

Ontologie

⇒ Moodle

- Jako pojmový model
- Jako slovník
- Jako nástroj pro organizaci znalostí v sémantickém webu



ÚISK – KSA T08

1

1

Informatická ontologie je:

- Thomas R. Gruber (1959) „Specifikace konceptualizace“
- Thomas R. Gruber + Willem Nico Borst “Formální explicitní specifikace sdílené konceptualizace skutečnosti”
- Sdílitelná a opakovatelně použitelná pojmová reprezentace vymezené domény (univerza diskursu)



ÚISK – KSA T08

2

2

Oblasti využití a uživatelé ontologií

- Komunikace znalostí
- Opakované použití a organizace znalostí
- Popis skutečnosti pro počítačové zpracování/podporu
- Automatické odvozování (inference) nových znalostí

■ Lidé

■ Software (agenti)

ÚISK – KSA T08

3

3

Komponenty ontologií

Ontologie = pojmy (konceptualizace) + axiomy +
+ ontologické závazky + software

- 1) Třída
- 2) Instance / individuum
- 3) Vlastnost třídy
- 4) Vztah
- 5) Axiom (výrok, tvrzení)
- 6) Ontologický závazek (pravidlo)

Znalostní báze

Zachycuje celý „svět“ poznání o dané oblasti: jeho prvky, vztahy mezi nimi, a pravidla fungování

ÚISK – KSA T08

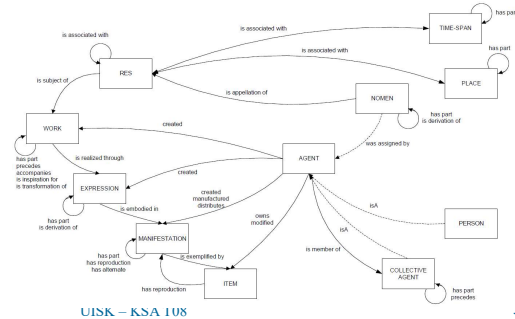
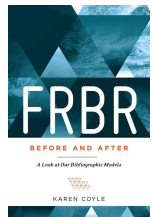
4

4

Komponenty ontologií

- 1) Třída
- 2) Instance / individuum

IFLA LRM



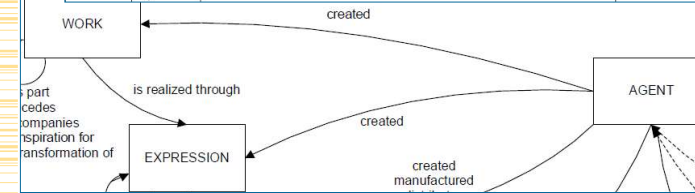
5

Komponenty ontologií

- 3) Vlastnost třídy – např. název

RDA/MARC21

245		Údaje o názvu	
	\$a	název	p
	\$b	další údaje o názvu	a
	\$c	údaj o odpovědnosti atd.	a
	\$n	číslo části/sekce díla	a
	\$p	název části/sekce díla	a



- 4) Vztah – např. vztah vytvoření

UISK – KSA T08

IFLA LRM

6

6

Specifikace požadavků na ontologii

Stanoví požadavky na **funkcionalitu / použitelnost** budoucího systému z pohledu **uživatele**.

Účel vytvořené ontologie:
pomůcka pro **organizaci zaznamenaných znalostí** (tj. dokumentů, informačních zdrojů) v rámci zvolené domény.

Metodika návrhu: 5W1H

- Proč se bude organizovat?
- Co se bude organizovat?
- Jak? Kdy? Kdo / co?

Úkol 9 **PROČ? CO?**
Cvičení 20, 21 **JAK?**

UISK – KSA T08

7

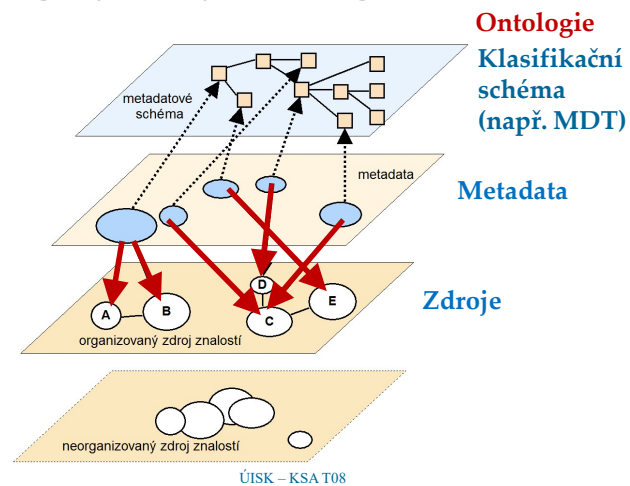
7

Specifikace požadavků na ontologii

	Funkce	Jednotka organizace
a) Systém organizace znalostí	indexace zdrojů a vícehlediskový přístup k nim	informační zdroje
b) Znalostní báze	poznání	zaznamenané znalosti a související výroky
c) Výkladový slovník	kategorizace hesel a vícehlediskový přístup k nim	slovníková hesla
d) Oborová encyklopedie		články /encyklopedická hesla

