

# Ontologie - úvod



**25. 2. 2011**

**PŘEDMĚT: ORGANIZACE ZNALOSTÍ**

**PŘEDNÁŠEJÍCÍ: SILVIE KOŘÍNKOVÁ PRESOVÁ**

# Co je ontologie? hledisko počítačové vědy



obecně akceptovaný výklad ontologie

→ **Explicitní specifikace konceptualizace.**

(Gruber)

Ontologie je popis (podobně jako formální specifikace programu) existujících pojmů a vztahů, který je určen pro inteligentní agenty nebo skupinu agentů. (Hjørland, 2007)



- modifikace (Borst cit. dle Svátek, 2002)
  - **Formální specifikace sdílené konceptualizace.**
    - konceptualizace – Systém pojmů modelující určitou část světa.
    - specifikace – Ontologie prezentuje konceptualizaci v konkrétní formě.
    - formalizace – Ontologie by měla být strojově zpracovatelná (reprezentace pomocí ontologického jazyka, který poskytuje potřebné strojově zpracovatelné kódování).
    - sdílení – Ontologie je výsledkem konsensu určité zájmové skupiny lidí.



- Proč byl termín ontologie (označení pro **filozofickou disciplínu**) převzat **oblastí umělé inteligence**?
  - *Artificial Intelligence (AI) deals with reasoning about models of the world. Therefore, it is not strange that AI researchers adopted the term ontology to describe what can be (computationally) represented of the world in a program". (Studer et al, 1998: 25, cit. v Arano)*
- 80. léta 20. stol. - ontologie v rámci znalostního inženýrství, postupně vznik nového oboru - ontologického inženýrství
- **ontologické inženýrství** - Moderní oblast informatiky, která se zaměřuje na návrh, implementaci a aplikaci ontologií. (Svátek, 2007)

# Výklady W3C



## Ontology

(1) collection of information, generally including information about classes and properties

(2) the information contained in an ontology document

(W3C Glossary and Dictionary)

Ontologie definuje termíny použité k popisu a reprezentaci nějaké znalostní domény. Ontologie jsou užívány lidmi, databázemi a aplikacemi, které potřebují sdílet informace z určité domény (tj. specifická předmětná oblast či část lidského poznání).

Ontologie zahrnuje strojově zpracovatelné definice základních pojmů z domény a vztahů mezi nimi.

(OWL Web Ontology Language Use Cases and Requirements, 2004)

# Jednoduché vers. strukturované ontologie (Slavic)



- **Jednoduché o.** – např. kategorizační schémata (v rámci např. předmět. katalogů), deskriptorové SJ, slovníky s definovanými hierar. vztahy
- **Strukturované o.** – Strojově čitelné zakódování hierarchických vztahů a informací o vlastnostech pojmů a omezeních, které se vztahují k hodnotám těchto vlastností (spojují pojem s instancí).

# Hledisko informační vědy



- **Vickery**: Rozdíl mezi ontologiemi a tradičními selekčním jazyky je v jejich plánovaném užití:
  - ontologie - znalostní báze ve znalostním inženýrství, aplikace umělé inteligence
  - tradiční SJ - navržené pro přímé užití člověkem
- **Lancaster**: termín **ontologie** apod. (taxonomie, hierarch. sítě) tendenční, v knihovnické oblasti postupné nahrazení termínu **klasifikace**
- **Soergel**: ontologie - znovuobjevení klasifikace

(Tennis, 2007 )

# Hledisko informační vědy



- **Garshol**: Považuje ontologie za vyvrcholení vývoje slovníků směrem k nejefektivnějším prostředkům popisu.
- **Jacob**: ontologie a tradiční SJ - mnoho podobností, rozdíly v možnostech aplikace v oblasti sémantického webu
- Hlavní výzkumné trendy vztahující se k ontologiím:
  - automatická indexace a tvorba tezaurů
  - transformace tezaurů do ontologií

(Tennis, 2007 )



# Výklad TDKIV



## Ontologie

- **Jazykové**, **metodické** a **softwarové** nástroje pro explicitní pojmovou reprezentaci skutečnosti, případně samotný výsledek této činnosti.
- Využití nachází především ve znalostním managementu a v pojmovém vyhledávání informací.

