

Životný cyklus informácie na sémantickom webe

Martin Svihla, Ivan Jelinek

{svihlm1, jelinek}@fel.cvut.cz

Research Group Webing

Department of Computer Science

Czech Technical University in Prague

Obsah

- Motivácia
- Kroky životného cyklu
- Anotácia webových zdrojov metadátami
- Zbieranie sémantických metadát
- Vytvorenie znalostnej bázy
- Dotazovanie znalostnej bázy
- Záver

Motivácia

- **Vytvorenie sémantického vyhľadávania:**
Nájdí webové stránky, na ktorých je spomenutá
“osoba s krstným menom Martin“
- **Ciele:**
 - Preskúmanie fáz životného cyklu informácie na sémantickom webe (SW)
 - Zistenie momentálnych možností technológií SW
 - Experimenty s jazykmi a vývojárskymi nástrojmi SW

Životný cyklus informácie (1/2)

Strana producenta informácie:

- Vytvorenie ontológie popisujúcej danú znalostnú doménu
- Vytvorenie metadát obsahujúcich informácie
- Publikovanie metadát a ich pripojenie k pôvodným webovým zdrojom

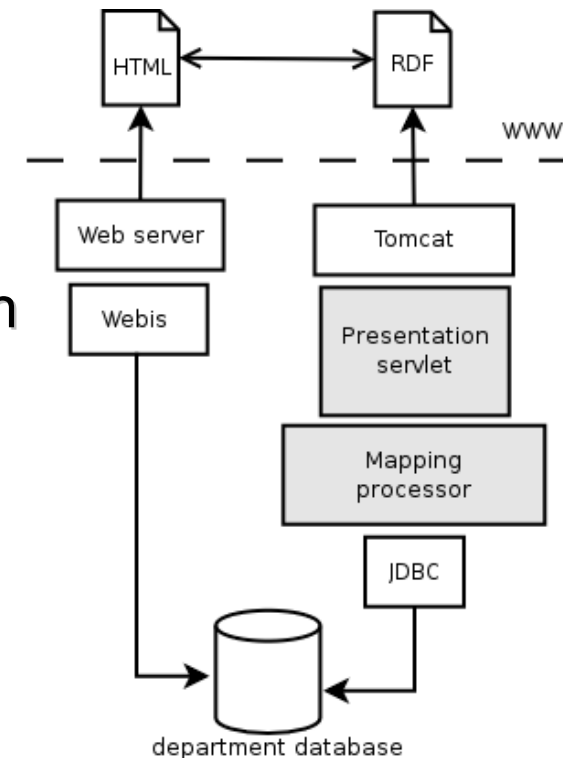
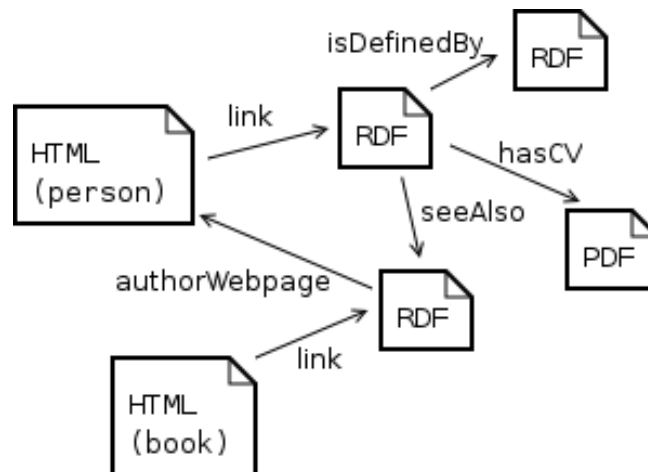
Životný cyklus informácie (2/2)

Strana konzumenta informácie:

- Zbieranie metadát z webu
- Vytváranie znalostnej bázy: analýza, indexovanie, predspracovanie a uloženie znalostí
- Použitie znalostí (sémantické vyhľadávanie)
- Konzument môže informácie ďalej publikovať

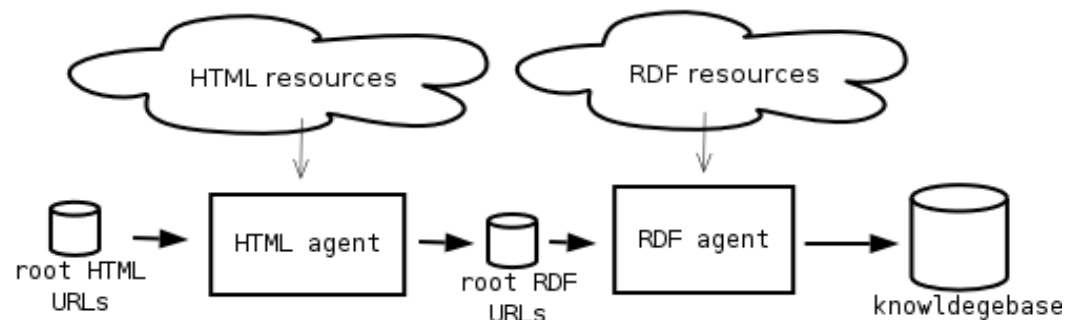
Anotácia webových zdrojov

- RDF metadáta sa generujú z rovnakej databázy ako HTML
- RDF a HTML majú rovnaký informačný obsah
- Metadáta sú spojené s príslušnými HTML stránkami



