

RDF ako algebraická štruktúra

Spracované podľa

dokumentov W3C

- <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>
- <http://www.w3.org/TR/rdf11-mt/#notation-and-terminology>

Abstraktná syntax RDF

RDF ako formálny jazyk

Čo sú **slová jazyka RDF**? Slová jazyka možno rozdeliť do 3 disjunktných množín :

- **IRI** (množina všetkých platných IRI - zobecnené URI)
- **bnode** (množina prázdnych uzlov)
- **literal** (množina platných literálov)

Pozn. Abstraktná syntax tieto 3 množiny už nijako bližšie nešpecifikuje. Ich prvky všeobecne nazýva uzly – nodes.

Čo je **veta jazyka RDF**?

- každá **rdf-trojica** - triple:
 $(s, p, o) \in (\text{IRI} \cup \text{bnode}) \times \text{IRI} \times (\text{IRI} \cup \text{bnode} \cup \text{literal})$
je veta.
- každá **konečná množina rdf-trojíc** - **rdf-graf** je veta.

Ďalšie termíny

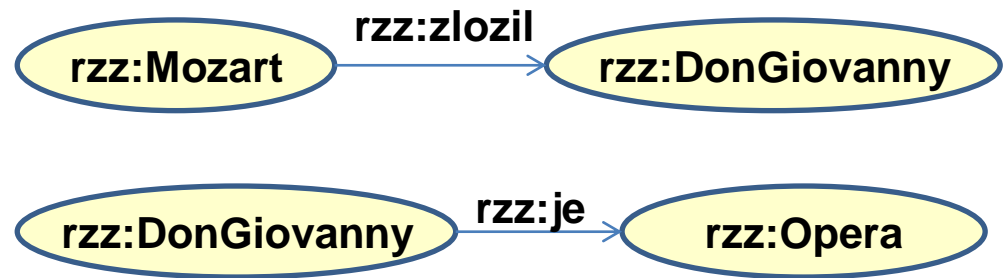
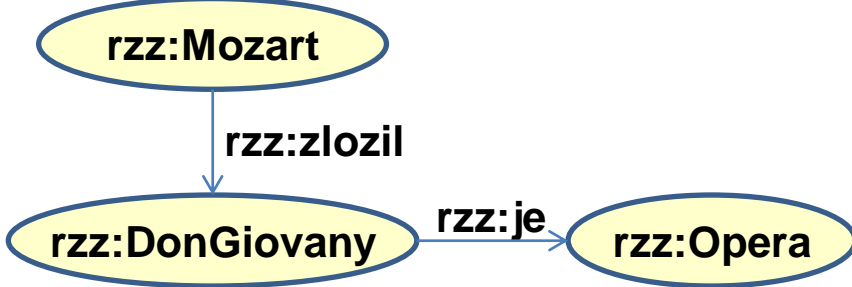
- **Meno** – je IRI alebo literál
- **Slovník RDF grafu (vocabulary)** – je množina mien použitých v grafe
- **Základná (ground) trojica** – je rdf-trojica, ktorá neobsahuje prázdne uzly
- **Základný(ground) graf** – je rdf-graf, ktorý neobsahuje prázdne uzly

Abstraktná sémantika RDF

Kedy možno považovať RDFy za identické, rovnaké?

Keďže RDF je množina aj graf, môžeme ho chápať a interpretovať čisto ako abstraktnú matematickú štruktúru a študovať jeho formálne vlastnosti prostriedkami teórie množín a teórie grafov (bez ohľadu na to čo konkrétny graf reprezentuje)

RDF ako graf:



identický ale sa neodporúča

RDF ako množina:

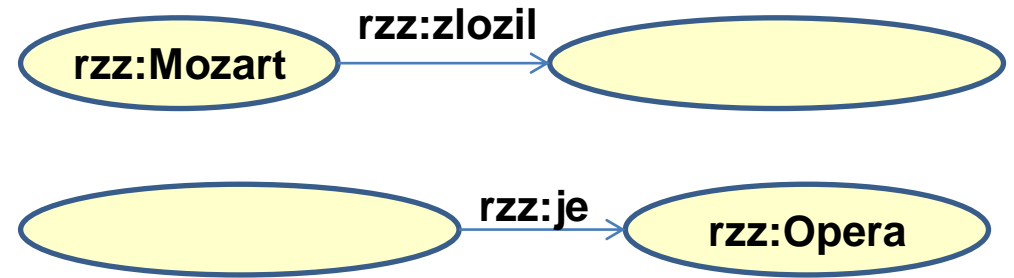
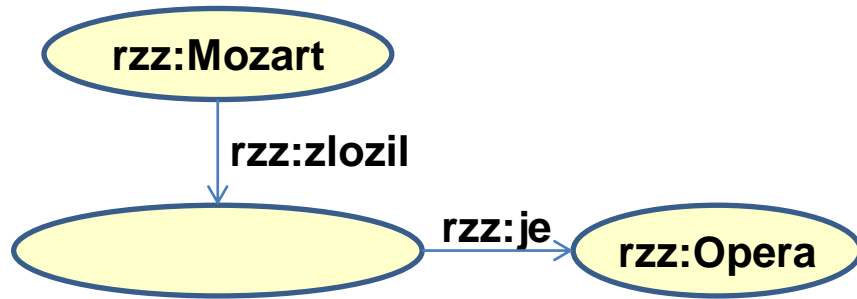
- nezáleží na poradí trojíc
- duplicitné trojice sú ignorované

Nasledujúce RDF dokumenty sú identické

| RDF1 | RDF2 | RDF3 |
|----------|----------|----------|
| s1 p1 o1 | s1 p1 o1 | s2 p2 o2 |
| s2 p2 o2 | s2 p2 o2 | s1 p1 o1 |
| | s1 p1 o1 | |

Problém prázdnych uzlov

Tieto dva RDF-grafy **nie sú identické**



zrejmé je to z ich prepisu do RDF dokumentu:

```
rzz:Mozart rzz:zlozil _:b1
_:b1 rzz:je rzz:Opera
```

```
rzz:Mozart rzz:zlozil _:b1
_:b2 rzz:je rzz:Opera
```

RDF-dokumenty (v turtle alebo inej syntaxe) používajú na označenie prázdnych uzlov lokálne identifikátory (identifikátory začínajúce prefixom _:). Pokiaľ sa dva grafy líšia len rôznymi identifikátormi prázdnych uzlov môžeme ich považovať za identické.

Nasledujúce RDF-dokumenty sú identické (sú zápisom RDF-grafu v vľavo)

```
rzz:Mozart rzz:zlozil _:b1
_:b1 rzz:je rzz:Opera
```

```
rzz:Mozart rzz:zlozil _:b2
_:b2 rzz:je rzz:Opera
```

Izomorfizmus

To, že dva grafy môžeme považovať za identické sa formálne dá vyjadriť pomocou pojmu **izomorfizmus**. Dva grafy sú izomorfné, ak existuje bijektívne zobrazenie medzi množinami ich prázdnych uzlov, ktoré zobrazuje jeden graf na druhý.

Definícia:

Two [RDF graphs](#) G and G' are *isomorphic* (that is, they have an identical form) if **there is a bijection** M between the sets of nodes of the two graphs, such that:

1. M maps blank nodes to blank nodes.
2. $M(lit)=lit$ for all [RDF literals](#) lit which are nodes of G .
3. $M(iri)=iri$ for all [IRIs](#) iri which are nodes of G .
4. The triple (s, p, o) is in G if and only if the triple $(M(s), p, M(o))$ is in G ,

See: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/#graph-isomorphism>

Príklad:

| | | | | | |
|------------|------------|-----------|------------|------------|-----------|
| rzz:Mozart | rzz:zlozil | _:b1 | rzz:Mozart | rzz:zlozil | _:b2 |
| _:b1 | rzz:je | rzz:Opera | _:b2 | rzz:je | rzz:Opera |

Riešenie: Sú izomorfné, pretože môžeme definovať bijekciu takto

$M: _ :b1 \rightarrow _ :b2$

Spájanie grafov

Jedna z požiadaviek na rdf je podpora pre integráciu. Pri spájaní RDF-grafov resp. dokumentov vzniká opat' problém interpretácie prázdnych uzlov: Ak máme prázdne uzly v dvoch rôznych grafoch označené rovnakým identifikátorom vzniká otázka je, či ide teda o jeden zdielaný uzol alebo dva rôzne uzly. Možné sú obe interpretácie.