# Příklady ontologií



- Různé typy ontologií, příklady základních a oborových/doménových ontologií viz Svátek, 2002, 2007a
- Semantic Web for Earth and Environmental Terminology ([SWEET](http://sweet.jpl.nasa.gov/ontology/)) - oborová ontologie  
<http://sweet.jpl.nasa.gov/ontology/>
- [WordNet](http://wordnet.princeton.edu/) <http://wordnet.princeton.edu/>  
– terminologická ontologie – původní verze pro angličtinu, vícejazyčné varianty, např. [EuroWordNet](http://www.ilc.uva.nl/EuroWordNet/)  
<http://www.ilc.uva.nl/EuroWordNet/>
- [OpenCYC](http://www.opencyc.org/) <http://www.opencyc.org/> - základní ontologie

# Tradiční selekční jazyky versus ontologie

(Soregel, 2004)



- **nedostatek pojmové abstrakce** - Rozdíl mezi pojmem a znakem není dostatečný, vychází zejména z hierarchické struktury selekčního jazyka.
- **omezené sémantické pokrytí** -
  - základní sémantické vztahy- hierarchické, asociace, ekvivalence,
  - nejsou rozlišeny typy pojmů (procesy, předměty, živé organismy apod.)
- **limitované možnosti automatického zpracování** -  
Nejednoznačná sémantika není vhodná pro strojové zpracování.

# Tradiční selekční jazyky versus ontologie

(Soregel, 2004)



**nedostatek konzistence** - aplikace vztahů je nedůsledná, vede k nejednoznačné interpretaci vztahů a vytvoření nesprávné sémantické struktury, např. hierarch. vztahy mohou být asociované a naopak

AGROVOC (verze z r. 2004)

- voda

NT **sladká voda**

NT **pitná voda**

NT **led**

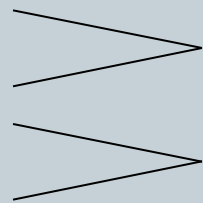
NT **vodní pára**

ASFA

RT led

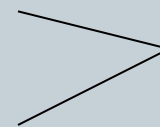
RT vodní pára

.....NT sladká voda



**typy vody**

**skupenství vody**



hier. vztah -  
různá klasif.  
kritéria

# Omezené sémantické pokrytí - příklady

(Soregel, 2004)



Agrovoc - nerozlišené  
vztahy hier. a asociace

- mléko

**NT** kravské mléko

**RT** mléčný tuk

- krávy

**RT** kravské mléko

- Cheddar

**RT** kravské mléko

rozlišené vztahy v ontologii

- mléko

<druhem je> kravské mléko

<obsahuje substanci>  
mléčný tuk

- krávy

<součástí je> kravské mléko

- Cheddar

<vyroben z> kravské mléko

# Pravidlo pro odvození (Soregel, 2004)



## Pravidlo 1

**část X** *<obsahuje substanci>* substance Y

**jestliže** zvíře W *<součástí je>* část X

**a** zvíře W *<přijímá potravu>* substance Y

## Pravidlo 2

**potravina Z** *<obsahuje substanci>* substance Y

**jestliže** potravina Z *<vyroben z>* část X

**a** část X *<obsahuje substanci>* substance Y

# Co můžeme odvodit?



- Z tezauru lze odvodit - asociovaným termínem k *sýru cheddar* je *kravské mléko*
- Z ontologie lze na základě pravidel a tvrzení odvodit, např.
  - Jestliže kráva přijímala krmivo obsahující rtuť, *sýr cheddar* vyrobený z mléka této krávy *<obsahuje substanci>* rtuť.

# Omezené sémantické pokrytí - příklady



## Možnosti vyhledávání

- tezaurus AGROVOC - vztah mezi krávou a mlékem - asociovaný
- ontologie - sémantický vztah vyjádřen - krávy <součástí je> kravské mléko.

Pokud chce uživatel rozšířit vyhledávání pomocí vztahů (např. všechny asociované či podřazené termíny), nemůže pomocí tezauru rozšířit vyhledávání diferencovaně, např. na části krávy, nemoci krav.



# Tezaurus vers. ontologie (Sklenák, 2010)



- podobnost

- terminologické pokrytí určité domény, vymezení vztahů mezi jednotlivými termíny, seskupení termínů do kategorií a podkategorií
- obojí se používá pro popis a organizaci informací

- rozdílnost

- ontologie obsahuje mnohem více vztahů, které jsou formálně definovány a jsou jednoznačné, interpretovatelné strojově i lidmi
- tezaurus – vztahy mezi termíny, ontologie – definování pojmu a vztahu mezi nimi, logický a formální popis

# Deskriptorový odstavec pro deskriptor **víno** tezaurus EuroVoc <http://eurovoc.europa.eu/>

60 ZEMĚDĚLSTVÍ A VÝŽIVA

MT [6021 nápoje a cukr](#)

BT1 [alkoholický nápoj](#)

BT2 [nápoje](#)

NT1 [alkoholizované víno](#)

NT1 [aromatizované víno](#)

NT1 [bílé víno](#)

NT1 [červené víno](#)

