

Ontologie - metodika tvorby



9. 3. 2012

PŘEDMĚT: ORGANIZACE ZNALOSTÍ

PŘEDNÁŠEJÍCÍ: SILVIE KOŘÍNKOVÁ PRESOVÁ

Metodologie tvorby ontologií



- různé metodologie
- Základní pravidla při tvorbě ontologií (Noy, 2009)
 1. Neexistuje jediný správný model domény – vždy existuje více realizovatelných variant.
 2. Vývoj ontologií je opakující se proces.
 3. Pojmy v ontologii by se měly vztahovat k pojmům a vztahům v oblasti, k níž se ontologie váže.

Výklad ontologie z hlediska tvorby ontologií



- Ontologie je formální, explicitní popis
 - **pojmu** v doméně - nazývají se **třídy** (někdy též pojmy)
 - **vlastností** každého pojmu popisující různé rysy a znaky pojmu - nazývají se **sloty** (někdy nazývány roles nebo properties)
 - **omezení slotů** - nazývají se **fasety** (někdy nazývány role restrictions, constraints)
- Ontologie spolu se souborem **instancí tříd** (**individuí**/jedinců) tvoří znalostní bázi (např. u expertních systémů) - 2 složky:
 - **TBOX** (Terminological Components): samotná ontologie - třídy a vlastnosti
 - **ABOX** (Assertional Components): instance tříd (Husáková, 2009)



- ukázky v editoru ontologií Protégé, verze 3.4 - **protégé-frames** – Vychází z modelu ontologie, která se skládá z množiny tříd hierarchicky uspořádaných, z množiny slotů asociovaných s třídami, z množiny instancí – individuálních případů pojmů. Slouží pro tvorbu znalostních systémů založených na rámcích (**rámcová schémata**).



- **editory ontologií** – Nástroje pro tvorbu a správu ontologií, obvykle používají některý z jazyků pro reprez. ontologií, některé umožňují export do jiných jazyků.
 - více viz Denny, 2002, 2004



- Třídy jsou středem zájmu většiny ontologií - popisují pojmy v doméně. Jsou hierarchicky organizovány do podtříd, které jsou specifitější než nadřazená třída.
 - Např. třída vín reprezentuje všechna vína. Lze vytvořit podtřídy - bílé v., červené v.
- Sloty popisují vlastnosti tříd a individuí.



- Třídy jsou středem zájmu většiny ontologií - popisují pojmy v doméně. Jsou hierarchicky organizovány do podtříd, které jsou specifitější než nadřazená třída.

- Např. třída vín reprezentuje všechna vína. Lze vytvořit podtřídy - bílé v., červené v.

- Sloty popisují vlastnosti tříd a individuí.

Např. Rulanské šedé výběr z hroznů 2000 (individuum, instance/výskyt třídy bílé víno) – suché, plné víno neutrální chuti, výrobce – Vinařství Hrbáč

- slot 1 obsah cukru → hodnota suché
- slot 2 plnost → hodnota plné víno
- slot 3 chuť → hodnota neutrální
- slot 4 výrobce → hodnota Vinařství Hrbáč



- **Individuum/jedinec** reprezentuje určitý objekt domény, který náleží určité třídě.
 - Jedná se o konkrétní výskyt - instanci třídy.
 - Jedinec přitom může patřit jedné nebo více třídám.
 - Je rozdíl mezi jedincem a instancí. Jedinec je instancí nějaké třídy.

(Husáková, 2009)

Příklad metodologie (Noy, 2009)



Základní kroky:

- Vymezení domény a rozsahu
- Zmapování existujících ontologií – převzetí stávající
- Vyjmenování důležitých termínů v ontologii
- Definování tříd a hierarchie
- Definování slotů/vlastností
- Definování faset/omezení slotů/vlastností
- Tvorba individuí

Vymezení domény a rozsahu



- Jakou doménu bude ontologie pokrývat? (reprezentace jídla a vína)
- Jaký je účel ontologie? (navrhnout vhodné kombinace jídla a vína)
- Na jaké otázky se bude pomocí ontologie odpovídat?
 - tzv. **kompetenční otázky** – slouží též k určení rozsahu ontologie (Obsahuje ont. dost informací? Vyžadují odpovědi danou zvolenou specifičností?)
 - Je Zweigeltrebe bílé nebo červené víno?
 - Jaké charakteristiky vína ovlivňují to, zda se hodí k dezertu?
 - Hodí se rulandské bílé k mořským rybám?
 - Je vhodnější k těstovinám bílé nebo červené víno?