8

Ontologie jako systém organizace znalostí



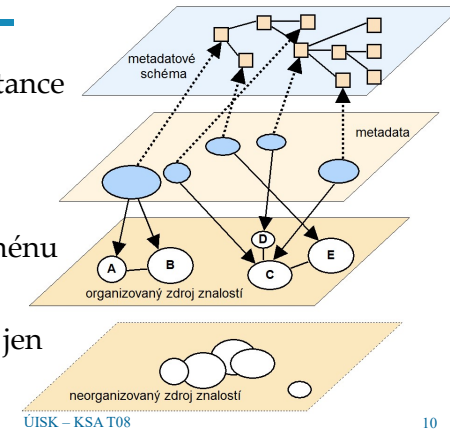
ÚISK – KSA T08

9

9

Projekt ontologie

- 1) V ontologii jsou třídy, nikoli instance
- 2) Pojmy mají v úplnosti zachycovat doménu
- 3) Každý pojem je jen v jedné fasetě



ÚISK – KSA T08

10

10

Sémantický web (web 3.0, web dat)



⇒ Moodle

znamená:

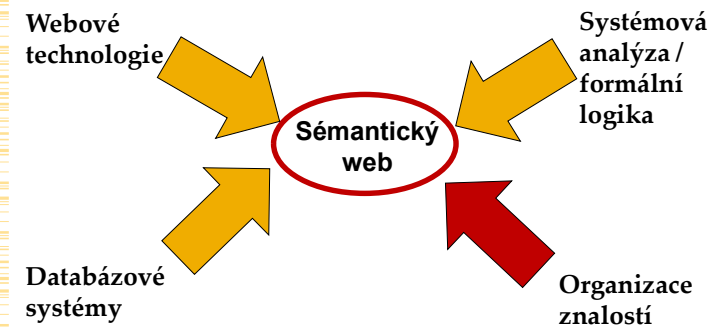
- 1) **Sémantická metadata** srozumitelná strojům
 - web 3.0, „chytrá data“
 - od informací ke znalostem
 - od klasifikačních schémat k ontologiím
- 2) **Propojená data**
 - web dat, jemnější granularita
 - LOD (*linked open data*)
 - od dokumentů k datům
(= *výroky predikátové logiky*)
- 3) **OWA, AAA, NUNA**

ÚISK – KSA T08

11

11

Zdroje pro sémantický web



ÚISK – KSA T08

12

12

Rekapitulace Systémová analýza – aplikace pro web

Třída

Prvek
⇒ Instance /
Individuum

Atribut
⇒ Predikát

Hodnota

1. **Prvek <je instancí> třídy.**
Třída/kategorie je množina prvků se stejnými vlastnostmi.
2. **Prvek/třída/kategorie <má> atribut.**
Atribut je specifikací prvku nebo třídy.
3. **Atribut <má> hodnotu.**
Třídy ani prvky nemají hodnoty.
4. **Vztah <je> atribut** patřící více prvkům nebo třídám.
5. **Metadata**
= atributy/prvky (dokumentů) + hodnoty
6. **Atribut nebo třída/kategorie ⇒ fasety**

ÚISK – KSA T08

13

13

Webové technologie

3 stadia standardizace (unifikace) WWW:

- 1) protokol http + URI/IRI (**identifikace**)
⇒ **komunikace**
technická interoperabilita
- 2) HTML + RDF ⇒ **syntaxe**
syntaktická interoperabilita
- 3) sémantická metadata ⇒ **sémantika**
sémantická interoperabilita

ÚISK – KSA T08

14

14

Cíl – sémanticky označený (=indexovaný)
informační prostor ⇒ **obsahová** metadata