Zbieranie metadát

- HTML agent
 - Prechádza HTML dokumenty
 - Zbiera odkazy na RDF dokumenty
 - HTML sa neukladá
- RDF agent
 - Používa odkazy zozbierané HTML agentom
 - Prechádza RDF dokumenty
 - Ukladá RDF



DATAKON 2005

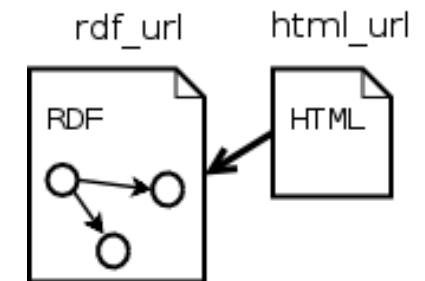
Životný cyklus informácie na sémantickom
webe

(7)

Vytváranie indexu znalostnej bázy

- **Informačný obsah domény – RDF**

- Každé RDF vyhlásenie je označené prostredníctvom URL dokumentu, z ktorého pochádza



- **Štruktúra domény**

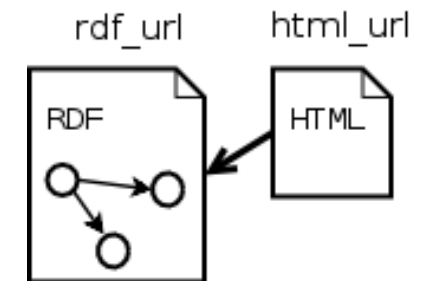
(odkazy medzi dokumentami)

- Množina usporiadaných dvojíc
[rdf_url, html_url]

Vytváranie indexu znalostnej bázy

- **Informačný obsah domény – RDF**

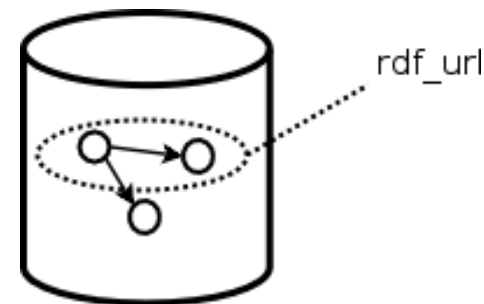
- Každé RDF vyhlásenie je označené prostredníctvom URL dokumentu, z ktorého pochádza



- **Štruktúra domény**

(odkazy medzi dokumentami)

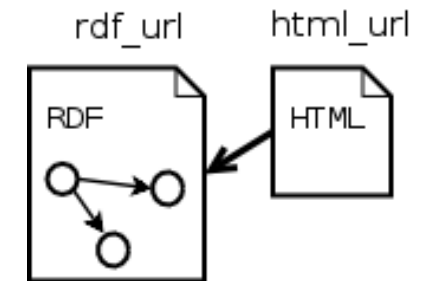
- Množina usporiadaných dvojíc [rdf_url, html_url]



Vytváranie indexu znalostnej bázy

- **Informačný obsah domény – RDF**

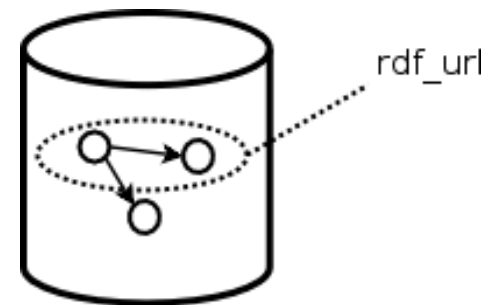
- Každé RDF vyhlásenie je označené prostredníctvom URL dokumentu, z ktorého pochádza



- **Štruktúra domény**

(odkazy medzi dokumentami)

- Množina usporiadaných dvojíc
[rdf_url, html_url]



[rdf_url, html_url]

