

Ontologie

Otakar Trunda

# [ Definice ]

- Mnoho různých definic:
  - Formální specifikace sdílené konceptualizace
  - Hierarchicky strukturovaná množina termínů popisujících určitou věcnou oblast
  - Strukturovaná slovní zásoba
  - Logická teorie v jazyce podporujícím „definování konceptů“
  - Grafová struktura nad pojmy

# [ Dělení ontologií ]

# Dělení podle oborových oblastí

- Terminologické ontologie
  - Tezaury
  - Strukturace termínů přirozeného jazyka
- Informační ontologie
  - Nadstavba nad datovými zdroji
  - Odpovídají DB schématům
- Znalostní ontologie
  - Reprezentace znalostí, logické odvozování

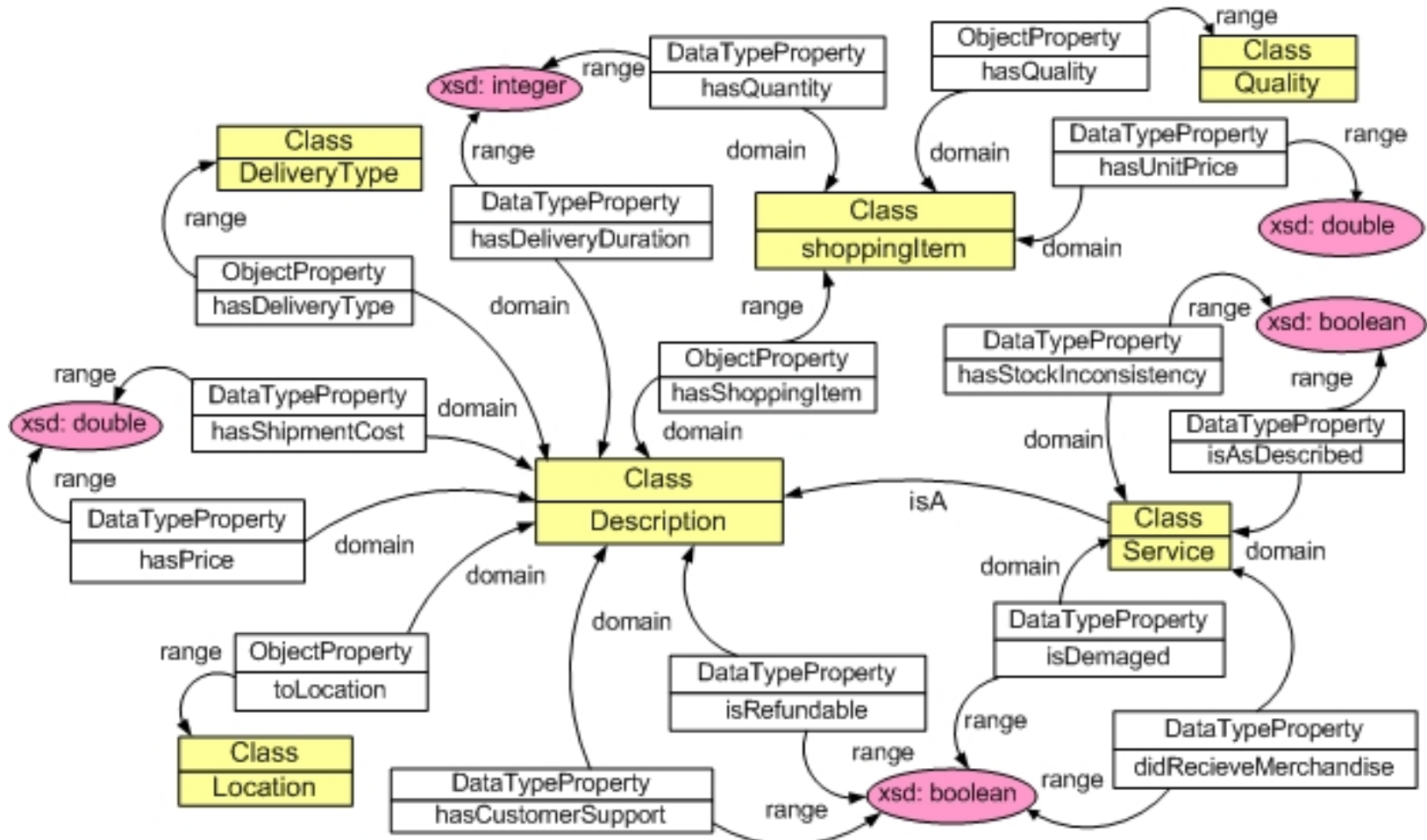
# Dělení podle předmětu formalizace

- Generické ontologie
  - Obecná problematika napříč věcnými oblastmi
- Doménové ontologie
  - Znalosti o určité doméně
- Úlohové ontologie
  - Způsob řešení určité úlohy
  - Diagnostika, plánování,...