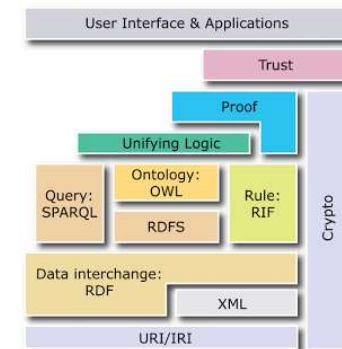
ÚISK – KSA T08

15

15

Technologie sémantického webu

- technické
- syntaktické
- sémantické



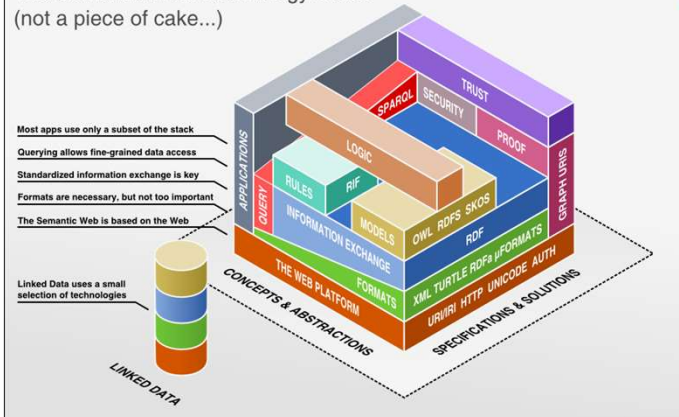
ÚISK – KSA T08

16

16

Technologie sémantického webu

The Semantic Web Technology Stack
(not a piece of cake...)



7

17

Nástroje sémantického webu = jazyky

- **RDF Resource Description Framework**
formát metadat
- **RDFS – RDF schema**
obsah metadat
- **SKOS – Simple knowledge organization system**
popis systémů organizace znalostí
- **OWL Web Ontology Language**
jazyk pro vyjádření obsahu ontologií

ÚISK – KSA T08

18

18

Nástroje sémantického webu = modely

- **IFLA LRM**
 - **BIBFRAME**
 - **CIDOC CRM**
 - **ISO 25964**
 - **Dublin Core**
- ...a další

ÚISK – KSA T08

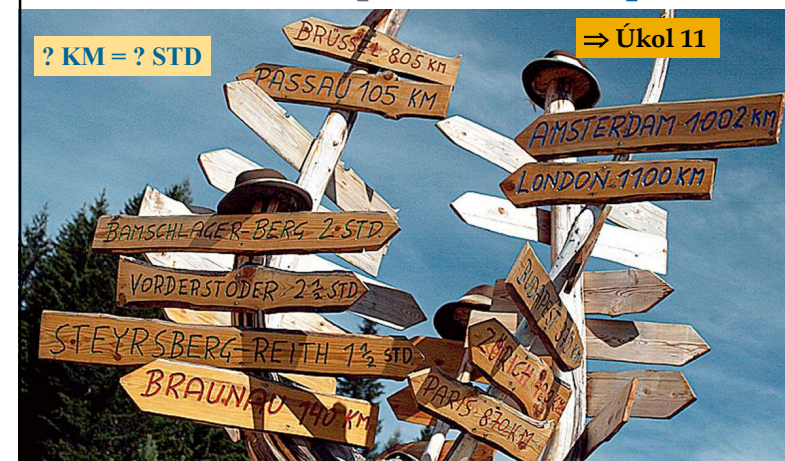
19

19

Problém: sémantická interoperabilita Řešení: mapování

? KM = ? STD

⇒ Úkol 11



20

Problémy / přednosti sémantického webu

...aneb stará témata s novými názvy

- OWA - open world assumption
- AAA - anyone can say anything about anything
- NUNA - non-unique naming assumption

21

Problém / přednost OWA

Tříprvková logika: pravda – nepravda – neznámé

Československá informatika: teorie a praxe. Praha: SNTL, 1971-1990.

Slovo a slovesnost: časopis pro otázky teorie a kultury jazyka. Praha: Melantrich, 1935- .

22

Problém / přednost AAA

Faseta je kategorie entit vytvořená podle jedné klasifikační charakteristiky (principium divisionis).

Faseta je specifická a samostatná množina pojmů, patřících k jedné základní kategorii, jež umožňuje organizovat obsah podle kritéria členění daného těmito základními kategoriemi.

Faseta je kategorie entit (množina podtříd) vytvořená uplatněním jedné klasifikační charakteristiky (principium divisionis), která je pro danou kategorii (třídu) podstatná, strukturální. Fasety vyjadřují vlastnosti použité pro seskupování pojmů podle jejich podstaty.

23

Problém / přednost NUNA

VIAF
Virtual International Authority File

Search

Select Field: All Headings Select Index: All VIAF Search Terms:

The screenshot shows a search result for Alexander Pushkin in the VIAF database. It displays a list of different name forms and dates associated with the entity, such as "Pushkin, Aleksandr Sergeevich, 1799-1837", "Пушкин, Александр Сергеевич, 1799-1837", and "Rouchkine, Alexandre Sergueievitch, 1799-1837". Each entry includes small icons for different languages and a VIAF ID.