Ukázka modulu Queries

The screenshot displays the Protégé 3.4 application window. The title bar reads "Vino Protégé 3.4 (file:IC:\Program%20Files\Protege_3.4\Vino.pprj, Protégé Files (.pont and .pins))". The menu bar includes File, Edit, Project, Code, Window, Collaboration, Tools, and Help. The toolbar contains various icons for file operations and editing. The main interface has several tabs: Classes, Slots, Forms, Instances, and Queries. The Queries tab is active, showing a query editor with the following content:

Query

Class	Slot	Class
bílá vína	jídlu	dezert

contains

More Fewer Clear Match All Match Any Find

Query Name
Jaké rulandské si dát k dezertu? Add to Que...

Query Library

- ▲ Jaké červené se hodí k vepřovému masu?
- ▲ Jaké víno se hodí k rybě?
- ▲ Jaké rulandské si dát k dezertu?
- ▲ Jaké polosuché víno se hodí k drůbežímu masu?

Search Results (1)

- ◆ Rulandské šedé 2007 (Rulandské šedé)

Zmapování existujících ontologií – převzetí stávající



- Proč převzít ontologii?
 - Šetří to úsilí.
 - Interakce s nástroji užívající jiné ontologie.
 - Vytvořená ontologie je již prověřená nějakou aplikací.
- Na vyjádření ontologie (ontologický jazyk) často nezáleží, protože tzv. knowledge-representation systems umí importovat a exportovat ontologie.

Seznamy dostupných ontologií - příklady



- Ontolingua ontology library
<http://www.ksl.stanford.edu/software/ontolingua/>
- DAML ontology library
<http://www.daml.org/ontologies/>
- Protege Ontology Library
http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library
- SchemaWeb <http://www.schemaweb.info/default.aspx>
- vyhledávací služby - Swoogle
<http://swoogle.umbc.edu/> , OntoSelect
<http://olp.dfki.de/ontoselect/> , Watson
<http://watson.kmi.open.ac.uk/WatsonWUI/>

Vyjmenování důležitých termínů v ontologii



- Vytvoření seznamu všech možných termínů.
- Zatím se neberou na vědomí vztahy mezi pojmy, role pojmů, např. zda půjde o třídu či slot.

Pomocné otázky

- O jakých termínech/pojmech budeme něco tvrdit či je vysvětlovat?
- Jaké mají tyto termíny vlastnosti?
- Co chceme říci o těchto termínech?

Definování tříd a hierarchie



- Krok nezávislý na způsobu tvorby hierarchie - nejprve **určení tříd** - tj. které termíny označují třídy.

Obecně lze říci - Roli tříd plní termíny označující nezávislé objekty spíše než termíny, které tyto objekty popisují.

např. **víno x cukernatost vína x jakost**

- Na množině termínů popisující třídy se vytvářejí hierarchické vztahy.
- Otázka, zda jde o třídu či individuum - instanci třídy (problematizováno dále).

Definování tříd a hierarchie



- Uspořádání pojmů do tříd a tříd nižších řádů (podtříd), obvykle na základě hierarchického vztahu, nejčastěji generického (rod-druh). —————> tvorba taxonomie
- **třída** - Skupina klasifikovaných pojmů vzniklá na základě shodných charakteristik za účelem definování sémantických vztahů mezi nimi. (TDKIV)
 - příklad - třída všech vín, třída bílých vín
 - > Některá vína jsou bílá, ale všechna bílá vína jsou vína.

Způsoby tvorby hierarchie



- **top-down** – Nejprve se definují nejobecnější pojmy a následně se specifikují.
- **bottom-up** – Definují se nejspecifičtější/nejužší pojmy, které se následně organizují do obecnějších/nadřazených tříd.
- **combination** – Vyberou se nejdůležitější pojmy a ty se následně generalizují či specifikují.

Ukázka vytvořené hierarchie tříd

The screenshot displays the Protégé 3.4 application window. The title bar reads "Vino Protégé 3.4 (file:IC:\Program%20Files\Protege_3.4\Vino.pprj, Protégé Files (.pont and .pins))". The menu bar includes "File", "Edit", "Project", "Code", "Window", "Collaboration", "Tools", and "Help". The toolbar contains various icons for file operations and editing. Below the toolbar, there are tabs for "Classes", "Slots", "Forms", "Instances", and "Queries".

