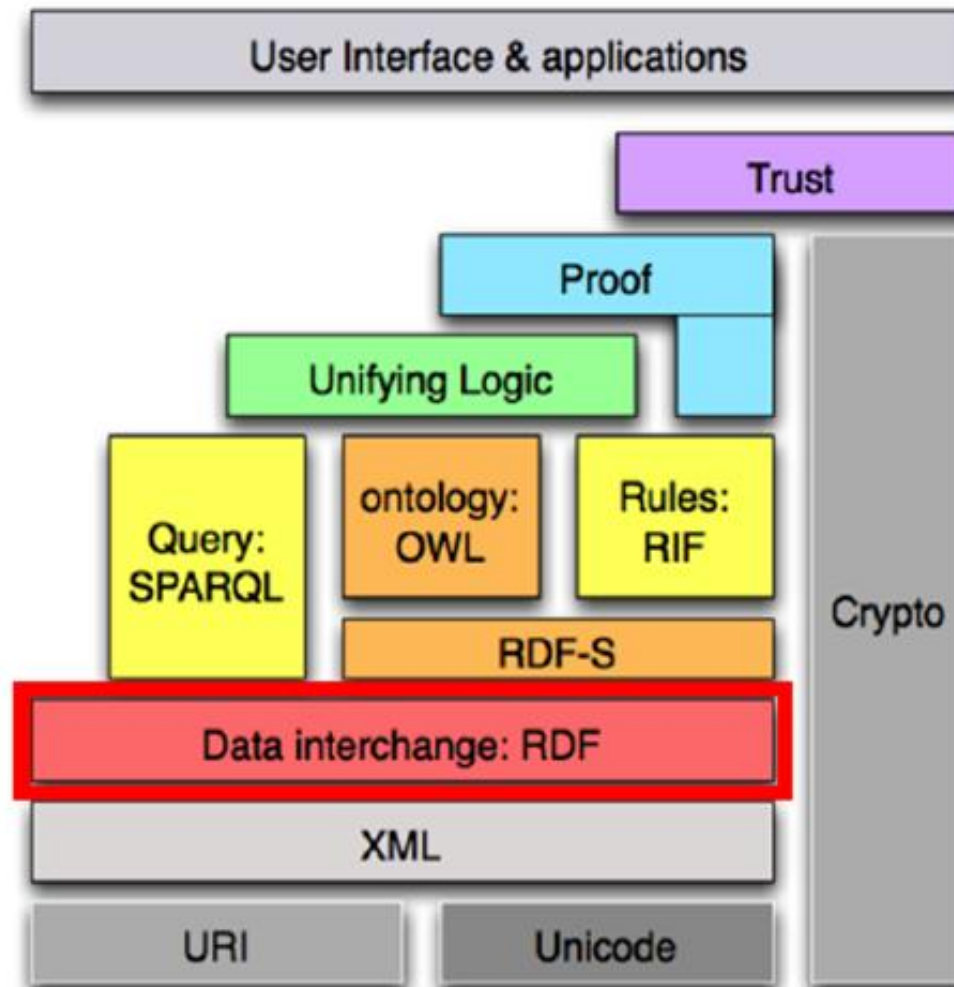


# Resource Description Framework

## RDF



# Resource Description Framework

## RDF

RDF je W3C recommendation – <http://www.w3c.org/RDF>

Definuje **abstraktný dátový model**, ktorý popisuje ako kódovať **poznatky** (informácie) **štruktúrovaným spôsobom**.

- pôvodne určený pre vkladanie metadát do webových stránok,
- v súčasnosti široko používaný aj pre iné účely.
- má univerzálny strojovo spracovateľný formát

### Čo si RDF zobralo z

- relačných databáz
  - formálny matematicko-teoretický prístup
- UML
  - použitie grafov na modelovanie
- XML
  - URI – univerzálny, globálny identifikátor (v rámci webu - čo umožňuje integráciu informácií naprieč celým WEBom).
  - Menné priestory
  - Dátové typy
  - Unicode

# Štrukturovaná informácia

## gramatická stavba vety

Informácia – **tvrdenie** – výrok (*proposition*)

- Mrkvička *sa narodil* 1. januára 1969.
- Mrkvička *prednáša* AIPr.
- AIPr *je* študijný predmet.

Jednoduché tvrdenia majú tvar vety s jednoduchou gramatickou stavbou (syntaxou)

podmet *prísudok* predmet

v terminológii sémantického webu (a RDF)

subjekt *predikát* objekt

# RDF

**RDF tripple** - informácia reprezentovaná ako trojica

**subjekt *predikát* objekt**

**RDF dokument** – množina (neusporiadaná) rdf-tripples.  
Reprezentuje teda istý súbor tvrdení - databázu znalostí.

**Tripple store** - Databáza špeciálne určená pre persistentné ukladanie informácií vo forme RDF tripples.

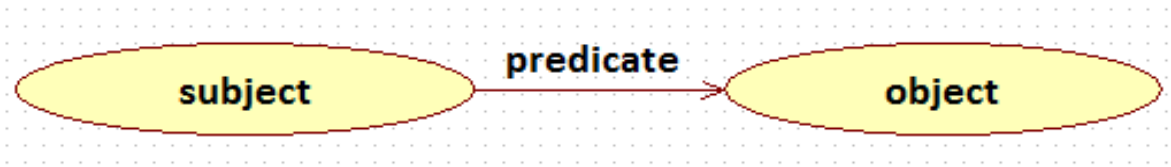
**Pozn.** Niekedy sa používajú aj štvorice – quads (quad store...)

