

Metody tvorby ontologií a sémantický web

Martin Malčík, Rostislav Miarka

Obsah

- Reprezentace znalostí
- Ontologie a sémantický web
- Tvorba ontologií

Hierarchie znalostí (D.R.Tobin)

- **Data** – jakékoliv znakové řetězce + relevance, zpracování, interpretace →
- **Informace** – data obohacená o jejich význam + aplikace, transformace, rozšíření →
- **Znalosti** – informace, které můžeme odvodit z jiných informací

Původní web

- V počátcích pouze „akademická záležitost“;
- 1962 – počítačová síť Arpa
- 1974 – vynález TCP/IP
- 1983 – spojení sítě a přenosového protokolu
- 1991 – Tim Berners Lee použil princip hypertextu Teda Nelsona (1976 Xanadu) a publikoval HTML – world wide web – syntaktický web
- 1998 – XML
- 2001 – Tim Berners Lee - sémantický web

Sémantický web

- Ideu Sémantického webu představil Tim Berners-Lee v roce 2001;
- Počet webových stránek na webu stále narůstá;
- Je stále složitější nalézt relevantní informace;
- Přejít od stávajícího webu k Sémantickému webu.

Cíl Sémantického webu

Cílem sémantického webu je zpřístupnění webových zdrojů automatickým procesům, a to prostřednictvím:

- rozšíření stávajících značkovacích jazyků o možnost zachycení sémantiky v metadatech popisujících obsah/funkce webových zdrojů;
- použití ontologií jako formálně specifikovaných slovníků tak, aby se formální specifikace staly automaticky přístupnými.
- Sémantický web je založen na:
 - RDF
 - URI

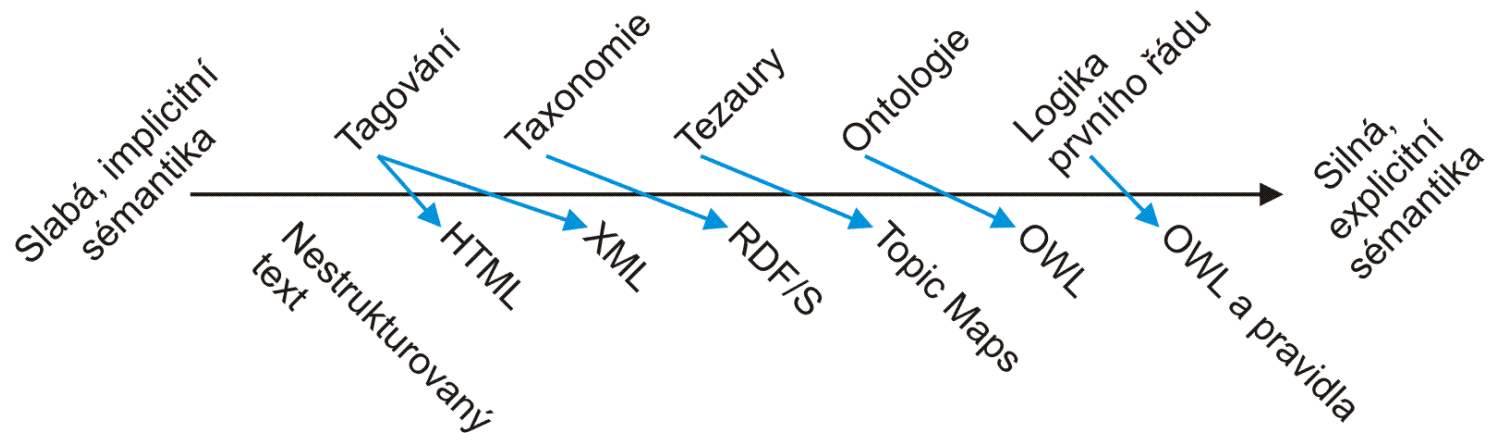
Syntaktický web -> sémantický web

Přechod od strukturálně propojených datových zdrojů na bázi klíčových slov k jejich propojení na bázi významových asociací;

Potřeba *sémanticky strukturovat* data na webu obrací pozornost ke konceptuální úrovni návrhů webových znalostníchází a aplikací

Sémantická struktura dat je reprezentována konceptuálním datovým modelem RDF.