NT1 [lahvové víno](#)

NT1 [místní víno](#)

NT1 [nešumivé víno](#)

NT1 [růžové víno](#)

NT1 [sekt](#)

NT1 [stolní víno](#)

NT1 [šumivé víno](#)

NT1 [výběrové víno](#)

RT [hroznové víno](#) [ 6006 ] [chaptalování](#) [ 6036 ] [vinné kvašení](#) [ 6036 ] [vinohradnictví](#) [ 5631 ]



# Praktické aplikace (Svátek, 2002)



- Znalostní management
- Elektronické obchodování
- Zpracování přirozeného jazyka
- Inteligentní integrace informací
- Pojmové vyhledávání informací
- Sémantické webové portály
- Inteligentní výukové systémy

Nasazení v praxi zaostává za rozsáhlým aplikačním výzkumem.

Formalizované znalosti jsou obtížně převoditelné do jednotné podoby, pro jejíž zpracování lze efektivně vyvinout hromadně šířený softwarový nástroj.

# Ontologie a sémantický web



**Semantic Web** - *The Web of data with meaning in the sense that a computer program can learn enough about what the data means to process it.* (W3C Glossary and Dictionary)

**Ontologie** - **klíčová technologie sémantického webu** - Způsob reprezentace významů dokumentů a umožňují, že webové aplikace a inteligentní agenti mohou tyto významy dále používat.

V rámci sémantického webu potřeba **strukturovaných** (ne jednoduchých) **ontologií**, tj. poskytujících strojově zpracovatelné popisy:

- tříd
- vztahů
- vlastností

# Ontologie a sémantický web



- **3 pilíře sémantického webu:** XML, RDF, agenti založení na ontologiích
- Smyslem těchto technických nástrojů je doplnit webové zdroje o strojům srozumitelné **sémantické anotace (metadata)**, aby mohly být rozvíjeny sofistikované služby založené na technikách automatického vyhledávání informací.

(Tennis, 2007)



- Projekty s českou účastí - **Research in Applied Ontology, Semantic Web and Information Extraction in the Czech Republic** <http://keg.vse.cz/ontoweb-cz/>
- Praktické aplikace - projekty:

**Svátek, 2002**

**KSL Ontology Server Projects** <http://www-ksl-svc.stanford.edu:5915/doc/ontology-server-projects.html>

**Magpie**

<http://projects.kmi.open.ac.uk/magpie/main.html#movies>

**Klastrování textových dokumentů s využitím ontologií /Matěj Štefaník. Teze disertační práce**

[http://is.muni.cz/th/50825/fi\\_r/teze.txt](http://is.muni.cz/th/50825/fi_r/teze.txt)

# Praktické aplikace - příklady

## The Friend of a Friend (FOAF) project - <http://www.foaf-project.org/>

- Vytvoření portálu založeného na ontologiích, který bude popisovat osoby, vztahy mezi nimi, jejich záliby a činnosti. Slovník FOAF lze použít i k popisu skupin, organizací aj.
- <http://www.foaf-search.net/>

```
<foaf:Person rdf:ID="me">
```

```
<foaf:name>Silvie Kořínková Presová</foaf:name>
```

```
<foaf:mbox_sha1sum>presova@seznam.cz</foaf:mbox_sha1sum>
```

```
<foaf:workplaceHomepage rdf:resource="http://kisk.phil.muni.cz/" />
```

```
<foaf:workInfoHomepage rdf:resource="http://kisk.phil.muni.cz/mgr-silvie-korinkova-presova-dis" />
```

```
<foaf:knows>
```

```
<foaf:Person>
```

```
<foaf:name>Jana Nováková</foaf:name>
```

```
</foaf:Person></foaf:knows></foaf:Person>
```



# Praktické aplikace - příklady



## DALOS

- <http://www.dalosproject.eu/>
- Projekt DALOS (Drafting Legislation with Ontology-based Support) - nástroje pro podporu návrhů zákonů a direktiv Evropské unie, jsou určeny tvůrcům evropské legislativy a občanům EU.
- Cílem je sjednotit terminologii legislativních textů ve vícejazyčném prostředí a podporovat diskuse vztahující se k interpretacím a implementacím evropské legislativy.
- Přístup k integraci pojmů právního jazyka v různých jazycích, vyžaduje definování mapovacích procedur mezi sémantickými slovníky. Tyto slovníky jsou tvořeny odkazy na prvky ontologie, které popisují základní entity a tvoří tak právní doménu.  
(Štefaník, 2009)

# Praktické aplikace - příklady



**CommerceNet project** <http://www.commerce.net/>

- oblast elektronického obchodování
- Cílem je zpřístupnit produkty a obchodní katalogy v takové formě, která umožní potencionálním zákazníkům nalézt zboží na základě jeho specifikace (x klíčová slova) a srovnávat zboží v různých el. obchodech.
- Rozvoj ontologií termínů a vztahů popisující třídy produktů.