**kontext subjekt predikát objekt**

Kontext spravidla vyjadruje (meta) informáciu o tvrdení. Napr. autor/zdroj tvrdenia.

# RDF - graf

**Trojicu** môžeme znázorniť graficky ako hranu s dvoma koncovými uzlami



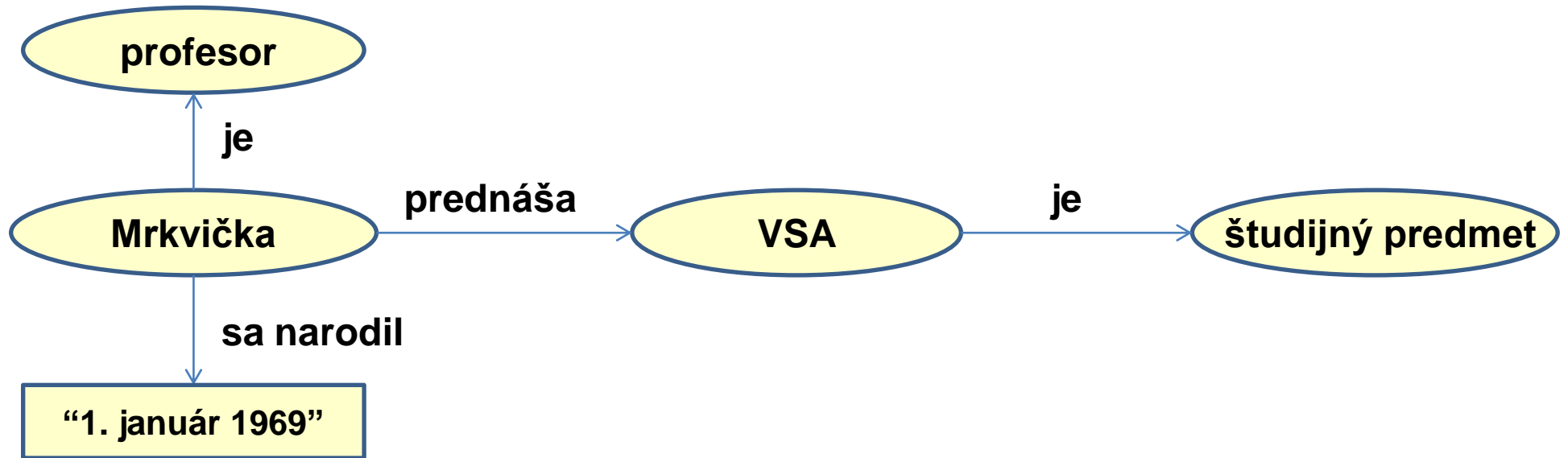
**RDF dokument** môžeme znázorniť potom ako **RDF graf**

*Mrkvička sa narodil 1. januára 1969.*

*Mrkvička je profesor.*

*Mrkvička prednáša VSA.*

*VSA je študijný predmet.*



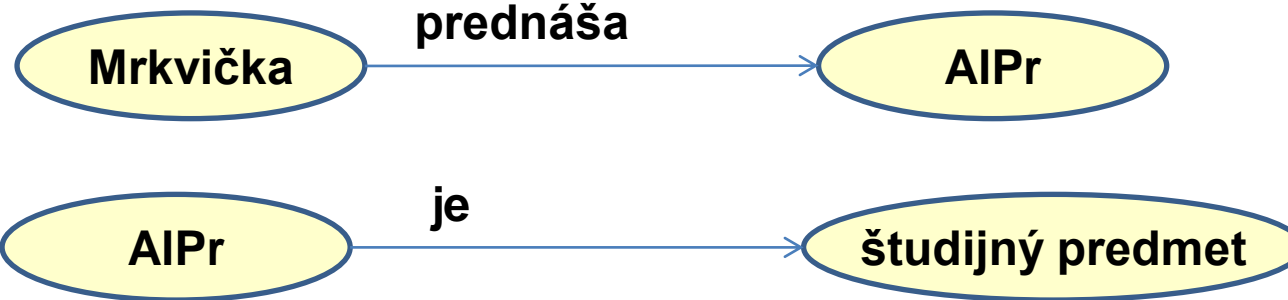
Uzly reprezentujú subjekty a objekty, hrany reprezentujú predikáty.

# Ako kresliť graf

Ten istý člen sa môže vyskytovať vo viacerých trojiciach na rôznych pozíciách.

Mrkvička *prednáša* AIPr.  
AIPr *je* predmet.