# [ Struktura ontologie ]

- Třídy
- Individua
- Sloty – relace, funkce a atributy
- Facety – omezení na sloty
- Primitivní hodnoty a datové typy
- Axiomy a odvozovací pravidla

# Struktura ontologie - příklad



# [ Facety – omezení slotů ]

- Typická omezení slotů:
  - Definiční obor, obor hodnot
  - Požadované vlastnosti relací. Např.:  
Tranzitivita, symetrie, funkčnost
  - Vzájemně inverzní sloty
    - JeRodičem inverzní k JeDítětem
  - Dědičnost na slotech
    - JeRodičem IS-A JePředkem



# [ Axiomy a pravidla ]

- Zejména u ontologií založených na deskripční logice
  - Třídy a sloty definované implicitně
- Axiomy: (např.)
  - Ekvivalence/Subsumpce tříd a relací
  - Disjunktnost tříd
  - Rozklad na podtřídy atd.
- Pravidla
  - Dopředné a/nebo zpětné řetězení
  - Umožňují provádět odvozování

# Základní odvozovací úlohy

- *Kontrola konzistence* znalostní báze zjišťuje, zda definice každého konceptu připouští náležití alespoň jednoho individua.
- *Kontrola individuí* zjišťuje, zda dané individuum spadá pod daný koncept.
- *Realizace* nalezne nespecifičtější koncept, pod který individuum spadá.
- *Vyhledání* nalezne všechna individua spadající pod zadaný koncept.
- *Kontrola subsumpce* zjišťuje, zdali je jeden zadaný koncept podkonceptem druhého.

# [ Rozdíly oproti OOP ]

- Třídy pouze jako unární relace
- Individua x instance
- Sloty nezávislé na třídách
- Dědičnost
  - OOP: znovupoužití kódu
  - Ontologie: zachycení reality
- Příslušnost k třídám a slotům může být zadána implicitně
- Relace na slotech

# Požadované vlastnosti ontologie

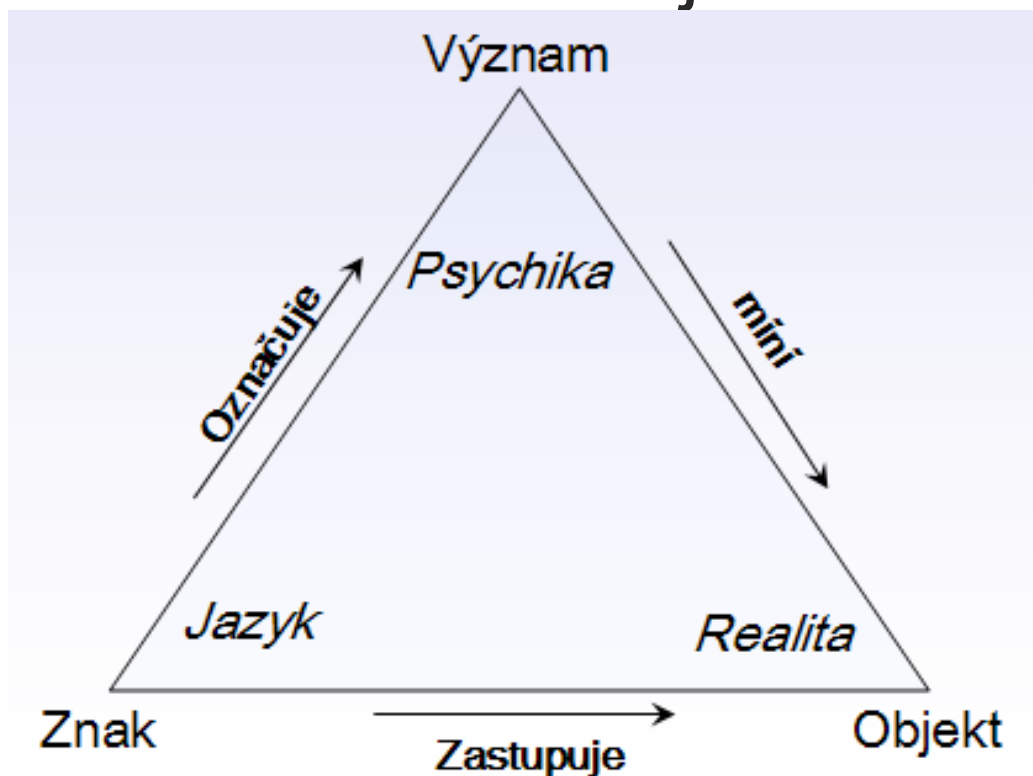
- Odpovídající realitě
- Konsensuální, obecně přijímaná
- Srozumitelná člověku, dále sdílitelná
- Znovupoužitelná, a to i nepředjímaným způsobem
- Bohatě strukturovaná
- Formálně popsána, strojově zpracovatelná
- Logicky konzistentní