VIAF ID: 66477450 (Personal)
Permalink: <http://viaf.org/viaf/66477450>

24

Současnost WWW – makroúroveň

„web of documents“, identifikované / propojené dokumenty

- narativní dokumenty – texty
- sémanticky nestrukturovaná, kompaktní data, jimž rozumí jen lidé

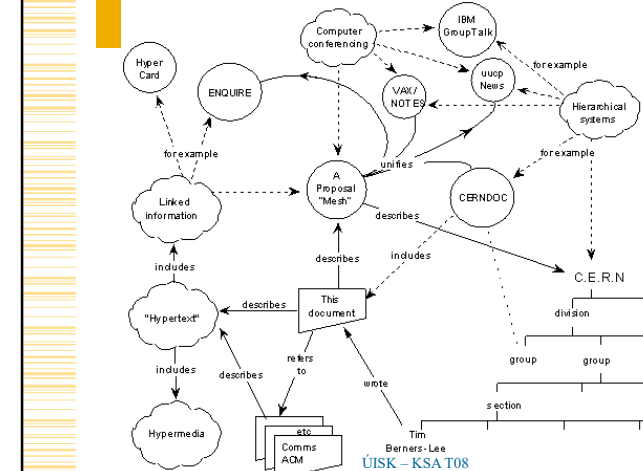
Technologie přístupu: plnotextové vyhledávání (Google)

ÚISK – KSA T08

25

25

Makroúroveň – web dokumentů (1989)



ÚISK – KSA T08

26

26

Mikroúroveň – web dat

„web of data“, identifikovaná / propojená data, „chytrá“ data

- transakční / strukturované dokumenty – záznamy
- sémanticky strukturovaná data, jejichž obsahu rozumí počítačové programy (software agent)

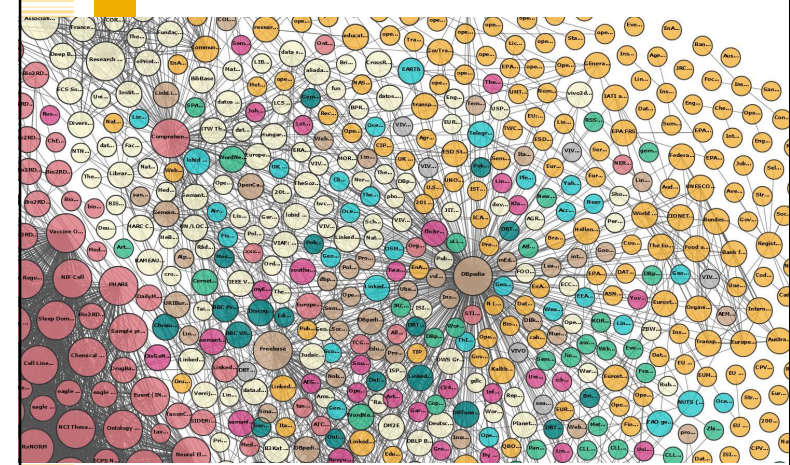
Technologie přístupu: obsahová/pojmová analýza ⇒ strukturovaná sémantická metadata (Diagram znalostí Google)

ÚISK – KSA T08

27

27

LOD – web dat



28

LOD - propojená otevřená data

- ★ data jsou dostupná na webu pod otevřenou licencí
- ★ data jsou poskytována ve strojově čitelném strukturovaném formátu
- ★ formát dat je otevřený
- ★ jako identifikátory objektů jsou použity URI
- ★ data jsou pomocí odkazů propojena na jiná související data

Ontologie jako metadatové schéma Příklad: tvorba znalostního grafu

Shiyali Ramamrita Ranganathan je autorem díla *Prolegomena to library classification* (1967) a tvůrcem Dvojtečkové klasifikace, založené na principu faset. Na jeho myšlenky navazuje CRG (Classification Research Group), která byla ustavena v roce 1952 a věnovala se zkoumání fasetové analýzy. Henry Evelyn Bliss (1870–1955) je tvůrcem plně fasetové Bibliografické klasifikace. 2. revidované vydání Blissovy Bibliografické klasifikace (BC2) připravuje od roku 1970 Bliss Classification Association (BCA) z podnětu Jacka Millse, člena CRG.

Obsahová / pojmová analýza = uplatnění systémové analýzy na obsah

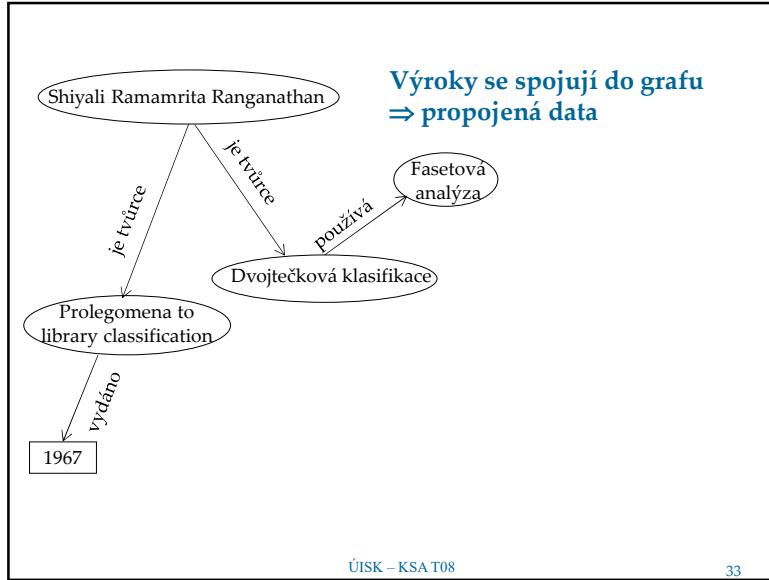
Jednotky [prvek] (atribut) hodnota
znalostí: [prvek] (vztah) [prvek]