The main interface is divided into two main panes:

- CLASS BROWSER:** Shows a class hierarchy for the project "Vino". The hierarchy is as follows:
 - :SYSTEM-CLASS
 - Jídla
 - bezmasá jídla
 - masitá jídla
 - předkrmy
 - těstoviny
 - Vína
 - bílá vína
 - červená vína
 - Frankovka
 - Svatovavřínecké
 - Zweigeltrebe
- CLASS EDITOR:** Shows the details for the selected class ".Jídla". It includes fields for "Name" (".Jídla"), "Role" (set to "Concrete"), and a "Template Slots" table with columns "Name" and "Ca...".

At the bottom of the interface, there is a "Superclasses" pane showing that ".THING" is the superclass of the selected class.

Definování slotů – vlastností tříd



- **sloty** – Vlastnosti tříd popisující různé rysy a znaky (vnitřní struktura pojmů) —→ popisují vlastnosti instancí třídy a vztahy mezi instancemi.
- Samotné třídy neposkytují dostatečný počet informací pro odpovědi (viz krok 2).
- Některé termíny v seznamu označují vlastnosti. Pro každou vlastnost ve vytvořeném seznamu se musí určit, jakou třídu popisuje.
 - třída VÍNA může mít následující sloty: vůně, plnost, cukernatost, barva, výrobce
 - třída VINAŘSTVÍ – sloty: lokace, produkce
- Slot by měl být připojen ke třídě nejvyšší úrovně.

Seznam slotů podtřídy ČERVENÁ VÍNA

The screenshot shows the Protégé 3.4 interface. The main window title is "Vino Protégé 3.4 (file:IC:\Program%20Files\Protege_3.4\Wino.pprj, Protégé Files (.pont and .pins))". The menu bar includes File, Edit, Project, Code, Window, Collaboration, Tools, and Help. The toolbar contains various icons for file operations and editing. The interface is divided into several panes:

- Classes**: A tabbed interface with "Slots" selected.
- CLASS BROWSER**: Shows the class hierarchy for the project "Vino". The hierarchy includes:
 - :THING
 - :SYSTEM-CLASS
 - Jídla
 - Vinaři
 - Vína
 - bílá vína
 - červená vína (selected)
- CLASS EDITOR**: Shows the editor for the class "červená vína" (instance of :STANDARD-CLASS). It includes:
 - Name**: červená vína
 - Role**: Concrete
 - Template Slots**: A table listing slots for the class.

Name	Cardin...	Type	Other Facets
barva	single	Symbol	allowed-values={bílá,červená} value=če...
jakost	single	Symbol	allowed-values={stolní,jakostní-odrůdov...
jméno	single	String	
jídlo	multiple	Class with supercla...	
obsah cukru	single	Symbol	allowed-values={suché,polosuché,sladk...
obsah tříslovin	single	Symbol	allowed-values={vysoký,střední,nízký}
plnost	single	Symbol	allowed-values={plné,střední,lehké}
synonyma	multiple	String	
výrobce	multiple	Instance of Vinaři	inverse-slot=produkce
vůně	multiple	Symbol	allowed-values={jemná,plná,diskrétní,de...

At the bottom of the window, there is a status bar that reads "Total Commander 7.04a - NOT REGISTERED".

Definování slotů – vlastností tříd



- typy vlastností
 - **vnitřní** – vůně, barva, plnost vína
 - **vnější** – cena, název vína
 - **části, pokud je objekt strukturován** – chody jídla
 - **vztahy k jiným objektům** – výrobce vína, produkce vinařství
- Podtřídy „dědí“ sloty/vlastnosti od nadřazené třídy, např. všechny sloty třídy VÍNA budou mít též podtřídy BÍLÁ VÍNA, ČERVENÁ VÍNA a jejich podtřídy.
- Lze přidat specifický slot podtřídě, např. třída ČERVENÁ VÍNA - slot trísloviny, dědí všechny její podtřídy.
- Je potřeba zajistit, aby každá třída, k níž je slot připojen opravdu danou vlastnost měla.

Definování faset slotů



- Sloty mají různé fasety - omezení vlastností (property constraints) popisující **typy hodnot**, **množství hodnot** (kardinalita – kolik hodnot současně může instance nabývat) a **jiné rysy hodnot**, které může slot nabývat, např.

hodnoty slotu plnost – plná, střední, lehká

→ typ hodnoty – symbol

→ kardinalita - single (konkrétní víno může mít jednu hodnotu)

Definování faset slotů – obvyklé fasety



- **kardinalita** – kolik hodnot může slot mít – single, multiple, minimum a maximum

např.