Expresivita a sémantika webu



RDF

- Resource Description Framework;
- Obecný rámec pro popis zdrojů na Internetu;
- Jde o obecný rámec pro popis, výměnu a znovupoužití metadat;

RDF trojice **subjekt – predikát – objekt**

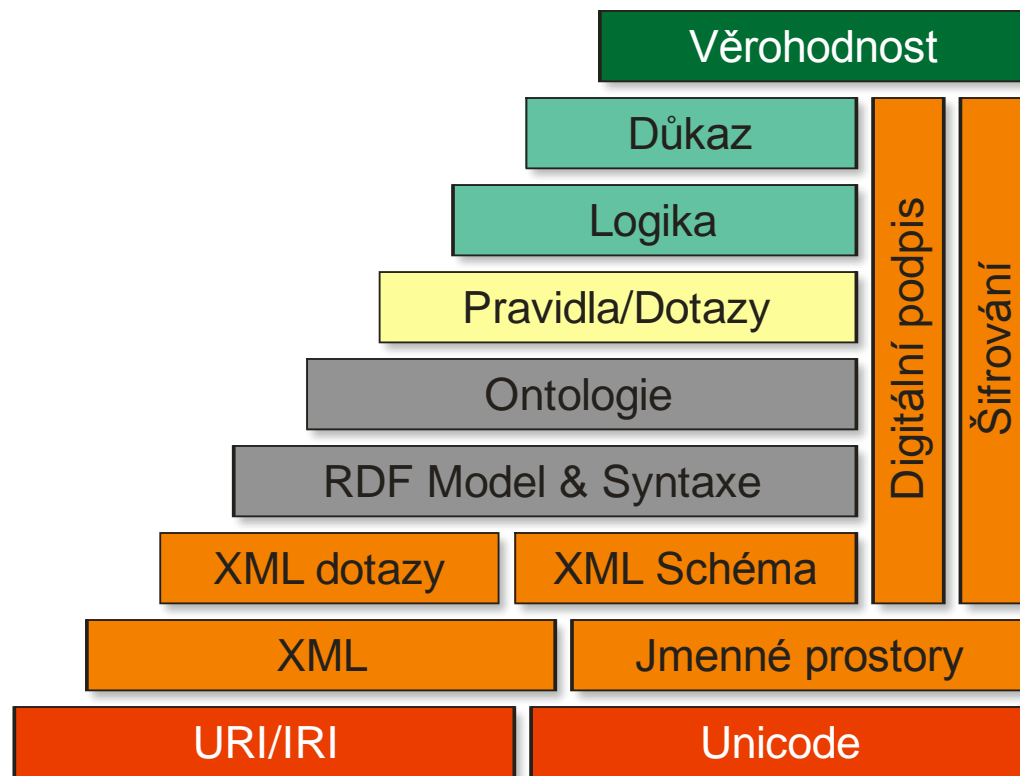
Subjekt má vlastnost určenou objektem;

- Grafová podoba – orientovaný graf
- Textová podoba – založena na jazyce XML

URI

- Uniform Resource Identifier
- URI identifikátory jsou použity k identifikaci zdrojů v RDF
- Jednoznačná identifikace zdrojů
- Podmnožinou URI jsou URL – adresy webových stránek nebo souborů na webu