[Shiyali Ramamrita Ranganathan] je (autorem) díla [Prolegomena to library classification] (1967) a (tvůrcem) [Dvojtečkové klasifikace], (založené na) principu [faset]. Na jeho myšlenky (navazuje) [CRG (Classification Research Group)], která (byla ustavena) v roce 1952 a věnovala se (zkoumání) [fasetové analýzy]. [Henry Evelyn Bliss] (1870–1955) je (tvůrcem) plně fasetové [Bibliografické klasifikace (BC1)]. 2. revidované vydání Blissovy Bibliografické klasifikace [(BC2)] (připravuje) od roku 1970 [Bliss Classification Association (BCA)] (z podnětu) [Jacka Millse], (člena) [CRG].

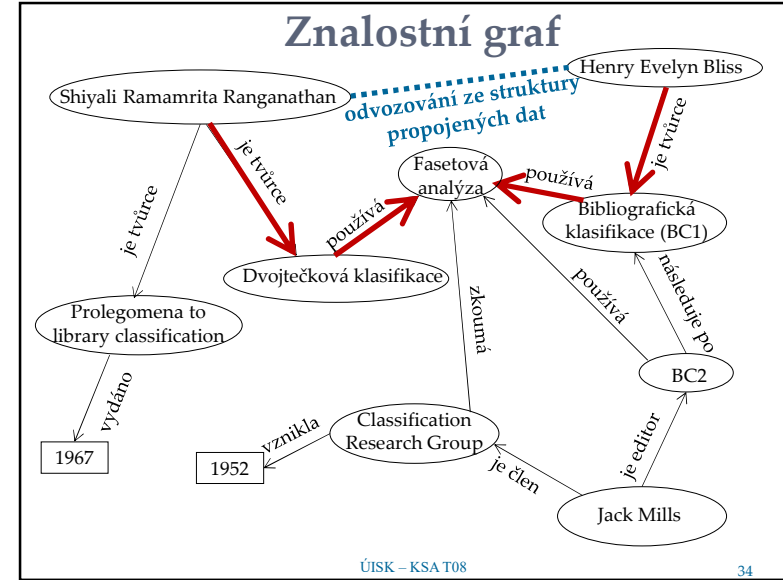
Výroky



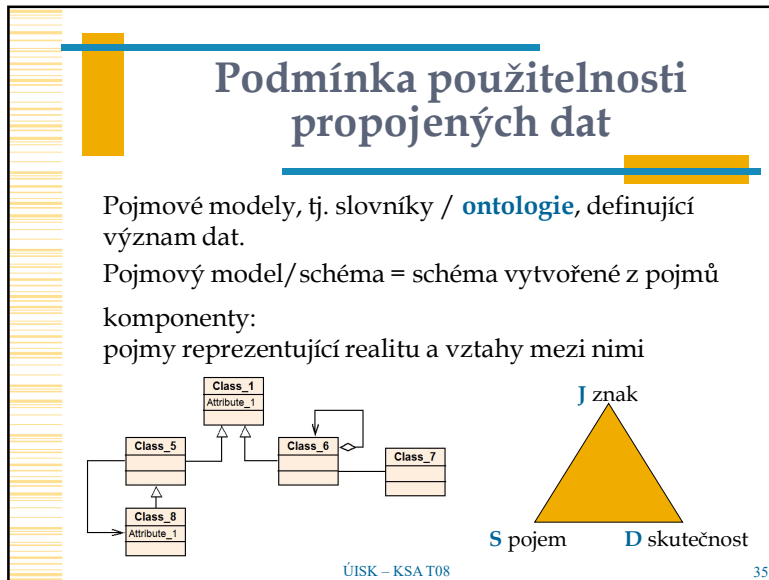
1. [Shiyali Ramamrita Ranganathan] (je tvůrce) [Prolegomena to library classification].
2. [Dvojtečková klasifikace] (používá) [fasetová analýza].
3. [Prolegomena to library classification] (bylo vydáno) v roce 1967.