# [ Využití ontologií ]

- Společné porozumění struktuře informací mezi lidmi nebo softwarovými agenty
  - Nejčastější využití. Sémantický web
- Znovupoužití doménové znalosti
  - Např. reprezentace času
- Explicitní popis předpokladů o doméně
- Oddělení znalosti o doméně od manipulace se znalostmi
- Analýza domény

# [ Odbočka - filozofická ontologie ]

- Ontologie - věda o jsoučnu, o tom co je
- Ogden-Richardsův trojúhelník



# [ Odbočka - filozofická ontologie ]

- Ontologie v informatice někdy považované za „aplikovanou filozofii“
- Témata tradiční (filozofické) ontologie:
  - Esenciální / neesenciální vlastnosti
  - Problém změny
    - (kdy je změna entity taková, že je nutno ji reklasifikovat)
  - Problém identity
    - (kdy různé „identifikátory“ odkazují k téže entitě)
    - Leibnizův zákon identity
  - Problém obecnin

# [ Odbočka - filozofická ontologie ]

## ■ Vlastnosti objektů:

- *Esenciální* - to, co dělá danou věc, tím čím je. (Například člověka člověkem, strom stromem atd.)
- Jednotliví lidé se navzájem liší *neesenciálními* vlastnostmi (věk, pohlaví, výška, váha, barva atd.)

## ■ Leibnizův zákon identity:

- Zákon stejnosti identických

$$\forall x \forall y [(x = y) \rightarrow (\forall P (P(x) \leftrightarrow P(y)))]$$

- Zákon identity stejných

$$\forall x \forall y [(\forall P (P(x) \leftrightarrow P(y))) \rightarrow (x = y)]$$



# [ Sémantický web ]

---

# Struktura současného webu

- Obrovské množství informací a služeb
- Stávající způsob vyhledávání je nedostatečný
- Informace dostupné pouze pro člověka, strojové zpracování obtížné (lingvistické aspekty apod.)
- Webové služby také určené primárně pro člověka, jejich využití agenty je problematické (standardizovaný formát)

# Textové vyhledávání na webu

Co vyhledávače sledují	Jak je to v HTML	Důležitost (váha)
titulek	<code>&lt;title&gt;text titulku&lt;/title&gt;</code>	obrovská
klíčová slova	<code>&lt;meta name=keywords content="slovo, slovo"&gt;</code>	značná, někdy žádná (Google)
popis (description)	<code>&lt;meta name=description content="Stručný popis"&gt;</code>	různá
nadpis 1. úrovně	<code>&lt;H1&gt;Nadpis&lt;/h1&gt;</code>	značná
ostatní nadpisy	<code>&lt;Hn&gt;Nadpis&lt;/hn&gt;</code>	sporná
začátek stránky	<code>&lt;body&gt;Několik prvních slov ...</code>	větší než malá
adresa URL	jméno souboru včetně cesty	různá
text odkazů mířících na tu stránku z jiného serveru	<code>&lt;a href="adresa"&gt;text odkazu&lt;/a&gt;</code>	u některých vyhledávačů obrovská (Google, Jyxo)
alty u obrázků	<code>&lt;img alt="zástupný text" ...&gt;</code>	malá
text stránky	prostě text	malá
katalogový popisek	co zadáte do případného spřízněného katalogu	různá (pouze na Seznamu)

