Základy jazyka RDF

Doc. Ing. Vojtěch Svátek, Dr.

Letní semestr 2011

http://nb.vse.cz/~svatek/rzzw.html

Struktura přednášky

- Abstraktní struktura jazyka RDF
- Syntaxe RDF, vztah ke XML
- RDF Schema – definice slovníků

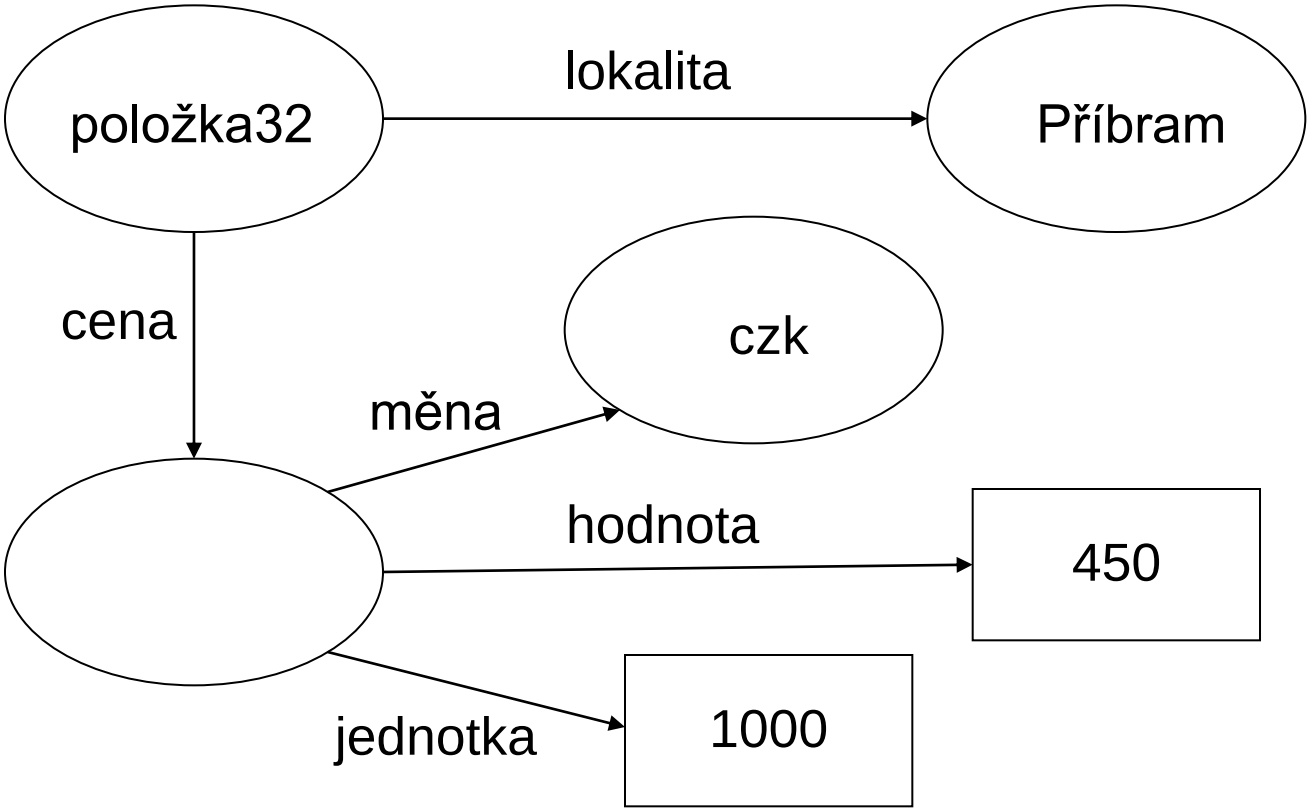
RDF

- “Resource Description Framework”
- Doporučení konsorcia W3C
  `http://www.w3.org/RDF/`
- Jednoduchý jazyk, v němž je možné vyjádřit tvrzení typu “Zdroj X nabývá pro vlastnost Y hodnoty Z” - tzv. *trojice* (“triple”) *subjekt-predikát-objekt*
- Např. (v tabulkovém zápisu):

subjekt	*predikát*	*objekt*
položka32	lokalita	Příbram
položka32	cena	X32
X32	měna	czk
X32	hodnota	450
X32	jednotka	1000