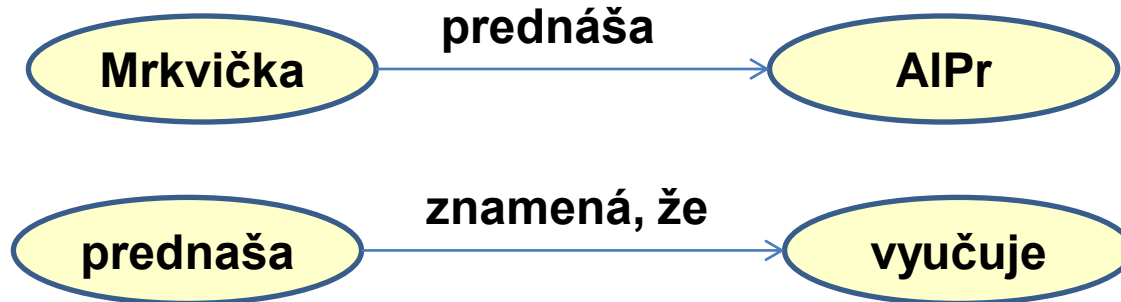
Môžeme síce nakresliť



ale kvôli prehľadnosti sa však odporúča spojiť uzly reprezentujúce ten istý člen



Nie je to však možné vždy, napr. keď sa člen vyskytuje ako predikát aj ako subjekt.



# Vetné členy

Čo môže stáť na mieste jednotlivých vetných členov?

- **IRI – internacionalized resource identifier**

- je to **globálny identifikátor** s definovanou syntaxou  
(rozšírenie **URI** umožňujúce používať aj non-ASCII znaky)
- môže sa nachádzať na ľubovoľnom mieste v trojici



- **Prázdny uzol** (blank node):

- môže sa vyskytovať na mieste **subjektu** alebo **objektu**



- **Literál**: môže sa vyskytovať len na mieste **objektu**



# Interpretácia IRI - Resource

URI (ako ho poznáme z webu) slúži ako globálny identifikátor stránok, dokumentov,... všeobecne akýchkoľvek zdrojov prístupných na webe: web-resourceov.

**IRI** je resource identifikátor, pričom **resource** ako ho chápe RDF je zobecnením pojmu web-resource.

- nemusí to byť len web-resource. URI (IRI) nemusí referencovať nejaký web dokument, nemusí sa na neho navigovať.
- môže to byť akýkoľvek predmet, pojem, vzťah,... konkrétny alebo abstraktný, reálny alebo virtuálny...

**Resource** je zjednodušene povedané môže to byť

**hocičo čo má svoju jednoznačnú identitu a význam – sémantiku**

Na rozdiel of web-resource, identita rdf-resource však nemusí byť známa. **rdf-resource nemusí mať definované IRI.**



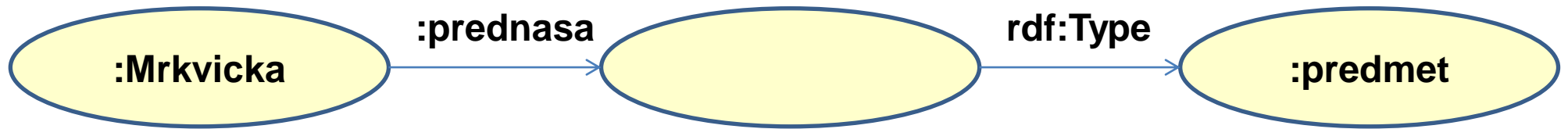
# Interpretácia prázdnych uzlov

Občas nastane situácia, že o existencii akéhosi objektu síce vieme, nevieme ho však presne pomenovať – identifikovať. Napríklad chcem vyjadriť, že

Mrkvicka prednáša **nejaký** predmet

ale už si nepamätám ktorý, teda nepoznám ani jeho URI.

Uzol reprezentujúci neznámy objekt ponechám **prázdny**



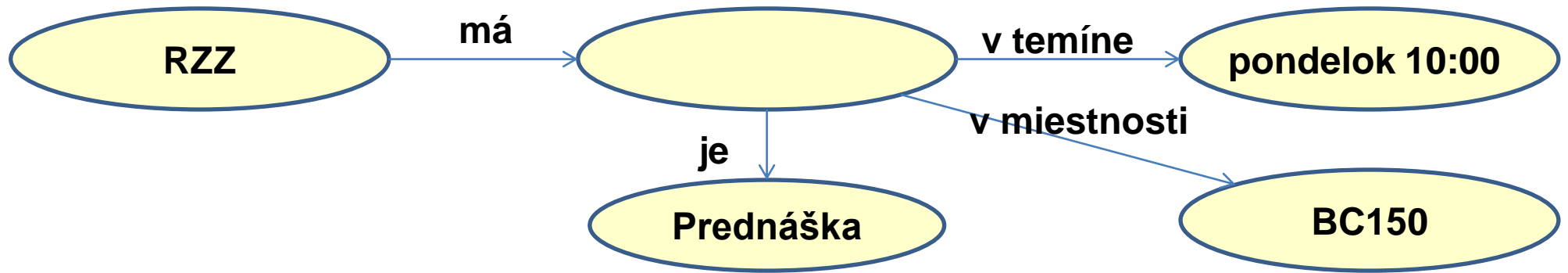
Prázdny uzol

- reprezentuje resource, ktorý nie je pomenovaný - **nemá URI**.
- interpretacia je **existenčné tvrdenie**. Hovoríme o niečom čo existuje aj keď sme to nepomenovali, a nemusíme poznať bližšie ani jeho identitu.

# Interpretácia prázdnych uzlov

Prázdne uzly umožňujú (nepriamo) vyjadriť aj viacnásobne vzťahy

## Prednášky z RZZ sú pondelok 10:00 v BC150



Prázdny uzol tu reprezentuje čosi - v AISe sa to nazýva rozvrhová akcia, čo má svoj čas a miesto.