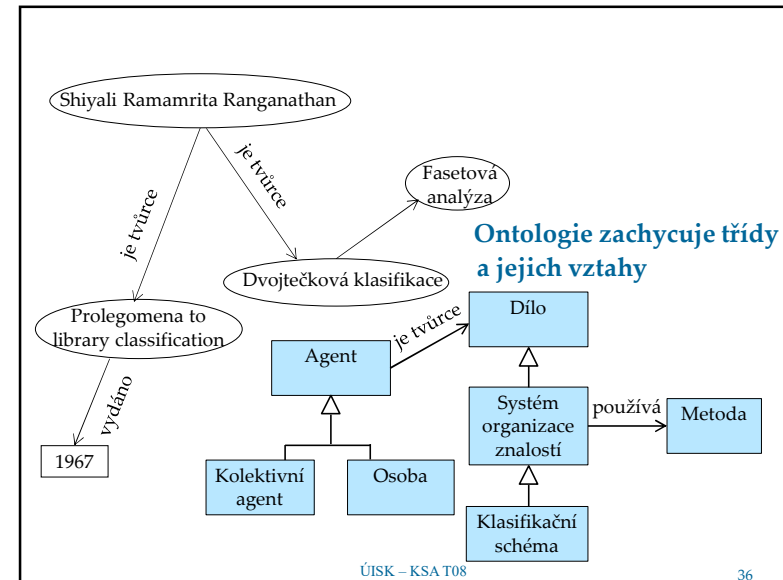
33



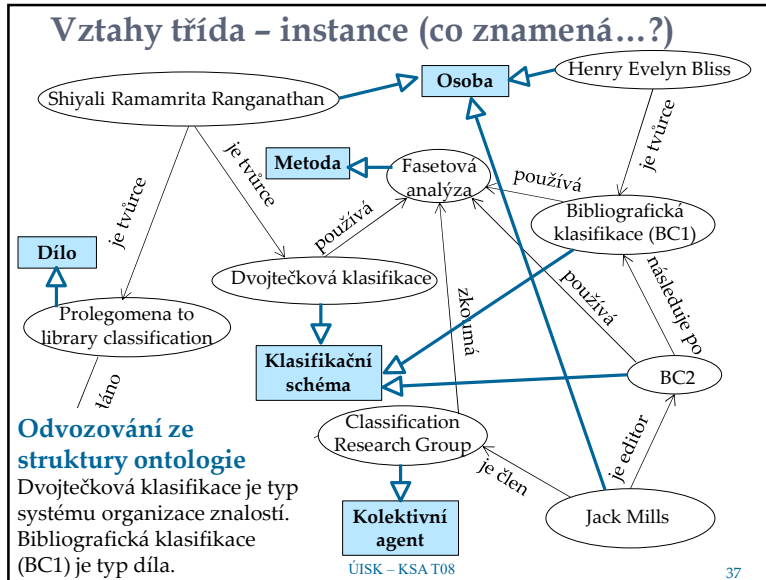
34



35



36



37

Funkce ontologií v sémantickém webu

= rozšíření funkcí řízeného slovníku

- 1) Vyšší stupeň **formalizace**
⇒ funkce jsou srozumitelné softwarovému agentu
- 2) Poskytují **kontext** potřebný pro pochopení smyslu zpracovávaných informačních objektů softwarovým agentem (tj. vztahy pojem–pojem / význam–význam, znak–pojem/význam) ⇒ možnost odvozování
- 3) **Obsah** / význam zahrnutých pojmů je určen prostřednictvím axiomů / ontologických závazků (tj. výroků srozumitelných softwarovému agentu) ⇒ zjednoznačnění

ÚISK - KSA T08 38

38

Zjednodušený pohled na ontologii v sémantickém webu

- 1) **Prvky** (⇒ subjekty/objekty)
- 2) **Vlastnosti** (⇒ predikáty)
 - a) binární = vztah subjekt–objekt
 - b) unární = vlastnosti/atributy

Výrok:

- a) říká něco o nějakých prvcích
- b) uvádí hodnotu vlastnosti nějakého prvku

ÚISK - KSA T08 39

39

Výroky v bibliografické znalostní bázi (instance tříd, hodnoty atributů)

⇒ Moodle

Hlavní záhlaví	● Hrabal, Bohumil, 1914-1997	Obalkyknih.cz:
Název	● Příliš hluchá samota : text / Bohumil Hrabal : [doslov Milan Jankovič]	
Vydání	1. vyd.	
Nakl. údaje	Praha : Odeon, 1989	
Popis (rozsah)	133 s. ; 21 cm	
EXEMPLÁŘE		
Číslo nár.bibl.	cnb000051861	
ISBN	80-207-0156-7 (váz.) : Kčs 16,00	
Další původce	● Jankovič, Milan, 1929-	
Poznámka	Vyšlo k autorovým 75. narozeninám 80000 výt.	
Předmět. heslo	● Hrabal, Bohumil, 1914-1997	
Forma, žánr	* studie	
Předmět. heslo	Hrabal, Bohumil (1914-1997 spisovatel čes.) - dílo - studie	
MDT	● 885 0-32	