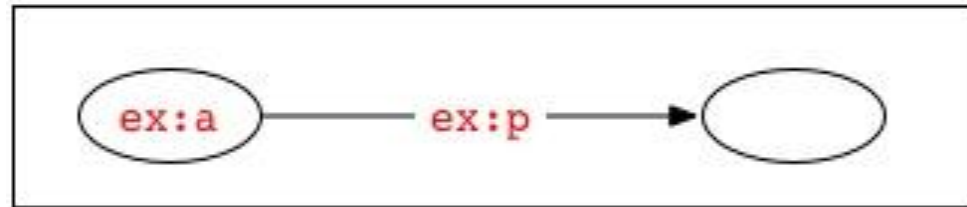
Zjednotenie (union) – považuje prázdne uzly s rovnakým identifikátorom za ten istý zdielaný uzol.

Zlúčenie (merge) – považujeme prázdne uzly z pôvodných grafov za rôzne (aj keď majú rovnaké identifikátory).

Pozn. v tomto prípade treba uzly vo výslednom grafe odlíšiť novými rôznymi identifikátormi.

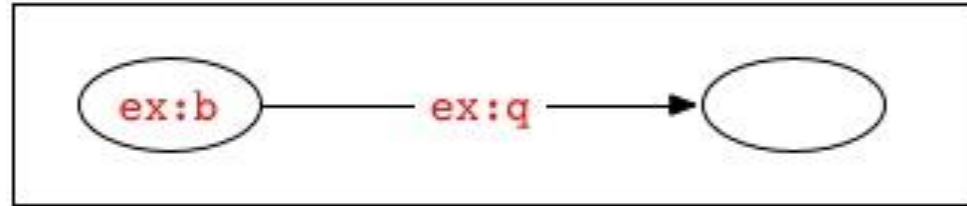
Dokument 1:

ex:a ex:p _:x .



Dokument 2:

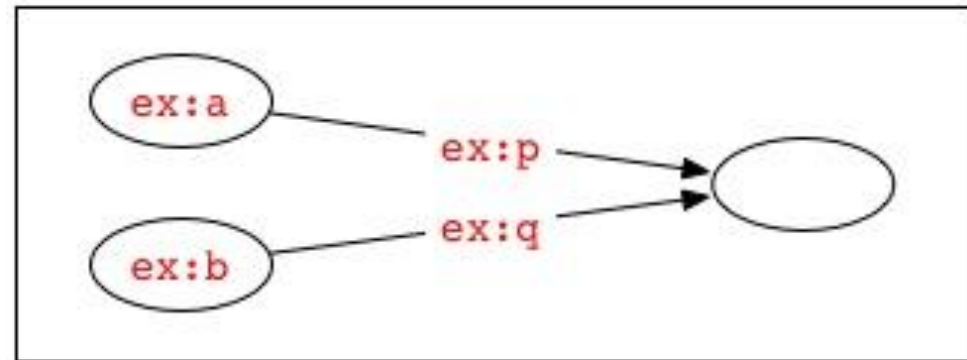
ex:b ex:q _:x .



Zjednotené dokumenty:

ex:a ex:p _:x .