Vytváranie indexu znalostnej bázy

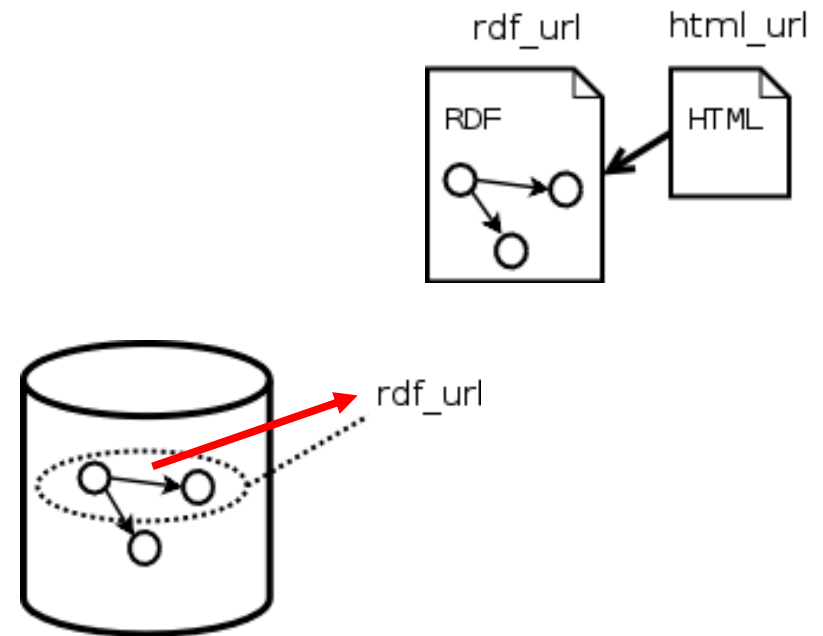
- **Informačný obsah domény – RDF**

- Každé RDF vyhlásenie je označené prostredníctvom URL dokumentu, z ktorého pochádza

- **Štruktúra domény**

(odkazy medzi dokumentami)

- Množina usporiadaných dvojíc
[rdf_url, html_url]



[rdf_url, html_url]

Vytváranie indexu znalostnej bázy

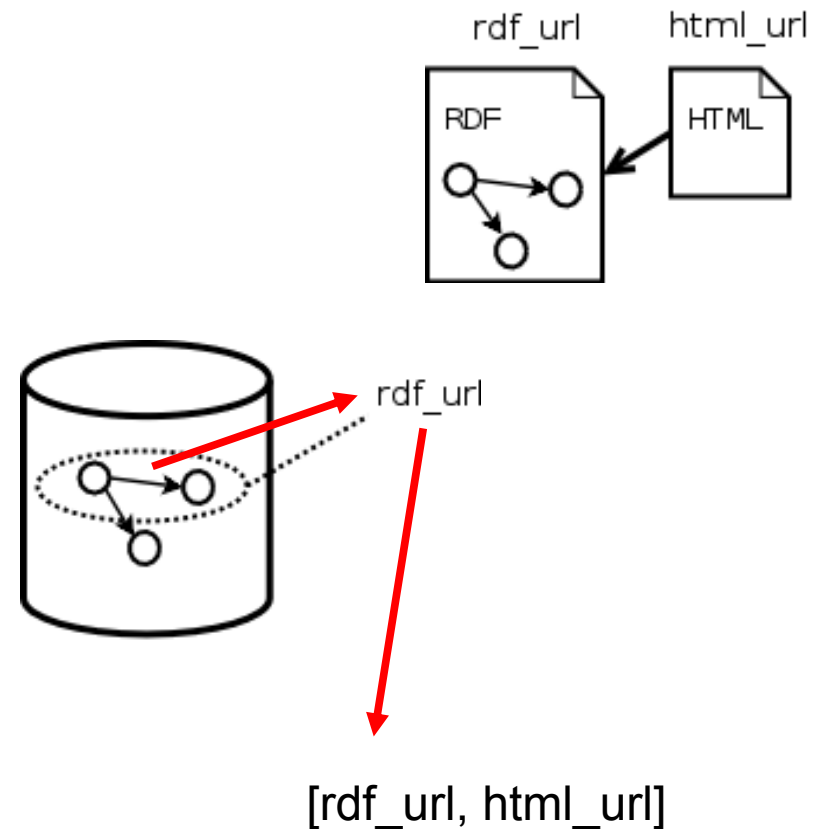
- **Informačný obsah domény – RDF**

- Každé RDF vyhlásenie je označené prostredníctvom URL dokumentu, z ktorého pochádza

- **Štruktúra domény**

(odkazy medzi dokumentami)

- Množina usporiadaných dvojíc [rdf_url, html_url]