Grafická notace RDF

odlišení zdrojů s URI, anonymních zdrojů („blank nodes“) a literálů



RDF - další možnosti

- Především “typování” zdrojů (rozdělení do tříd) pomocí konstrukce *rdf:type*
- Dále pak mj.
  - sdružování zdrojů do *kontejnerů* a *kolekcí*
  - *reifikace*

Reifikace v RDF

- Využívá předdefinované vlastnosti „subject“, „predicate“, „object“, a typ zdroje „statement“
- Pomocí reifikace se tvrzení se stane adresovatelným objektem
- To vede k možnosti formulovat tvrzení o tvrzeních, např. pro označení autora daného tvrzení
- Nevýhody
  - Nová tři tvrzení (pro subjekt, predikát a objekt) nejsou přímo spojena s původním tvrzením – režie s udržováním integrity
  - Nárůst rozsahu dat
- Dnes chápána jako zastaralá, její roli hraje zpravidla formalismus *pojmenovaných grafů* (named graphs)

RDF versus (nativní) XML

- *modulární* (trojice na sobě nezávislé)
  - datové zdroje je možné libovolně slučovat a rozdělovat
- subjekty, predikáty i některé objekty jsou *zdroje* s jednoznačným *identifikátorem* – URI
  - možnost propojování nezávisle vzniklých zdrojů
- trojice = fakta o světě, kterým lze přiřadit *pravdivostní hodnotu*; nejde jen o strukturu dat jako v případě XML stromů
  - trojice lze korektně interpretovat jako logické formule a *odvozovat* nad nimi
  - samotné RDF ovšem stále nestačí pro strojové odvozování nových informací; potřebujeme zdrojům dodat sémantickou informaci pomocí *RDFS* nebo *OWL*

XML syntaxe RDF

- RDF lze zapisovat (serializovat) pomocí XML, např.:

```
<rdf:RDF xmlns:r="http://www.reality.cz/"
 <rdf:Description
  about="http://www.real-a.cz/polozka32">
   <r:Lokalita
rdf:resource="http://www.mistopis.cz/Pribram"/>
  </rdf:Description>
</rdf:RDF>
```

Subjekt

Predikát

Objekt

- Vedle toho existují stručnější notace, např. Notation3 (autor Tim Berners-Lee):

```
real_a:polozka32 r:Lokalita mist_cz:Pribram .
```