40

RDF – Resource Description Framework

„RDF je jazyk, navržený lidmi pro vyjadřování lidských myšlenek ve formě přístupné strojovému zpracování.“ Thomas Baker, 2012

- aplikace XML – jazyk pro psaní metadat pro zdroje na webu
- „obálka“, „kontejner“ na jakákoli metadata
- **síťová (grafová) struktura: orientovaný ohodnocený graf**

ÚISK – KSA T08 41

41

Syntaxe RDF

Výrazy v jazyce RDF tvoří 3 typy elementů:

- **zdroj** (angl. *resource*), tj. prvek subjekt/objekt uzel grafu **“podmět / předmět”**
- **vlastnost / predikát** (angl. *property / predicate*) (binární) vztah subjekt-objekt nebo (unární) atribut hrana grafu **“přísudek”**
- **výrok** (angl. *statement*) trojice RDF (angl. *RDF triple*) výrok = zdroj + vlastnost + zdroj/hodnota výroky jsou vždy binární a orientované **“věta”**

ÚISK – KSA T08 42

42

RDF trojice (triple)

[prvek] (atribut) hodnota
[prvek] (vztah) [prvek]

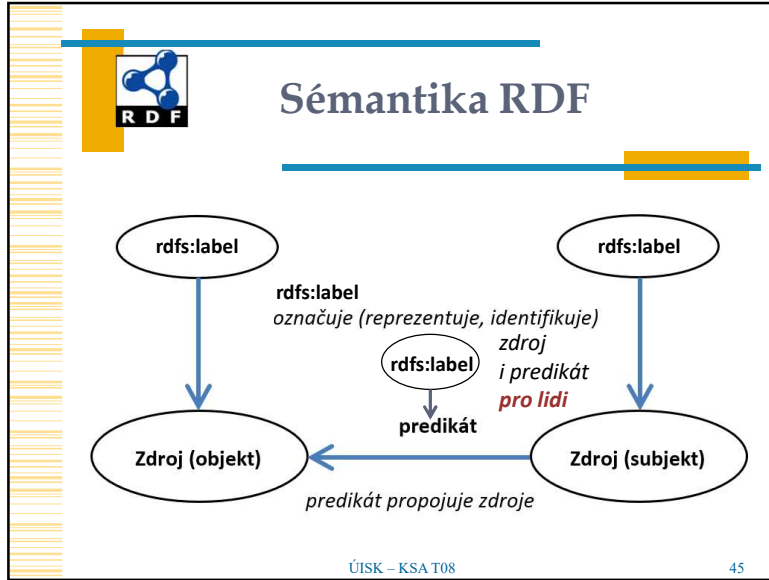
ÚISK – KSA T08 43

43

Sémantika RDF

ÚISK – KSA T08 44

44



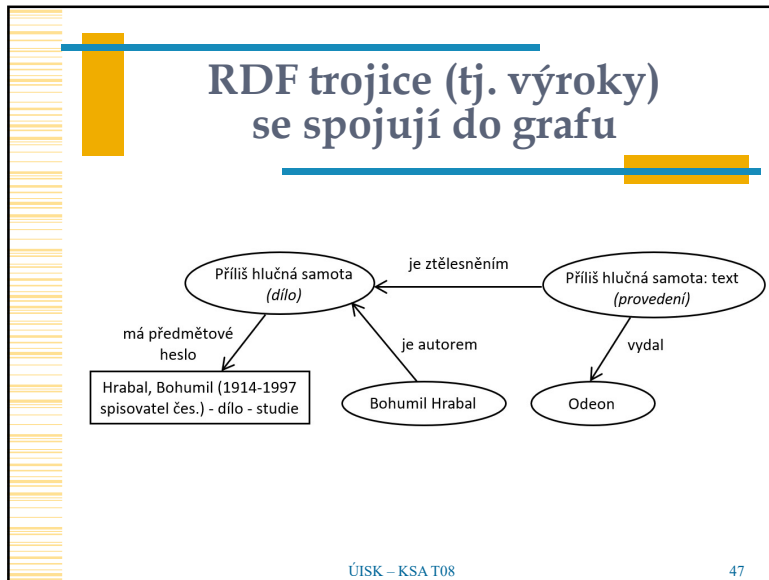
45

Hlavní záhlaví	Hrabal, Bohumil, 1914-1997
Název	Přiliš hluchá samota : text / Bohumil Hrabal : [doslov Milan Jankovič]
Vydání	1. vyd.
Nakl. údaje	Praha : Odeon, 1989
Popis (rozsah)	133 s. ; 21 cm
	Digitalizovaný dokument
EXEMPLÁŘE	
Číslo nár.bibl.	cnb000051861
ISBN	80-207-0156-7 (váz.) : Kčs 16,00
Další původce	Jankovič, Milan, 1929-
Poznámka	Vyšlo k autorovým 75. narozeninám 80000 výt.
Předmět. heslo	Hrabal, Bohumil, 1914-1997
Forma, žánr	studie
Předmět. heslo	Hrabal, Bohumil (1914-1997 spisovatel čes.) - dílo - studie
MDT	885.0-32