**Pozn.** Hoci slovné spojenia **v termíne** resp. **v miestnosti** nie sú (prísne vzaté) slovesá, reprezentujú predikáty s významom **byť v miestnosti** resp. **konať sa v termíne**.

Graf teda môžeme interpretovať:

*RZZ má **akúsi** rozvrhovú akciu.*

*Táto akcia je prednáška.*

*Táto akcia je v miestnosti BC150.*

*Táto akcia sa koná v termíne pondelok 10:00.*

# Interpretácia literálov

Mozart má krstné meno **Wolfgang**.

Samotné slovo „**Wolfgang**“ neoznačuje žiadnu konkrétnu osobu, objekt alebo pojem. **Mimo kontextu nemá svoju identitu ani význam**, nemôže to byť teda resource, je to len textový reťazec – literál.

## Literály

- sú znakové reťazce

“Wolfgang” “1756” “27.Jan.1756”

- reprezentujú dátové hodnoty
- môžu byť asociované s **dátovým typom**

“1756-01-27”<sup>^xsd:date</sup>

ktorý definuje spôsob ich interpretácie – sémantiku

- literály **bez dátového typu** sú chápané ako textové reťazce – **plain-text**

“1756-01-27”<sup>^xsd:date</sup> je dátum ale “1756-01-27” je len textový reťazec

- literály bez dátového typu môžu byť **jazykovo lokalizované**

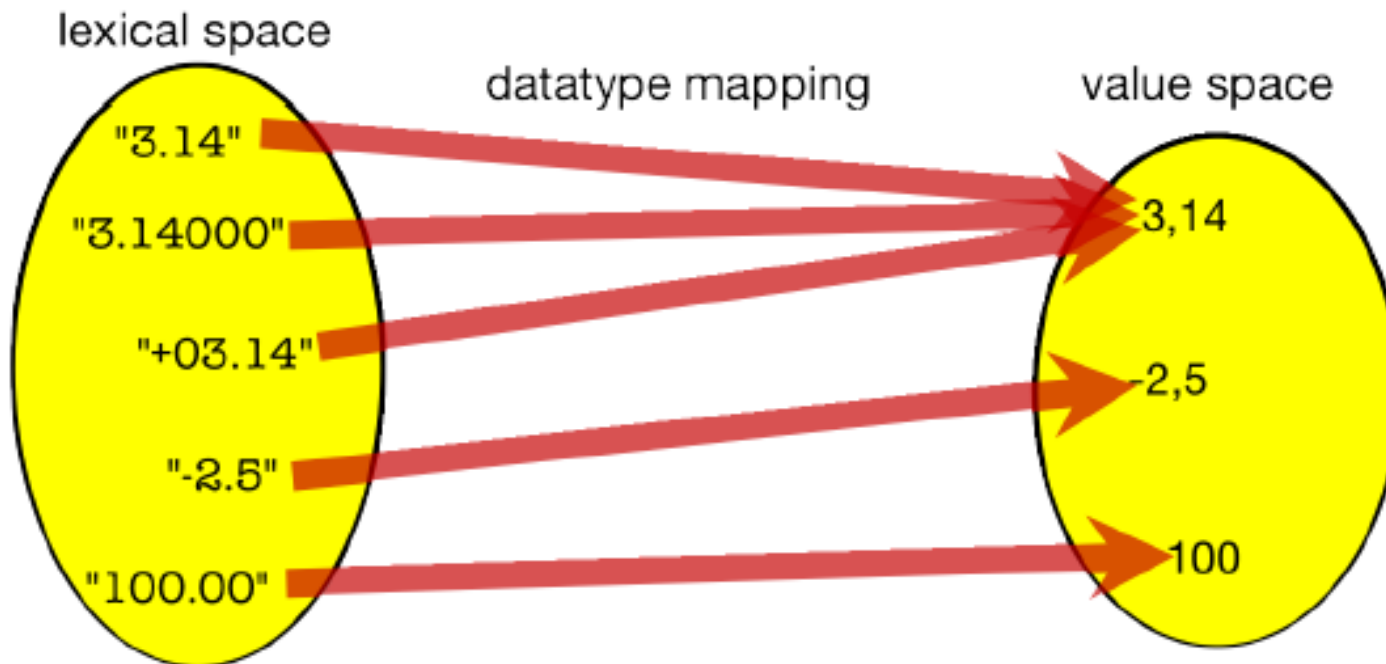
“Bratislava”