Struktura sémantického webu



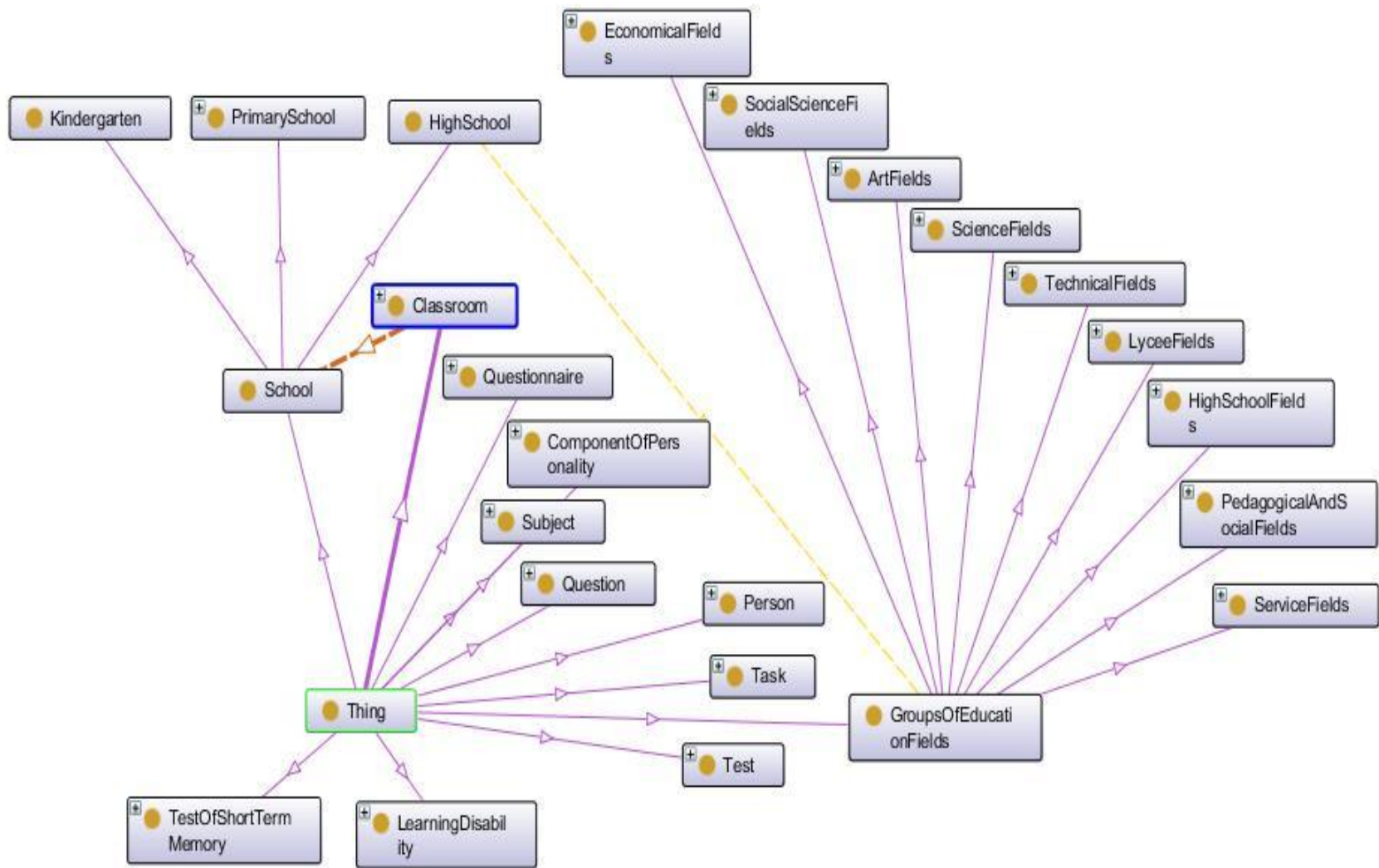
Obr. 1 Schéma vrstev sémantického webu (Zdroj: W3C. Přeložil MM)

Ontologie

- Filosofické pojetí bytí
- Ontologie lze charakterizovat jako formalizované reprezentace znalostí určené k jejich sdílení a znovupoužití.
- Domluvená sdílená terminologie pro určitou aplikační doménu
- *Formální ontologie* je explicitní formální reprezentace konceptualizace v určité doméně, založená na všeobecně sdílené významové struktuře a terminologickém slovníku domény pro danou oblast.

