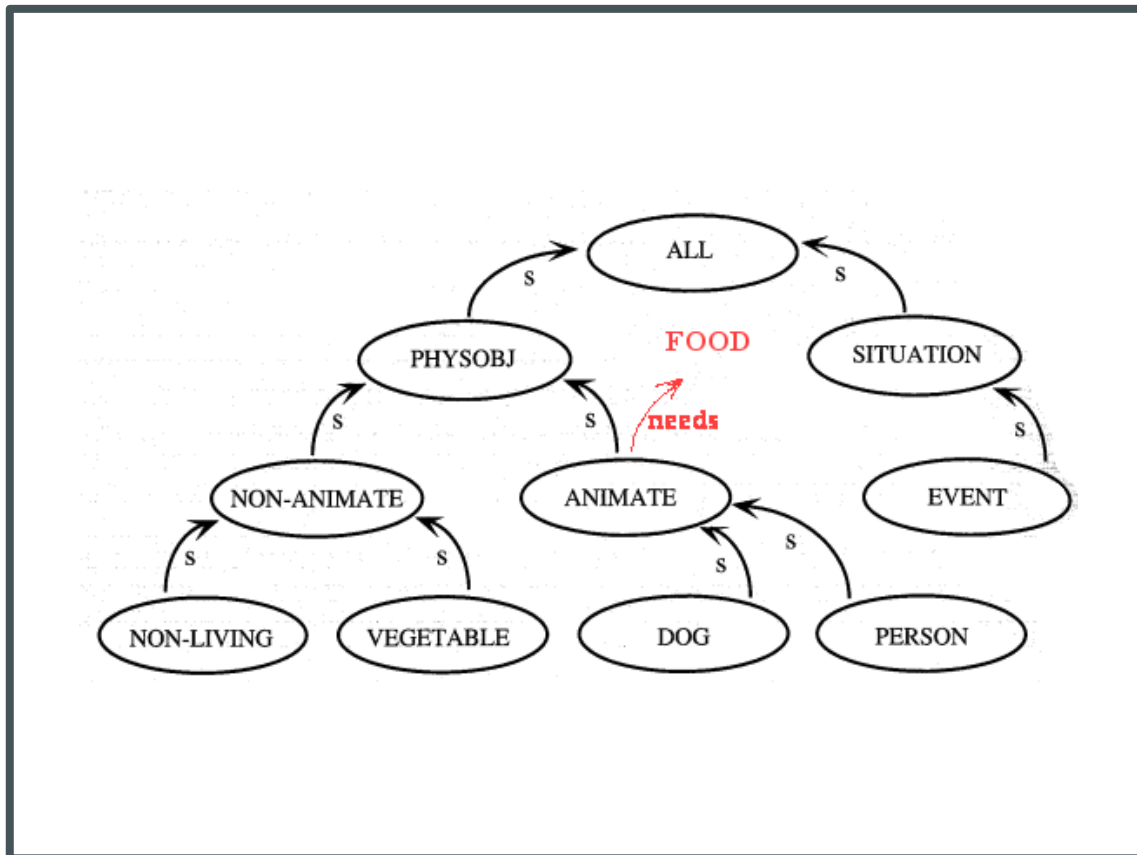


ZNALOST SVĚTA



ZNALOST JAZYKA

# SÉMANTICKÉ SÍTĚ



- Různé typy znalostí
  - Znalost jazyka
  - Znalost „světa“
    - Encyklopedická znalost
    - Common sense
- Pro člověka poměrně srozumitelné
- Pro počítačové programy dobře zpracovatelné
  - API (application programming interface)

# APLIKAČNÍ PROGRAMOVÉ ROZHHRANÍ

## APPLICATION PROGRAMMING INTERFACE (API)

- Způsob, jak spolu komunikují dva počítačové programy
- Jazyk pro komunikaci:
  - XML, JSON, RDF(S)/OWL
- Jazyky jsou standardizované, význam jednotlivých relací nikoli.

```
"query": {  
  "pages": {  
    "22760983": {  
      "pageid": 22760983,  
      "ns": 0,  
      "title": "Linguistics",  
      "revisions": [  
        {  
          "contentformat": "text/x-wiki",  
          "contentmodel": "wikitext",  
          "*": "{{short description|Study of human language}}"        }  
      ]  
    }  
  }  
}
```

```
<query>  
  <pages>  
    <page _idx="22760983" pageid="22760983" ns="0" title="Linguistics">  
      <revisions>  
        <rev contentformat="text/x-wiki" contentmodel="wikitext"  
xml:space="preserve">{{short description|Study of human language}}      </rev>  
    </page>  
  </pages>  
</query>
```

# RDF(S)/OWL: JAZYKY SÉMANTICKÉHO WEBU

- Sémantický web: webové stránky nejsou jen text, ale obsahují sémantické anotace, tudíž jsou zpracovatelné počítačovými programy
- Jazyky pro sémantické anotace: RDF, RDFS, OWL a varianty
- Tyto jazyky popisují propozice jako trojice tvrzení (subjekt, predikát, obj
- RDF(S) obsahuje relace podtřída a umožňuje typovou kontrolu
- OWL umožňuje definovat typy relací (symetrické, tranzitivní)