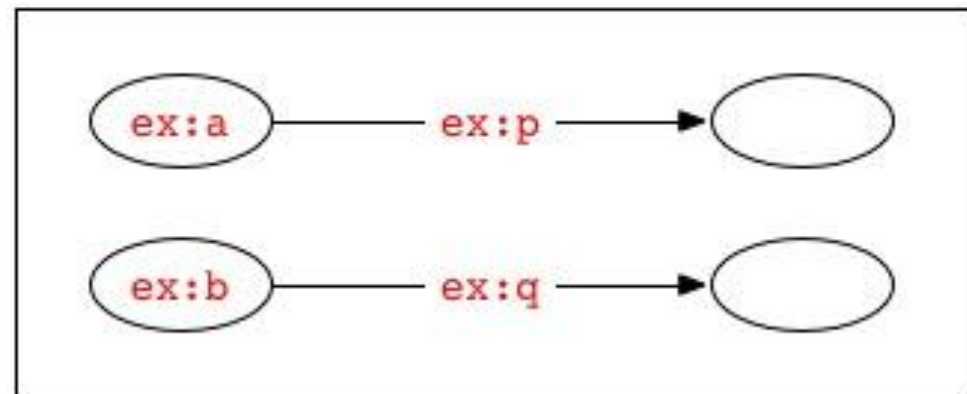
ex:b ex:q _:x .



Zlúčené dokumenty:

ex:a ex:p _:x1 .

ex:b ex:q _:x2 .



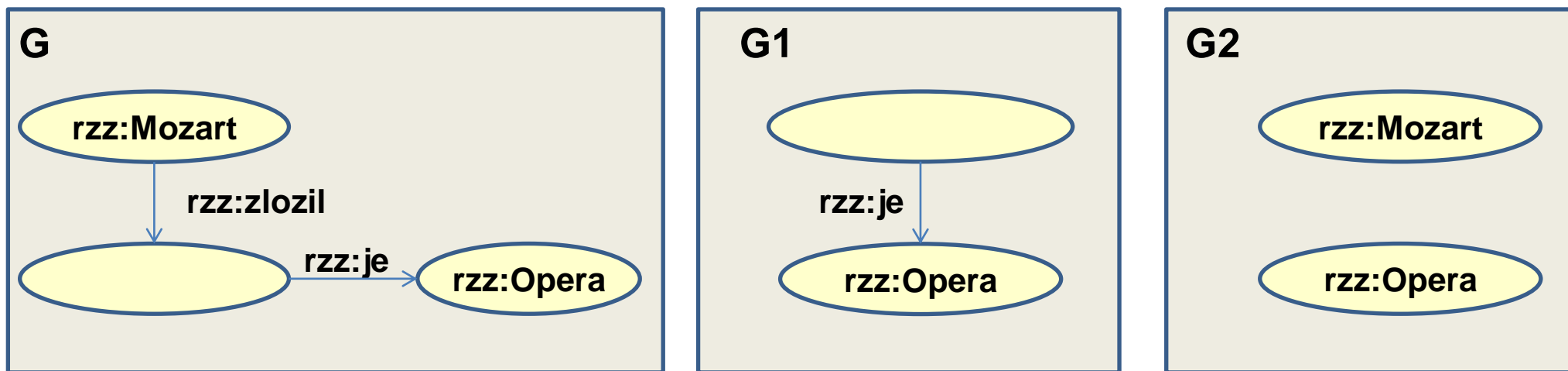
Podgraf

Keď graf chápeme ako množinu (množinu rdf-trojíc) môžeme definovať nasledujúce pojmy známe z TM.

Prázdny graf – je prázdna množina rdf-trojíc

Podgraf rdf-grafu G – je akákoľvek podmnožina množiny rdf-trojíc grafu G.

Vlastný podgraf grafu G je podgraf vlastná podmnožina množiny rdf-trojíc grafu G.

