

PLIN037 Sémantika a počítače

Zuzana Nevěřilová
xpopelk@fi.muni.cz

Centrum zpracování přirozeného jazyka, B203
Fakulta informatiky, Masarykova univerzita

11. dubna 2019

Parafráze

Reprezentace znalostí

Jazyky sémantického webu

OWL: Inference

Dotazy

Aplikace rodiny RDF

Parafráze

Čím jsou zajímavé:

- schopnost parafrázovat vyžaduje schopnost porozumět
- aplikace ve strojovém překladu
- aplikace ve výuce jazyka
- aplikace v detekci obsahového překryvu
- aplikace v detekci plagiátů

Klasifikace parafrází

- obtížný úkol, se kterým se každý vyrovná po svém (\rightarrow nízká mezianotátorská shoda)
- anotační manuál (který buď nezachytí všechny případy, nebo ho nikdo nebude číst)
- řešení neshody (třetí anotátor)
- řešení náhodné shody? (výpočet Cohen κ nebo Fleiss κ)

Klasifikace parafrází

- obtížný úkol, se kterým se každý vyrovná po svém (\rightarrow nízká mezianotátorská shoda)
- anotační manuál (který buď nezachytí všechny případy, nebo ho nikdo nebude číst)
- řešení neshody (třetí anotátor)
- řešení náhodné shody? (výpočet Cohen κ nebo Fleiss κ)

lepší je formulovat lépe úkol pro anotátory, nejlépe jako klasifikační úlohu pro velmi málo tříd

Reprezentace znalostí

K sémantické analýze potřebujeme znalost jazyka a znalost světa.

Znalost světa je dostupná ve znalostních bázích:

- lexikální sítě (wordnety)
- valenční slovníky
- doménové znalosti (doménové ontologie)
- encyklopedické znalosti (DBPedia)
- common-sense (ConceptNet, CyC)
- metadata (DublinCore)

Jazyky sémantického webu

Delší dobu trvající snahy o standardizaci přinesly několik různých způsobů zápisu faktů a odvozovacích pravidel do znalostních bází, např. KIF:

```
(documentation AudiA6 EnglishLanguage "The A6 model of
cars made by Audi beginning in 1995")
(subclass AudiA6 AudiAutomobile)
(subclass AudiAutomobile Automobile)
(termFormat EnglishLanguage AudiA6 "Audi A6")
(termFormat EnglishLanguage AudiAutomobile "Audi")
```

Jazyky sémantického webu (RDF, OWL)

- ukládají data ve formě trojic
- mají několik standardizovaných formátů (NT, XML, Turtle).

Sémantický web

Sémantický web (také Web 3.0) – web, kterému budou počítačové programy „rozumět“.¹

- aktivita W3C (autorem myšlenky je sir Tim Berners-Lee)
- standardizovaný jazyk pro popis sémantických vztahů: RDF



¹ <http://en.wikipedia.org/wiki/File:Tim_Berners-Lee-Knight-crop.jpg>

Web a sémantický web

```

<h1>CD Nejlepší kapela: The Worst Of</h1>
<p>Nejlepší kapela je vítězem soutěže Evropská unie má
    talent. Na svém novém CD The Worst Of hrají ty
    nejlepší pecky.</p>

<h1>CD <item id="id1" type="band-name">Nejlepší kapela
    </item>: <item id="id2" type="artwork-name">The
    Worst Of</item><relation id="rel1" type="author"
    subj="id1" obj="id2"/></h1>
<p><item id="id3" ref="id1" type="band-name">Nejlepší
    kapela</item> je <relation id="rel2" type="winner"
    subj="id3" obj="id4">vítězem</relation> soutěže <
    item id="id4" type="contest-name">Evropská unie má
    talent</item>. Na svém novém CD <item id="id5"
    rel="id2" type="artwork-name">The Worst Of</item>
    hrají ty <item id="id6" type="mark-positive">
    nejlepší</item> pecky.</p>

```

Jazyky sémantického webu: rodina RDF

RDF = Resource Description Framework

RDFS = RDF Scheme

OWL = Web Ontology Language

⋮

Jazyky sémantického webu: rodina RDF

trojice: subjekt–predikát–objekt

Tim Berners-Lee – vynalezl – web

Rodina RDF: několik způsobů zápisu

Tim Berners-Lee – vynalezl – web

N3 (Notation3):

```
@PREFIX dc: <http://purl.org/dc/elements/1.1/>.
<http://en.wikipedia.org/wiki/Tim_Berners-Lee>
  dc:creator <http://en.wikipedia.org/wiki/
    World_Wide_Web>
```

XML:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
    ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://en.wikipedia.org/
  wiki/Tim_Berners-Lee">
  <dc:creator rdf:resource="http://en.wikipedia.org/
    wiki/World_Wide_Web"/>
</rdf:Description>
</rdf:RDF>
```

Rodina RDF: validace zápisu