“Pressburg”@de

“Bratislava”@sk

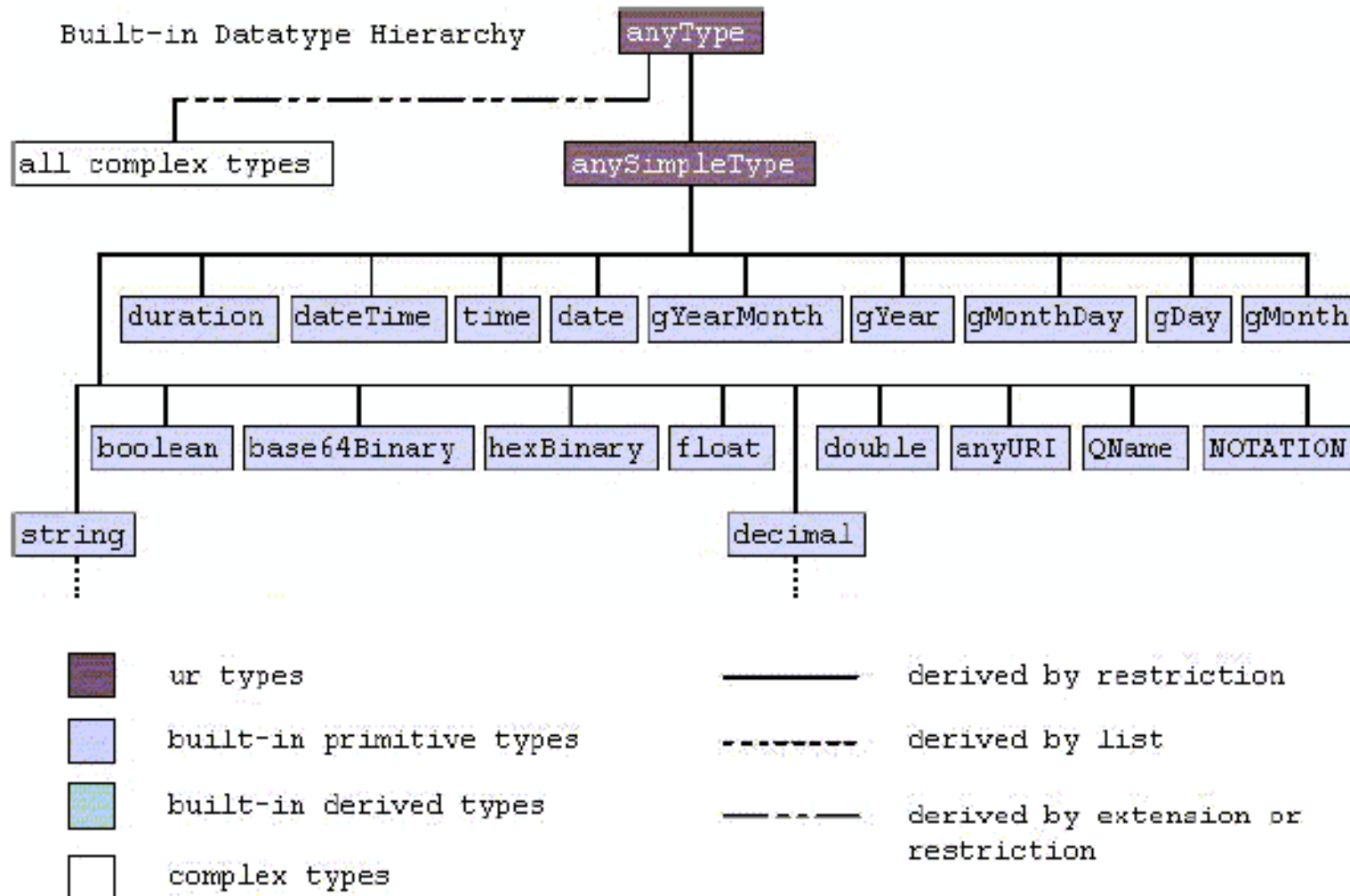
# Datatypes

Example: `xsd:decimal`



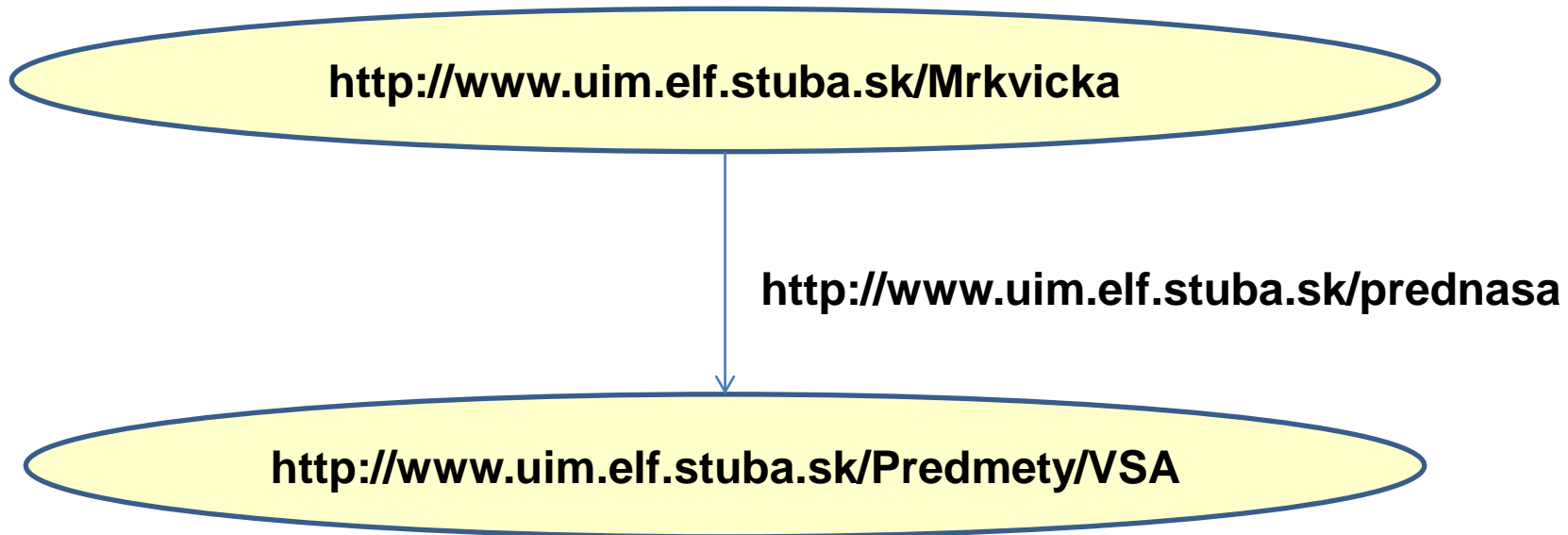
For `xsd:decimal` holds "3.14"=" +03.14"  
But not for `xsd:string` !