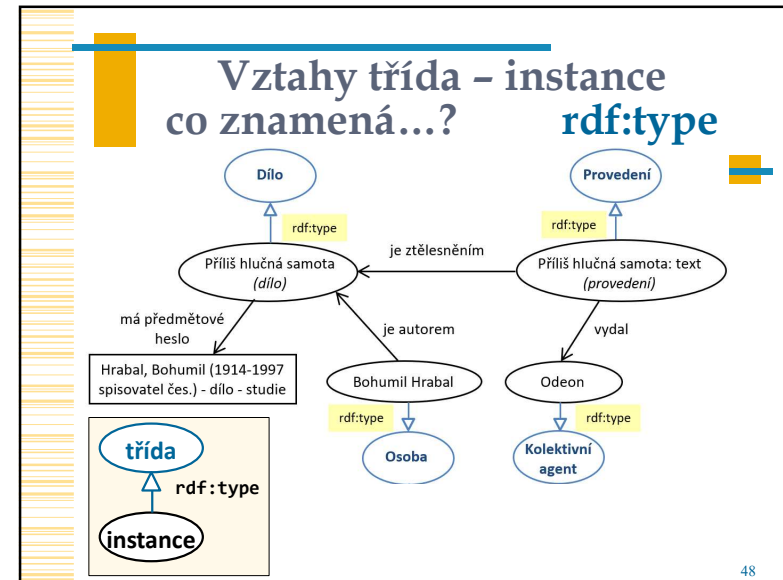
Obalkyknih.cz:

ÚISK – KSA T08 46

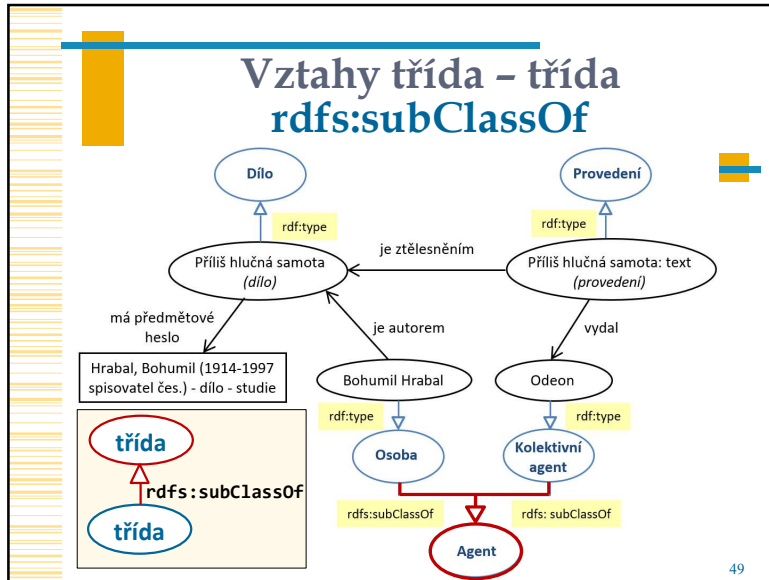
46



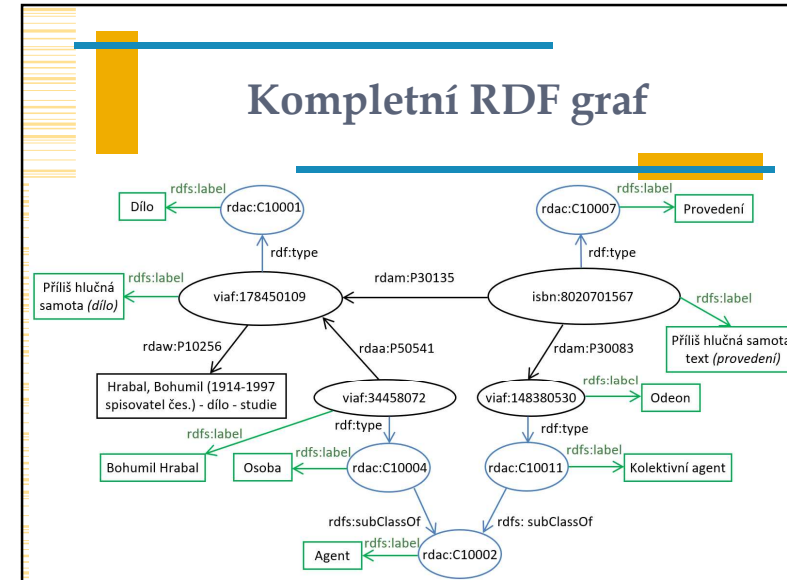
47



48



49



50