# Povinná literatura



- Svátek, V., Vacura, M. 2007a. **Ontologické inženýrství** [online]. In DATAKON 2007, Brno, 20.-23. 10. 2007, 32 s. [cit. 2009-02-20]. Dostupné z WWW:  
<http://nb.vse.cz/~svatek/dkon07final.pdf>
- Svátek, V. 2002. **Ontologie a WWW** [online]. In DATAKON 2002, Brno, 19.-22. 10. 2002, 35 s. Dostupné z WWW:  
<http://nb.vse.cz/~svatek/onto-www.pdf>

# Doporučená literatura

- Arano, S. 2005. **Thesauruses and ontologies** [online]. *Hipertext.net*. 2005, č. 3, [cit. 2009-02-20]. Dostupné z WWW: <http://www.hipertext.net/english/pag1009.htm>  
ISSN 1695-5498
- Gilchrist, A. 2003. **Thesauri, taxonomies and ontologies an etymological note**. *Journal of Documentation*, 2003 , roč. 59, č. 1, s. 7-18.
- Gruber, T.R. 1993. **A Translation Approach to Portable Ontology Specifications**. *Knowledge Acquisition*, 1993, č. 5, s. 199.
- Hjørland, B. 2007. **Ontology** [online]. Last edited: 24-05-2007

[cit. 2011-02-10]. Dostupné z WWW:  
[http://www.iva.dk/bh/lifeboat\\_ko/CONCEPTS/ontology.htm](http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/ontology.htm)

# Doporučená literatura



- Chowdhury, G. G. ; Chowdhury, S. 2007. *Organizing information : from the shelf to the Web*. London : Facet, 2007. **Kap. 10 Ontology**, s. 171-185. ISBN 978-1-85604-578- dostupné v ÚK – registrační pult
- Kayed, A. aj. 2008. **Building Ontological Relationships : A New Approach**. *Journal of the American Society for Information Science and Technology*, 2008, roč. 59, č. 11, s. 1801–1809.
- **OWL Web Ontology Language : Use Cases and Requirements** [online]. 2004. W3C Recommendation 10 February 2004, Copyright © 2004 W3C [cit. 2010-04-27]. Dostupné z WWW: <http://www.w3.org/TR/2004/REC-webont-req-20040210/>
- Sklenák, V. 2010. **Sémantický web** [online]. In *Sborník konference Inforum*. © 1995-2010 AiP [cit. 2010-04-27]. Dostupné z WWW: <http://www.inforum.cz/archiv/inforum2003/prispevky/Sklenak>

# Doporučená literatura



- Slavic, A. 2005. **Knowledge Organization Systems, Network Standards and Semantic Web**. IN *Informacijske znanosti u procesu promjena*. Zagreb: Zavod za informacijske studije, 2005. s. 5-22. Dostupné též z WWW:  
[http://dlist.sir.arizona.edu/1326/02/semweb\\_kos\\_EN\\_2.pdf](http://dlist.sir.arizona.edu/1326/02/semweb_kos_EN_2.pdf)
- Soregel, D. aj. 2004. **Reengineering thesauri for new applications : The AGROVOC Example** [online]. FAO, Office of Knowledge Exchange, Research and Extensi, 2004. [cit. 2011-01-27]. Dostupné z WWW:  
<http://www.fao.org/docrep/008/af234e/af234e00.htm>
- Svátek, V. 2007b. **Research in Applied Ontology, Semantic Web and Information Extraction in the Czech Republic** [online]. Last update Jun 20, 2007 [cit. 2009-03-27]. Dostupné z WWW:  
<http://keg.vse.cz/ontoweb-cz/>

# Doporučená literatura



- Štefaník, M. 2009. **Klastrování textových dokumentů s využitím ontologií** [online]. Teze disertační práce. V Brně dne 15.1.2009 [cit. 2011-01-27]. Dostupné z WWW:  
<http://is.muni.cz/th/50825/fir/teze.txt>
- Tennis, J. T., Calzada-Prado, J. 2007. **Ontologies and the Semantic Web : Problems and Perspectives for LIS Professionals.** *IBERSID: Revista de Sistemas de Información y Documentación*, č. 1, s. 303-311.
- **W3C Glossary and Dictionary** [online]. Copyright © 2000-2003 W3C [cit. 2011-01-27]. Dostupné z WWW:  
<http://www.w3.org/2003/glossary/>