# [ Sémantický web ]

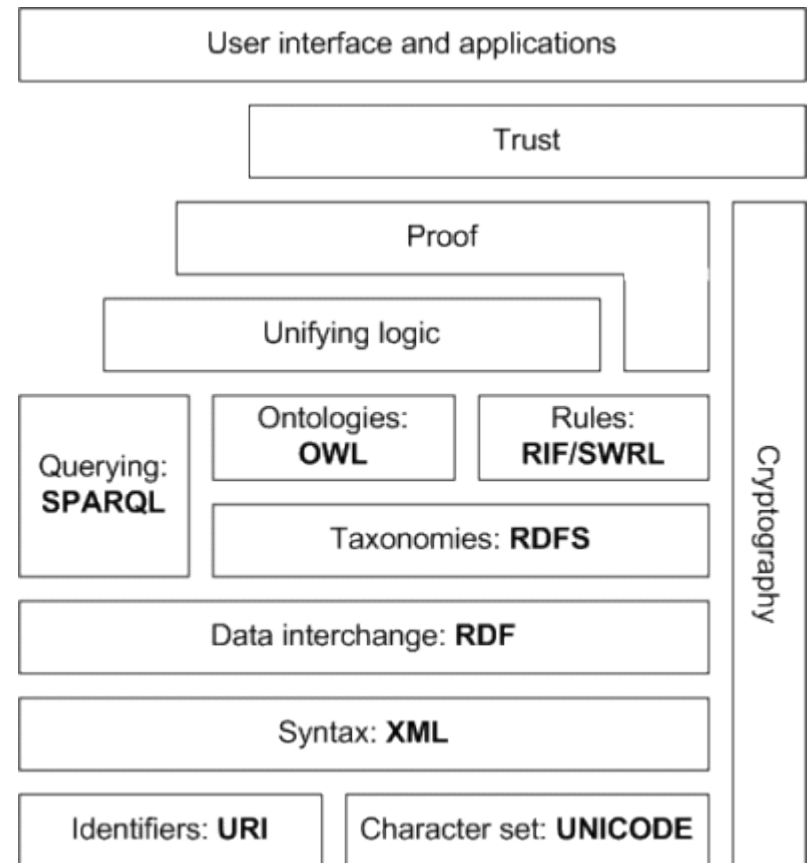
- Idea: informace srozumitelné jak pro člověka, tak pro stroje
- Aplikace by měly „rozumět“ informacím na webu, umět je vyhledávat, zpracovávat a použít
- Je třeba formálně definovat sémantiku informací pro strojové zpracování
- Zde se dají využít (doménové) ontologie

# Sémantický web - realizace

- Potřebné nástroje:
  - Jazyk pro popis ontologií
  - Jazyk pro popis sémantiky webových dat (na základě příslušné ontologie)
  - Integrace tohoto jazyka se stávajícími protokoly (html)
- Databáze standardních ontologií

# Sémantický web - realizace

- Popis ontologií: jazyk OWL
- Popis sémantiky dat: jazyk RDF
- Integrace RDF do HTML: pomocí metadat a formátu XML



# [ Tvorba ontologií ]

- Editor Protégé
- Tvorba ontologie je obecně iterativní proces, neexistuje jeden správný postup
- Různé metodiky, návody, seznam častých chyb, apod.
- Typická posloupnost kroků:
  - Ujasnění účelu a rozsahu ontologie
  - Specifikace terminologické části
  - Odlišení ontologických typů
  - Specifikace taxonomie
  - Vytvoření netaxonomických relací, atributů a instancí
  - Specifikace pokročilých axiomů
  - Nasazení a údržba ontologie

# [ Ujasnění účelu a rozsahu ]

- Formulace obecných scénářů a případů užití
- Konkrétní vzorové **kompetenční otázky**
  - Co by mělo být možné s pomocí ontologie zodpovědět?
- Vyvarovat se bezbřehosti (tzv. „hugeness problem“)
- Současně ale zachovat otevřenost pro prozatím neřešené aplikace (znovupoužití)



# [ Specifikace terminologické části ]

- Východiskem pro vývoj ontologie bývá seznam relevantních **termínů**
- Například seznam všech podstatných jmen a sloves v textu popisujícím danou doménu
- Možnost již v této fázi využít podporu automatickým nástrojem – text-mining

# [ Odlišení ontologických typů ]

- Třídy, instance, relace, atributy
- Zpravidla existuje více možností!
- Typicky podstatná jména odpovídají třídám a atributům, slovesa pak relacím (neplatí vždy)
- Postupuje se od návrhu tříd a jejich hierarchie
- Je možné využít některé již definované ontologie

# [ Učení ontologií ]

- Zejména pro extrakci relevantních termínů
- Učení probíhá na základě textu, popisujícího cílovou doménu
- Používané techniky:
  1. Syntaktická analýza - zejména odlišení podstatných a přídavných jmen a sloves
  2. Vytvoření distribučního modelu pro slova a víceslovné výrazy
  3. Použití statistických metod strojového učení a dobývání znalostí (text-mining), zejména shlukové analýzy

# [ Extrahované znalosti ]

- Termíny definující třídy
- Hierarchie tříd
  - Buď pomocí slovníku subsumpcí nebo použitím tzv. Harrisovy distribuční hypotézy
- Atributy tříd a jejich obory hodnot
- Obecné relace mezi třídami
  - Využití pokročilé lingvistické analýzy
- Takto získaný model slouží jako základ pro další zpracování člověkem!

[ Děkuji za pozornost ]