Vytváranie indexu znalostnej bázy

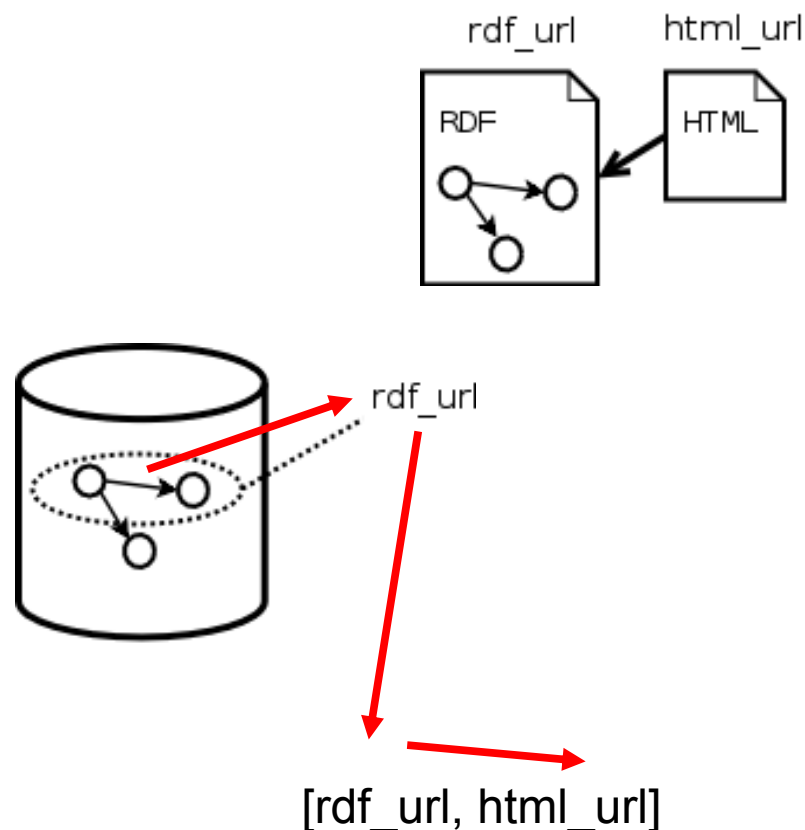
- **Informačný obsah domény – RDF**

- Každé RDF vyhlásenie je označené prostredníctvom URL dokumentu, z ktorého pochádza

- **Štruktúra domény**

(odkazy medzi dokumentami)

- Množina usporiadaných dvojíc [rdf_url, html_url]



Sémantické vyhľadávanie (1/2)

- Znalostná báza je dotazovaná trojicou: [podmet, prísudok, predmet]
 - napr.: [foaf:person, foaf:firstName, "Martin"]

Search query

<input type="text" value="http://semwebis.felk.cvut.cz/CTUOntology0.1#Person"/>	<input type="text" value="http://xmlns.com/foaf/0.1/firstName"/>	<input type="text" value="Martin"/>	<input type="button" value="Search!"/>
---	--	-------------------------------------	--

```
@prefix ql: <http://www.w3.org/2004/12/ql#> .
<> ql:select { (?s ?p ?o) . };
ql:where {
  ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://semwebis.felk.cvut.cz/CTUOntology0.1#Person>;
  <http://xmlns.com/foaf/0.1/firstName> "Martin" .
  ?s ?p ?o.
}.
```

- 1. výsledkom je množina RDF vyhlásení...
-

Sémantické vyhľadávanie (2/2)

- 2. Pre každé vyhlásenie nájdeme URL jeho RDF dokumentu...

@prefix : <http://sw.deri.org/2004/06/yars#> .

@prefix ql: <http://www.w3.org/2004/12/ql#> .

<> ql:select {

 (?c) .

}; ql:where {

 { <http://webing.felk.cvut.cz/sewebis/?username=svihlm1>

<http://xmlns.com/foaf/0.1/firstName> "Martin". } :context ?c .

}.

- 3. Pre každé RDF nájdeme URL príslušného HTML dokumentu...

Záver

- Implementované všetky kroky životného cyklu informácie na sémantickom webe
- Mnoho zjednodušení
- Široký záber projektu – povrchnosť
- Prototyp vyhľadávača je vo veľmi skorej fáze vývoja

Pozorovania:

- Technológia (RDF, OWL) pre tento jednoduchý príklad postačuje
- Vývojárske nástroje nie sú väčšinou pripravené na skutočné nasadenie

Možné smery výskumu

Široký záber projektu odhalil možné smery výskumu:

- Inteligentní agenti
- Indexovanie znalostnej bázy
- UI pre komplexnejšie dotazy, iné dotazovacie jazyky
- Vyhľadávanie v heterogénnych znalostných doménach

...d'akujem za pozornosť

{svihlm1, jelinek}@fel.cvut.cz

DATAKON 2005

Životný cyklus informácie na sémantickom
webe

(18)

References

Protege: <http://protege.stanford.edu/>

METAmorphoses: <http://www.svihla.net/metamorphoses/>

YARS (Yet Another RDF Store): <http://sw.deri.org/2004/06/yars/>

Jena2: <http://jena.sourceforge.net>

WebSphinx: <http://www.cs.cmu.edu/~rcm/websphinx/>