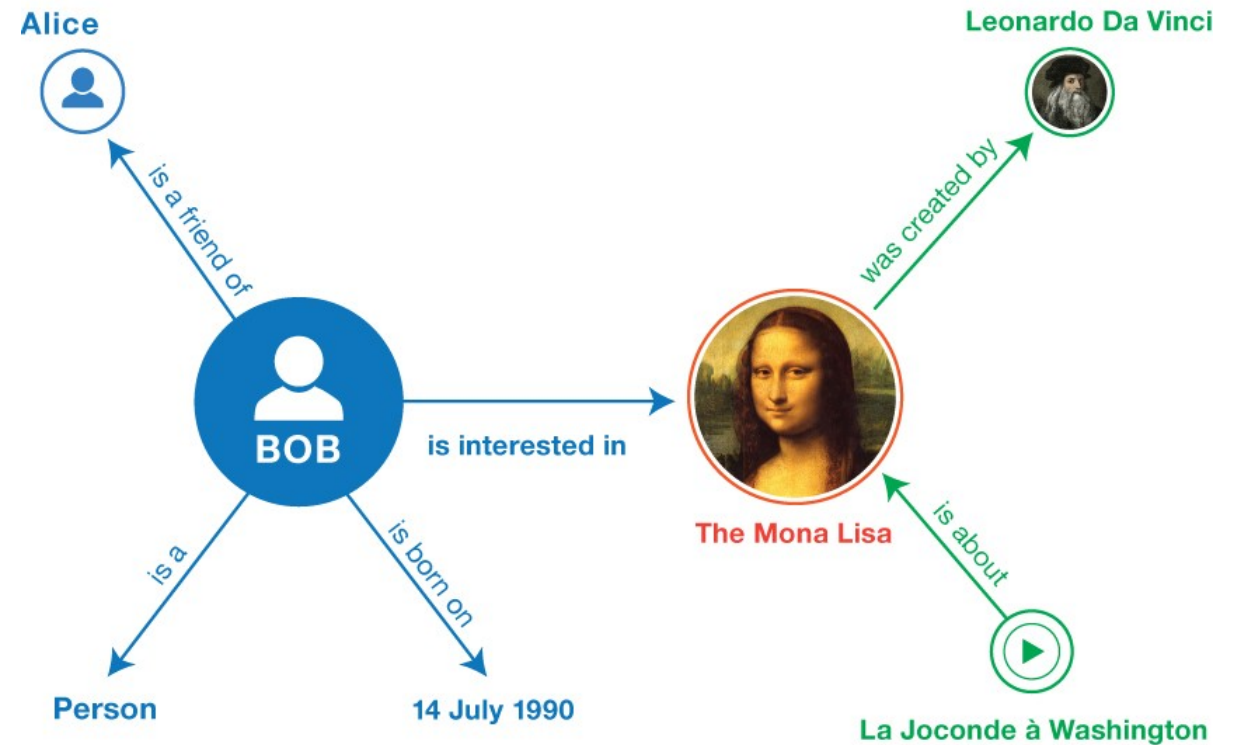


Pomocí RDF(S)/OWL definujeme odvozování faktů.

# RDF: PŘÍKLAD

- Bob is interested in The Mona Lisa.
- Bob is born on 14 July 1990.
- Bob is a Person.
- Bob is a friend of Alice.
- La Joconde a Washington is about The Mona Lisa.
- The Mona Lisa was created by Leonardo Da Vinci.

<https://www.w3.org/TR/rdf11-primer/>



# RDF(S)/OWL: PŘÍKLAD INFERENCE

- Symetrie: A is sibling of B  $\rightarrow$  B is sibling of A.
- Tranzitivita: A has part B, B has part C  $\rightarrow$  A has part C.
- Inverzní relace: A has part B  $\rightarrow$  B is part of A.



# SPARQL: DOTAZOVACÍ JAZYK PRO RDF(S)/OWL

```
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX : <http://dbpedia.org/resource/>
```

```
SELECT ?name ?birth ?death ?person
WHERE { ?person dbo:birthPlace :Berlin .
        ?person dbo:birthDate ?birth .
        ?person foaf:name ?name .
        ?person dbo:deathDate ?death .
FILTER (?birth < "1900-01-01"^^xsd:date) . }
ORDER BY ?name
```

- Najdi všechny osoby (jméno, datum narození, úmrtí) takové, že se narodily v Berlíně před rokem 1900.
- Hledáme v dbPedi
- Využíváme ale schémata i jiných ontologií
  - FOAF – friend of a friend (relace mezi lidmi)
  - XSD – XML schema (datové typy)

<http://sites.linkeddata.center/help/devop/examples/sparql-examples>



# OKÉNKO DO ROKU 2069

Neměl v pokoji standardní infromatický panel, musel dolů do vestibulu, chtěl-li získat údaje, které potřeboval. Jako naschvál byla kabinka obsazena nějakou starší dámou, které trvalo pátrání v informačních zdrojích tak dlouho, že Morgan už málem na dveře budky zabouchal. Ale nakonec se loudalka přece jen s nějakým omluvným zamumláním z kabiny vybatolila a Morgan stál tváří v tvář nashromážděným vědomostem a tvořivým silám celého lidstva.

Již za svých studentských let zvítězil v několika takových pátračských šampionátech, kdy o závod s časem lovil a vyhrabával nejzasutější informace požadované testy; formulovali je až geniálně sadističtí sudí. („**Jaké bylo množství srážek v hlavním městě nejmenšího jednonárodního státu světa v den, kdy v kolejním baseballovém mistrovství byl zaznamenán druhý největší počet celých oběhů?**“ – tenhle mu snad nejvíc utkvěl v paměti i v srdci.)

Arthur C. Clarke: RAJSKÉ FONTÁNY, 1979

# LITERATURA

- Lee Feigenbaum: **SPARQL By Example: A Tutorial**. VP Technology & Standards, Cambridge Semantics. 2009 <https://www.w3.org/2009/Talks/0615-qbe/>
- W3C: **Web Ontology Language (OWL)**. W3C Semantic Web. 2012. <https://www.w3.org/OWL/>
- OpenLink Software: **Virtuoso SPARQL Query Editor**. 2020. <http://dbpedia.org/sparql>