# XML Schema Datatypes



# Ako zvolit' vhodné IRI pre Mrkvičku?

IRI má byť **globálne jedinečný identifikátor** →  
použiť meno internetovej domény, ktorú máme zaregistrovanú,  
napr. [stuba.sk](http://stuba.sk)



## Problémy:

- príliš dlhé identifikátory
- stále nejednoznačná interpretácia (*Mrkvicka je osoba alebo zelenina*)

# Prefixy pre menné priestory

Pre zjednodušenie zapisu RDF grafu si môžeme definovať skratku - prefix pre všetky URI ktoré s ním začínajú napr.

**elf:** pre <http://www.uim.elf.stuba.sk/>

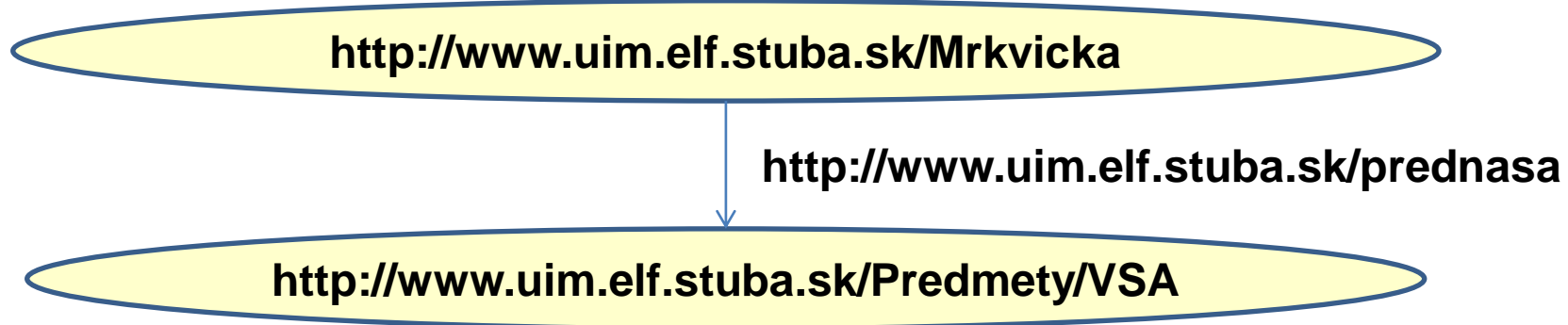


*Všimnite si, že prefix oddeluje od zvyšku URI dvojbodka.*

Môžeme si tiež definovať **default**-ný menný priestor pre RDF-graf. Ako prefix má prázdny reťazec.



# (ne)Jednoznačná interpretácia URI



- *Čo je Mrkvička, človek alebo zelenina?*

**Odporúčenie:** ak používate meno web-domény (ktorú vlastníte) na definovanie identifikátorov pre pojmy - ресурсы, vytvorte v nej si web-stránku (aj príp. podstránkami), na ktorých bude neformálne popísaný význam vami definovaných identifikátorov.

- <http://www.uim.elf.stuba.sk/Predmety/VSA> je stránka predmetu VSA, Mrkvička však prednáša predmet nie stránku, tú by mohol vytvoriť...

**Odporúčenie:** Jasne rozlišujte medzi URI resoursu a URI web-stránky, ktorá o ňom hovorí. Vhodné je používať jednotnú stratégiu pre tvorbu URI stránok a resoursov.

*Napr. na identifikáciu resoursov preprezentujucich predmety použime URI s prefixom: <http://www.uim.elf.stuba.sk/Predmety/Vocab#>*



# Aké URI použiť pre Mozarta?

- [https://sk.wikipedia.org/wiki/Wolfgang\\_Amadeus\\_Mozart](https://sk.wikipedia.org/wiki/Wolfgang_Amadeus_Mozart)

asi nie - je to stránka slovenskej wikipedie o ňom.

- Definovať vlastné napr. <http://www.uim.elf.stuba.sk/Mozart>

tiež nie - to môže urobiť každý a budeme mať množstvo rôznych URI pre ten istý pojem/resource

- Vyhladať a použiť IRI, ktoré pre daný pojem/resource už definovala nejaká autorita v danej oblasti (*ja to naozaj nie som* 😊).