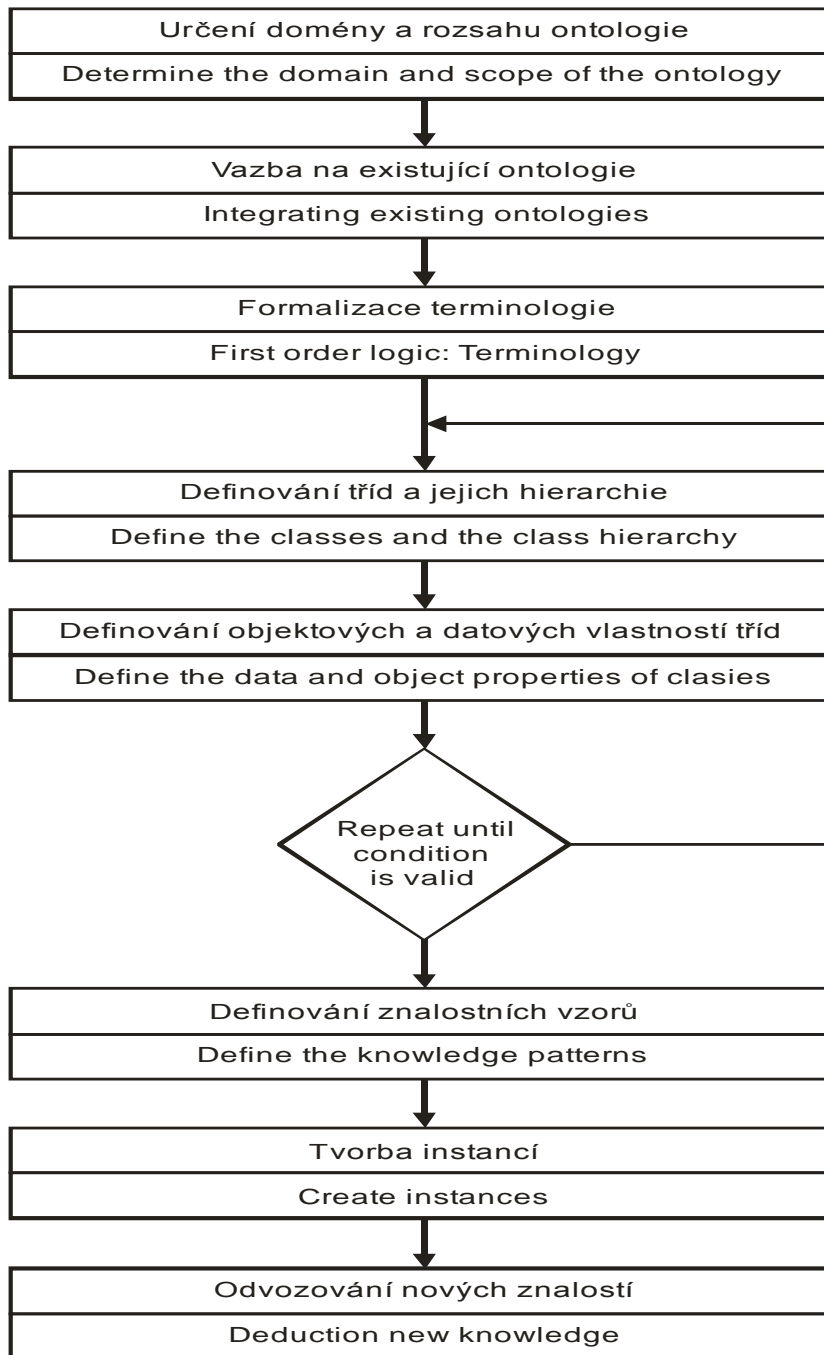
Ontologie

- *Ontologie může co do expresivity představovat:*
- *taxonomie (znalosti strukturované hierarchicky nebo genericky stylem předeek – potomek);*
- *tezaury (slova a synonyma),;*
- *konceptuální modely (se složitějším zachycením struktury znalostí) ;*
- *logické teorie (s velmi bohatými, složitými, konzistentními a smysluplnými znalostmi).*



Možnosti tvorby ontologií

- Automatické a poloautomatické ontologie – tvorba a extrakce s použitím různých druhů zdrojů a různých technologií.
 - ze souborů např. na webu
 - z databází – např. Wikipedie dbpedia <http://dbpedia.org>
- Expertní tvorba ontologií, kde je výzkum zaměřen na problémy buďto spojené s procesem návrhu ontologie nebo problémy specifické pro vzdělávací technologii.



Expertní tvorba ontologií I.

- Určení domény a rozsahu ontologie: Definování domény, oblastí použití a uživatelů ontologie.
- Vazba na existující ontologie.
- Formalizace terminologie: Termíny navržené k zařazení do ontologie musí být specifikovány s použitím predikátové logiky prvního řádu, případně v jazyce KIF (Knowledge Interchange Format). V rámci vymezeného univerza diskursu se definují objekty, které jsou reprezentovány prostřednictvím konstant a proměnných. Pro objekty se definují atributy jakožto unární predikáty a vztahy mezi nimi se definují pomocí n-árních (např. binárních) predikátů.
- Definování tříd a jejich hierarchie: V rámci námi zpracovávaných domén jsme postupovali postupem *top-down*. Ze seznamu termínů jsme vybrali termíny označující objekty se samostatnou existencí a tyto označili jako třídy. Mezi třídami jsme definovali generické hierarchické vztahy s uplatněním dědičnosti
- Definování vlastností tříd: Z termínů, které nebyly označeny jako třídy, se vyberou ty, jež představují vlastnosti některé z definovaných tříd. Rozlišují se typy vlastností: objektové popisující zejména vztahy mezi třídami a datové, popisující datový typ.

Expertní tvorba ontologií II.

- Definování znalostních vzorů: Znalostní vzory slouží ke znovupoužití znalostí. Ve vytvořené ontologii je možné nalézt koncepty a vztahy, které mají shodnou strukturu, ale liší se pouze ve svých názvech
- Tvorba instancí: Vytvoření individuálních instancí tříd a naplnění jejich vlastností hodnotami.
- Odvozování z ontologie: Ve výsledné ontologii je možné provádět dotazování a odvozování. Je možné vytvářet dotazy na třídy a na individua.

Děkuji za pozornost

Ukázka ontologie v Protégé