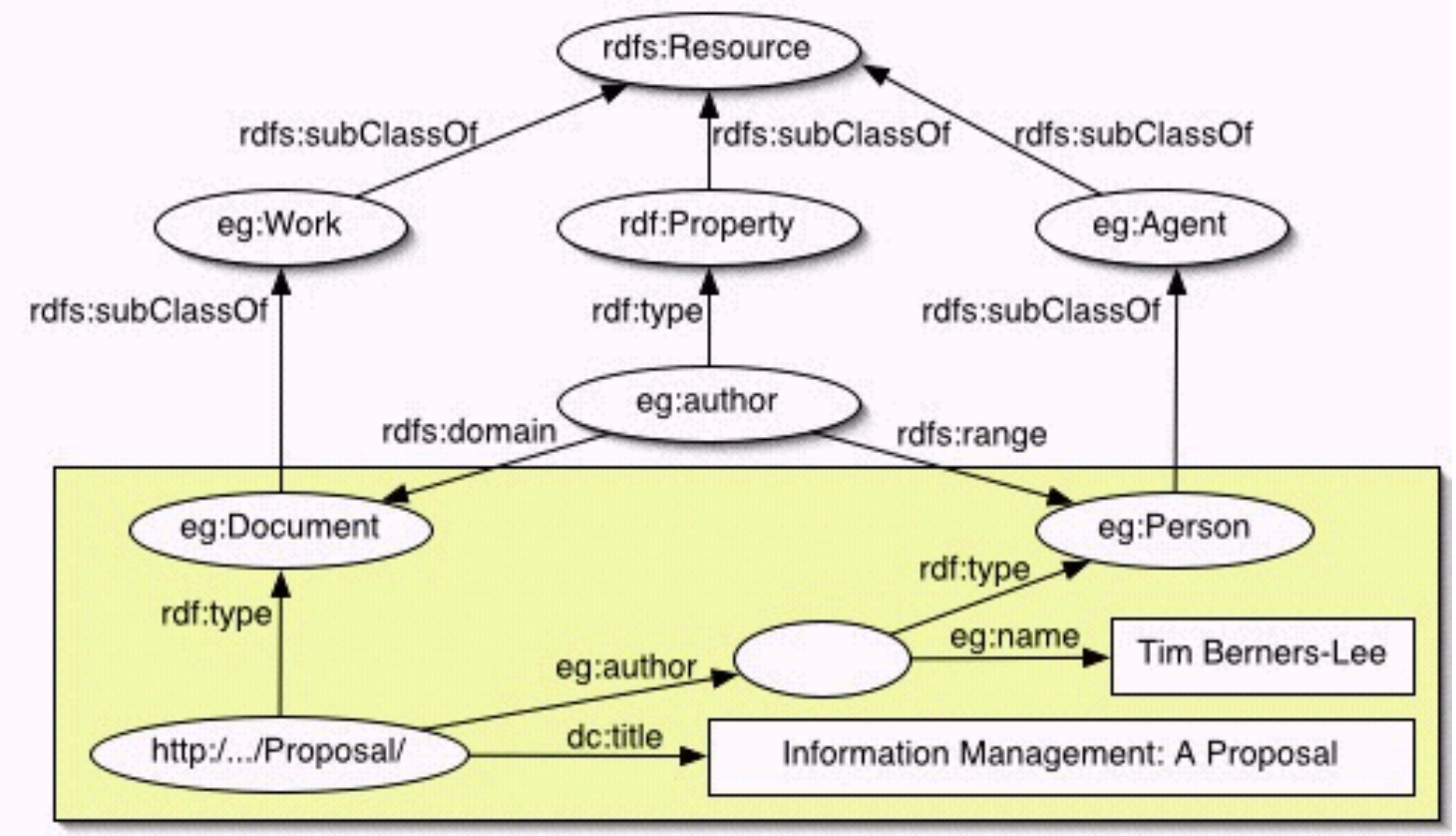
RDF a odvozování

- Nová tvrzení můžeme odvodit zejména tehdy, když konkrétní zdroje přiřadíme k obecným *třídám* jakožto jejich *instance* pomocí *rdf:type*
- Vlastnosti definované u tříd se pak promítají do jejich instancí
- Struktura tříd a jejich vlastnosti mohou být definovány ve *slovnících / ontologiích*
- Hlavní jazyky pro reprezentaci slovníků / ontologií:
  - *RDF Schema*: jednoduchý hierarchický jazyk
  - *OWL*: jazyk s bohatými vyjadřovacími možnostmi, založen na *deskripční logice*

RDF Schema

- Nejde o „datové schéma“ pro RDF, ale o jazyk pro „popis slovníků“ čili tvorbu jednoduchých ontologií („vocabulary description language“)
- Standard zahrnuje možnost specifikovat:
  - vztah třídy a podtřídy, vlastnosti a “podvlastnosti”
    - subclass(Okres,Území)
    - subproperty(sousedí,je_blízko)
  - definiční obor a obor hodnot vlastnosti
    (pozor, nejde o omezení, ale odvozovací pravidla!)
    - domain (lokalita) = Nemovitost
    - range (lokalita) = Území

RDFS - příklad



RDF/S vs. objektový přístup

- Objektový přístup
  - Prvotní jsou třídy a jejich hierarchie
  - Atributy, metody (v programování) a asociace (v modelování) jsou závislé na třídách
- RDF/S a OWL
  - Prvotní jsou individua, jejich množiny a vztahy mezi nimi
  - Na nich jsou teprve závislé (a s jejich pomocí definované) třídy a jejich hierarchie
  - Vlastnosti jsou nezávislé na třídách (třídy mohou být v OWL definované pomocí vlastností, ne naopak!)
  - Mezi objektovými a datovými vlastnostmi je menší rozdíl než mezi asociacemi a atributy v OOP
- Detaily se ozřejmí později, ale touto optikou je vhodné vykládané stále vnímat!