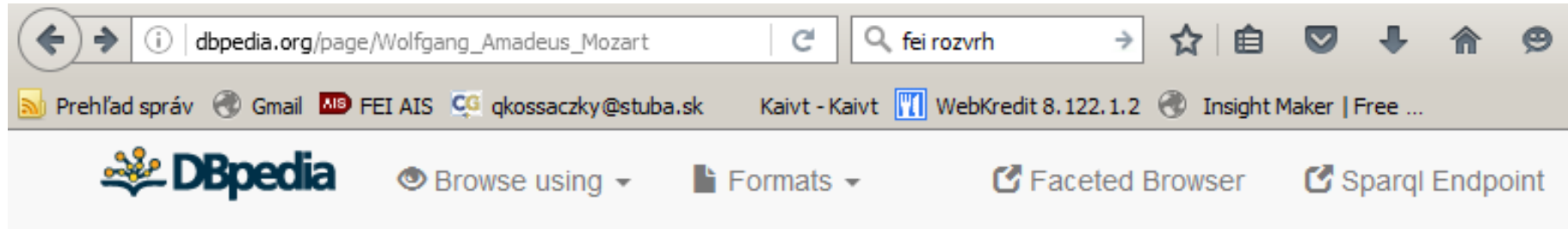
[http://dbpedia.org/resource/Wolfgang\\_Amadeus\\_Mozart](http://dbpedia.org/resource/Wolfgang_Amadeus_Mozart)

# Príklad - DBpedia

Dbpedia definovala vlastné URI pre Mozarta

[http://dbpedia.org/resource/Wolfgang\\_Amadeus\\_Mozart](http://dbpedia.org/resource/Wolfgang_Amadeus_Mozart)

A vytvorila stránku, ktorá toto uri popisuje.



The screenshot shows a web browser window with the address bar containing 'dbpedia.org/page/Wolfgang\_Amadeus\_Mozart'. The search bar contains 'fei rozvrh'. The browser's address bar also shows several extensions: 'Prehľad správ', 'Gmail', 'FEI AIS', 'qkossaczky@stuba.sk', 'Kaivt - Kaivt', 'WebKredit 8.122.1.2', and 'Insight Maker | Free ...'. Below the browser window, the DBpedia logo is visible, along with navigation options: 'Browse using', 'Formats', 'Faceted Browser', and 'Sparql Endpoint'.

## About: Wolfgang Amadeus Mozart

An Entity of Type : [agent](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

Wolfgang Amadeus Mozart (German: [ˈvɔlfɡaŋ amaˈdeːʊs ˈmoːtʰsʰʌʁt], English see fn.; 27 January 1756 – 5 December 1791), baptised as Johannes Chrysostomus Wolfgangus Theophilus Mozart, was a prolific and influential composer of the Classical era. Born in Salzburg, Mozart showed prodigious ability from his earliest childhood. Already competent on keyboard and violin, he composed from the age of five and performed before European royalty.

Property	Value
<a href="#">dbo:abstract</a>	▪ Wolfgang Amadeus Mozart (German: [ˈvɔlfɡaŋ amaˈdeːʊs ˈmoːtʰsʰʌʁt], English

# Terminologické slovníky

*V súvislosti s jednoznačnosťou URI ide v skutočnosti o dva problémy:*

- 1) Aby sa **jedno URI** nepoužívalo v rôznych významoch, na identifikáciu **rôznych vecí** - *jednoznačnosť URI*.
- 2) Aby sa pre **jeden pojem** nepoužívali **viaceré URI** - *jedno-jednoznačnosť URI*.

*Aby sa vyhlo 2. problému nie je rozumné aby URI pre všeobecne známe objekty/pojmy definoval hocikto.*

Vhodné je zveriť túto úlohu všeobecne rešpektovanej **AUTORITE**, ktorá definuje sémantiku a URI pre objekty/pojmy z nejakej oblasti, vytvára a spravuje **TERMINOLOGICKÉ SLOVNÍKY** (controlled vocabulary, tezaury, ontologie,...)

*Takéto authority aj slovníky už existujú, dajú sa rozdeliť na dva druhy*

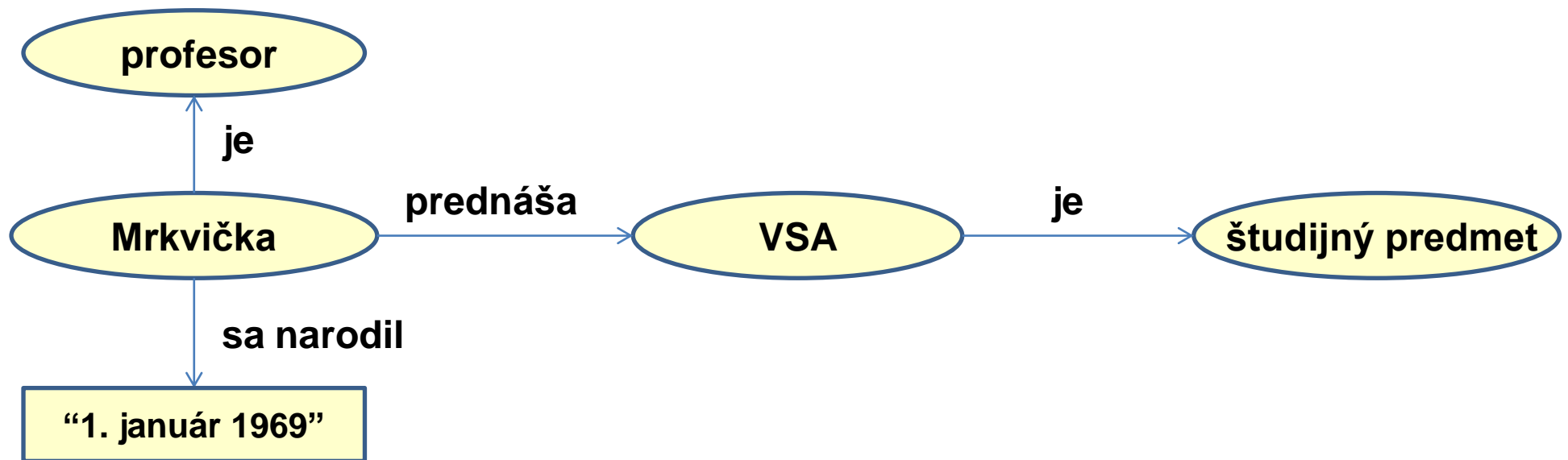
- **Všeobecné**. napr. DUBLIN CORE
- **doménovo špecifické** napr. SNOMED\_CT