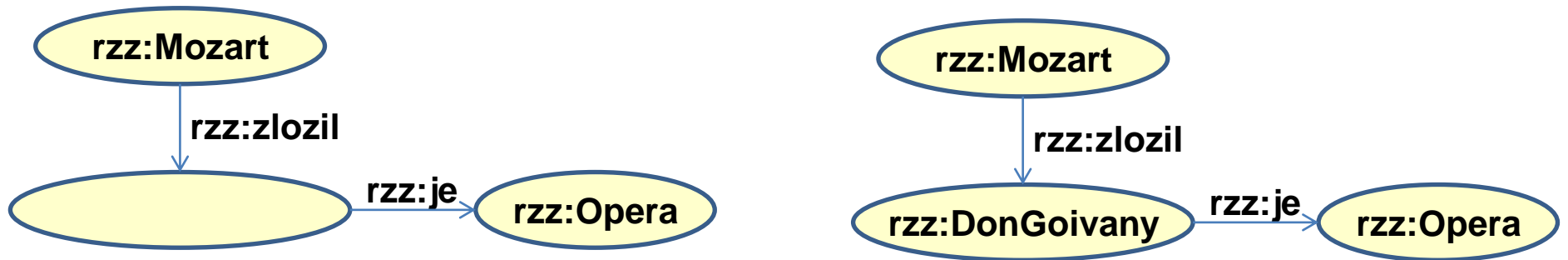


G1 je vlastný podgraf rdf-grafu G,

G2 neobsahuje žiadnu hranu, uzly ktoré nemajú neležia žiadnej hrane sú zbytočné.

Pozn. Všimnime si , že ak hrany interpretujeme ako tvrdenia, teda celý graf predstavuje istú množinu informácií/vedomostí, potom každý jeho podgraf je podmnožinou tejto množiny informácií.

Inštancia grafu



Podobne aj v tomto príklade, graf vľavo predstavuje podmnožinu informácií reprezentovaných grafom vpravo. Nie je však podgrafom. Evidentné je to z ich prepisu so RDF-dokumentu.

rzz:Mozart rzz:zlozil _:b1
_:b1 rzz:je rzz:Opera

rzz:Mozart rzz:zlozil rzz:DonGiovany
rzz:DonGiovany rzz:je rzz:Opera

Definícia: Nech G je rdf-graf a M je nejaká množina obsahujúca IRI, literály aj prázdne uzly. Nech f je zobrazenie z množiny prázdnych uzlov grafu G do M . Potom graf G' , ktorý vznikne nahradením niektorého alebo viacerých prázdnych uzlov b prvkom $f(b)$, nazývame **inštanciou grafu G** .

Pozn. Množina M môže obsahovať IRI, literály a uzly grafu G , môže však obsahovať aj také, ktoré v ňom nie sú.

V príklade z obrázku je graf v pravo inštanciou grafu vľavo, pretože funkciu f môžeme definovať nasledovne: **$f: _ :b1 \rightarrow rzz:DonGiovany$**

Pozn. Pozorný čitateľ si mohol všimnúť, že dva grafy sú izomorfné práve vtedy keď sú si navzájom inštanciami.

Príklady inštancie

Graf

| | | |
|------|------|-------|
| ex:a | ex:q | _:x |
| ex:b | ex:p | _:y |
| _:y | ex:p | “aaa” |

Inštancia

| | | |
|------|------|-------|
| ex:a | ex:q | ex:z |
| ex:b | ex:p | _:y |
| _:y | ex:p | “aaa” |

Inštancia

| | | |
|------|------|-------|
| ex:a | ex:q | ex:z |
| ex:b | ex:p | ex:w |
| ex:w | ex:p | “aaa” |

Inštancia

| | | |
|------|------|-------|
| ex:a | ex:q | ex:z |
| ex:b | ex:p | _:n |
| _:n | ex:p | “aaa” |

Inštancia

| | | |
|------|------|-------|
| ex:a | ex:q | _:n |
| ex:b | ex:p | _:n |
| _:n | ex:p | “aaa” |

Inštancia

| | | |
|------|------|---------|
| ex:a | ex:q | “Hello” |
| ex:b | ex:p | _:y |
| _:y | ex:p | “aaa” |

nie je inštancia. Prečo?

| | | |
|------|------|-------|
| ex:a | ex:q | ex:z |
| ex:b | ex:p | ex:w |
| _:y | ex:p | “aaa” |

nie je inštancia. Prečo?

| | | |
|---------|------|---------|
| ex:a | ex:q | ex:z |
| ex:b | ex:p | “Hello” |
| “Hello” | ex:p | “aaa” |