```
<http://www.w3.org/RDF/Validator/>
```

Rodina RDF: složitější příklad

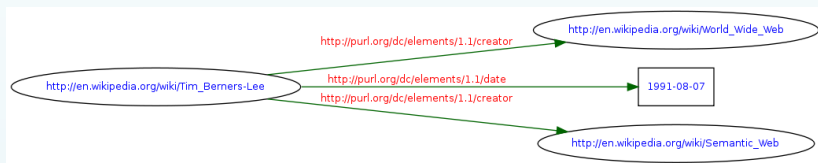
```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
    ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://en.wikipedia.org/
  wiki/Tim_Berners-Lee">
  <dc:creator rdf:resource="http://en.wikipedia.org/
    wiki/World_Wide_Web"/>
  <dc:date>1991-08-07</dc:date>
  <dc:creator rdf:resource="http://en.wikipedia.org/
    wiki/Semantic_Web"/>
</rdf:Description>
</rdf:RDF>
```

Rodina RDF: složitější příklad

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
    ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://en.wikipedia.org/
  wiki/Tim_Berners-Lee">
  <dc:creator rdf:resource="http://en.wikipedia.org/
    wiki/World_Wide_Web"/>
  <dc:date>1991-08-07</dc:date>
  <dc:creator rdf:resource="http://en.wikipedia.org/
    wiki/Semantic_Web"/>
</rdf:Description>
</rdf:RDF>

```



RDF: není trojice málo?

Není. Díky reifikaci (zpředmětování, reification).

subjekt	predikát	objekt
Tim Berners-Lee	dc:creator	WWW

RDF: není trojice málo?

Není. Díky reifikaci (zpředmětování, reification).

subjekt	predikát	objekt	
Tim Berners-Lee	dc:creator	WWW	⇐ 1991-08-07

RDF: není trojice málo?

Není. Díky reifikaci (zpředmětování, reification).

subjekt	predikát	objekt
tvrzení1	rdf:type	rdf:Statement
tvrzení1	rdf:subject	Tim Berners-Lee
tvrzení1	rdf:predicate	dc:creator
tvrzení1	rdf:object	WWW

RDF: není trojice málo?

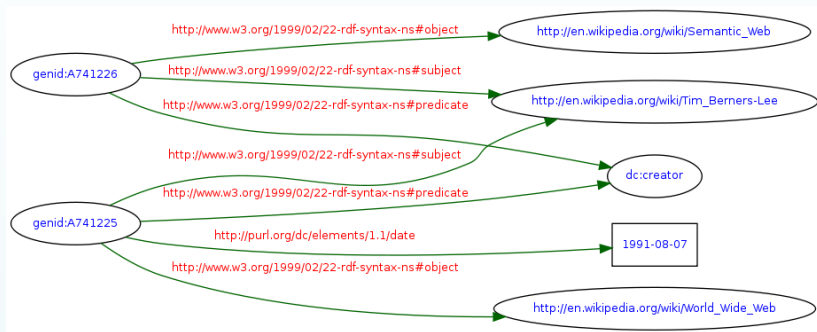
Není. Díky reifikaci (zpředmětování, reification).

subjekt	predikát	objekt
tvrzení1	rdf:type	rdf:Statement
tvrzení1	rdf:subject	Tim Berners-Lee
tvrzení1	rdf:predicate	dc:creator
tvrzení1	rdf:object	WWW
tvrzení1	dc:date	1991-08-07

RDF: reifikace

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
    ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description>
  <rdf:subject rdf:resource="http://en.wikipedia.org/
    wiki/Tim_Berners-Lee"/>
  <rdf:predicate rdf:resource="dc:creator"/>
  <rdf:object rdf:resource="http://en.wikipedia.org/
    wiki/World_Wide_Web"/>
  <dc:date>1991-08-07</dc:date>
</rdf:Description>
<rdf:Description>
  <rdf:subject rdf:resource="http://en.wikipedia.org/
    wiki/Tim_Berners-Lee"/>
  <rdf:predicate rdf:resource="dc:creator"/>
  <rdf:object rdf:resource="http://en.wikipedia.org/
    wiki/Semantic_Web"/>
</rdf:Description>
</rdf:RDF>
```

RDF: reifikace



Sémantický web?

“This simple idea . . . remains largely unrealized.”

[?]

OWL: Inference

Tvrzení o vlastnostech:

- `rdf:property` je symetrická
- `rdf:property` je antisymetrická
- `rdf:property` je tranzitivní
- `rdf:property` je ekvivalence
- `rdf:property` má definiční obor (domain)
- `rdf:property` má obor hodnot (range)

OWL: Inference

pomocí SubClassOf

```
Class: bus_driver
  EquivalentTo:
    person
    that drives some bus
```

```
Class: driver
  EquivalentTo:
    person
    that drives some vehicle
```

```
Class: bus
  SubClassOf:
    vehicle
```


Dotazy

Dotazovací jazyky: SPARQL, RDQL, ... (všechny podobné jako SQL)

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
```

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
SELECT ?creator
```

```
WHERE {
```

```
  ?creator dc:creator <http://en.wikipedia.org/wiki/World_Wide_Web>.
```

```
}
```

Aplikace rodiny RDF

FOAF, dbPedia, GoPubMed, WordNet as RDF