Navyše samotný štandard RDF poskytuje slovníky **RDF** a **RDFS** definujúce význam - sémantiku pre **najobecnejšie abstraktné pojmy**.

***Pozn.** ďalšie riešenie 2. problému poskytujú ontologie – umožňujú povedať že dve URI predstavujú ten istý objekt/pojem.*

# Taxonómie

# Predikát byt'



Sloveso **je** sa tu vyskytuje viac krát v zmysle byt' – **mat' typ** (byt' inštanciou druhu).  
Matematicky vyjadrené: **byt' prvkom množiny**.

Rovnakú sématicku má slovo je aj vo vetách.

Mrkvicka **je** profesor.

Acylpirín **je** liek.

Keďže sloveso byt' sa v tom to zmysle používa veľmi často vo všetkých oblastiach (doménach) URI pre vzťah s takouto sémantikou definuje priamo štandard RDF.

**rdf:type**

Kde rdf je prefix pre menný priestor

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

# Jedinci a triedy (prvky a množiny)

**rdf:type** je URI pre **vzt'ah** vyjadrujúci príslušnosť jedinca do triedy

:Mrkvicka rdf:type :Profesor

Resorce na pozícii subjektu je chápaný ako **jedinec** (individual).

Resource na pozícii objektu je chápaný ako **trieda** (class) – množina jedincov.

Jeden jedinec môže mať aj viac typov t.j. patriť do viacerých tried.

:Mrkvicka rdf:type :Muz

:Mrkvicka rdf:type :Profesor

Ten istý resource sa môže vyskytovať na mieste jedinca aj triedy

:Mrkvicka rdf:type :Profesor

:Profesor rdf:type :AkademickaHodnost

# Nie je je ako je

Pozor na významový rozdiel:

Profesor	je	akademická hodnosť.
Profesor	je	učiteľ.

V prvom prípade by sme mali použiť `rdf:type` v druhom nie. Prečo?

# Podtriedy

Výrokom **Profesor je učiteľ.**

chceme v skutočnosti vyjadriť skutočnosť, že

**každý** profesor je učiteľ.

Skutočnosť, že **každý** jedinec jednej triedy je aj členom inej triedy, vyjadrujeme v RDF pomocou **vzťahu** **rdfs:subclassOf**, teda:

**:Profesor rdfs:subclassOf :Ucitel**

V jazyku matematiky by sme povedali, že: jedna množina **je podmnožinou druhej** množiny.

Z takto chápaného významu (sémantiky) vzťahu podtrieda vyplývajú niektoré jeho dôležité vlastnosti.



# Prečo nie je je ako je - odpoveď

Predikát **rdfs:subclassOf** má špecifickú vlastnosť:

Zo skutočností:

:Mrkvicka	<b>rdf:type</b>	:Profesor
:Profesor	<b>rdfs:subclassOf</b>	:Ucitel

vyplýva, že:

:Mrkvicka	<b>rdf:type</b>	:Ucitel
-----------	-----------------	---------

Predikát **rdf:type**, túto vlastnosť nemá, z tvrdení:

:Mrkvicka	<b>rdf:type</b>	:Profesor
:Profesor	<b>rdf:type</b>	:AkademickaHodnost

by totiž nemalo nevyplývať, že:

:Mrkvicka	<b>rdf:type</b>	:AkademickaHodnost
-----------	-----------------	--------------------

# Vlastnosti podtriedy

**Tranzitivita:** zo skutočností

Pes                      rdfs:subclassOf              Cicavec  
Cicavec                  rdfs:subclassOf              Živočích

vyplýva:

Pes                      rdfs:subclassOf              Živočích

Vzťah podtriedy umožňuje (vd'aka tranzitivite) vytvárať celé hierarchie tried – **taxonómie**.

Napr. taxonómie živočíšnych druhov.

**Reflexivita:** Pes              rdfs:subclassOf              Pes

Vzťah podtriedy umožňuje tiež vyjadriť, že dve rôzne URI označujú tú istú triedu:

Pes rdfs:subclassOf Dog  
Dog rdfs:subclassOf Pes

# Porovnanie terminológie

RDF	Teória množín	OOP/OOM
Trieda (Class)	množina	trieda
Jedinec (Individual)	prvok	objekt
rdf:type	$\in$	inštancia
rdfs:subclassOf	$\subseteq$	podtrieda