- víno má **jeden charakter jakosti** (buď hodnota stolní, či jakostní, nebo přívlastkové) - **single**, ale může být z **více typů hroznů (multiple)**
- vinaři vyrábějí různá vína - **produkce (multiple)**

Definování faset slotů – obvyklé fasety



- **typ hodnoty**
 - **string** (řetězec znaků)
 - **number**
 - **boolean** (pravda/nepravda)
 - **enumerated/symbol** (seznam všech specifických možných hodnot) – obsah cukru – možné hodnoty – suché, polosuché, sladké
 - **instance** – umožňují definovat vztahy mezi instancemi/individui, musí být definovány třídy, ze kterých instance pocházejí, např. třída VINAŘI má slot/vlastnost produkce s hodnotou typu instance pocházejících ze třídy VÍNA

Ukázka definice slotu

The screenshot displays the Protégé 3.4 software interface. The main window title is "Vino Protégé 3.4 (file:\C:\Program%20Files\Protege_3.4\Vino.pprj, Protégé Files (.pont and .pins))". The menu bar includes File, Edit, Project, Code, Window, Collaboration, Tools, and Help. The toolbar contains various icons for file operations and editing. The main workspace is divided into several panes:

- Classes:** A list of classes including výrobcce, produkce, barva, cukernatost hroznů, jméno, obsah třísloviny, plnost, vůně, and several system classes like :ANNOTATED-INSTA, :ANNOTATION-TEXT, and :ASSOCIATED-FACE.
- Slots:** The "SLOT EDITOR" pane is active, showing the configuration for the slot "vůně" (instance of :STANDARD-SLOT).
 - Name:** vůně
 - Value Type:** Symbol
 - Allowed Values:** jemná
 - Cardinality:** required, multiple
 - Minimum:** (empty), **Maximum:** (empty)
 - Inverse Slot:** (empty)
 - Template Values:** (empty)
 - Default Values:** (empty)
 - Domain:** Vína
- Superslots:** (empty)

Definování faset slotů



- **domain of the slot (doména)** - Třídy, se kterými je slot spojen nebo třídy, jejichž vlastnosti jsou popisovány.
 - doménou slotu výrobce je třída VÍNA, doménou slotu cukernatost je třída VÍNA
- **range of a slot (rozsah)** - Třídy obsahující prvky, které jsou pro daný slot hodnotou typu instance.
 - třída VINAŘSTVÍ je rozsahem slotu výrobce
 - Zweigeltrebe pozdní sběr barrique 2004 (instance třídy Vína) - Vinařství Hrbáč je hodnota slotu výrobce a současně instancí třídy Vinařství

Ukázka definice slotu

The screenshot displays the Protégé 3.4 interface for defining a slot. The main window title is "Vino Protégé 3.4 (file: C:\Program%20Files\Protege_3.4\Vino.pprj, Protégé Files (.pont and .pins))". The menu bar includes File, Edit, Project, Code, Window, Collaboration, Tools, and Help. The toolbar contains various icons for file operations and editing. The interface is divided into several panes:

- Classes:** A list of classes including "výrobce ↔ produkce", "produkce ↔ výrobce", "barva", "cukernatost hroznů", "jméno", "obsah tříslovin", "plnost", "vůně", and several annotated instances.
- SLOT EDITOR:** The active pane for editing the "výrobce" slot. It shows:
 - Name:** "výrobce"
 - Value Type:** "Instance"
 - Allowed Classes:** "Vinaři"
 - Cardinality:** "at most 1" (with "required" and "multiple" options unchecked)
 - Inverse Slot:** "produkce"
 - Domain:** "Vína"
- SLOT BROWSER:** Shows the slot hierarchy for the "Vino" project.
- Superslots:** A section for defining superslots, currently empty.

Určení třídy pro rozsah a doménu slotu

- Volba nejobecnější třídy v dané hierarchii - ne vyjmenovávání podtříd.
- Slot **produkce** se vztahuje k třídě **VINAŘSTVÍ**. Nabývá konkrétní hodnoty na úrovni instance (vinařství produkuje konkrétní víno) - ta spadá přímo pod třídu reprezentující jednotlivou odrůdu vína. Jako doména se však uvede třída Vína.

vína

červená vína

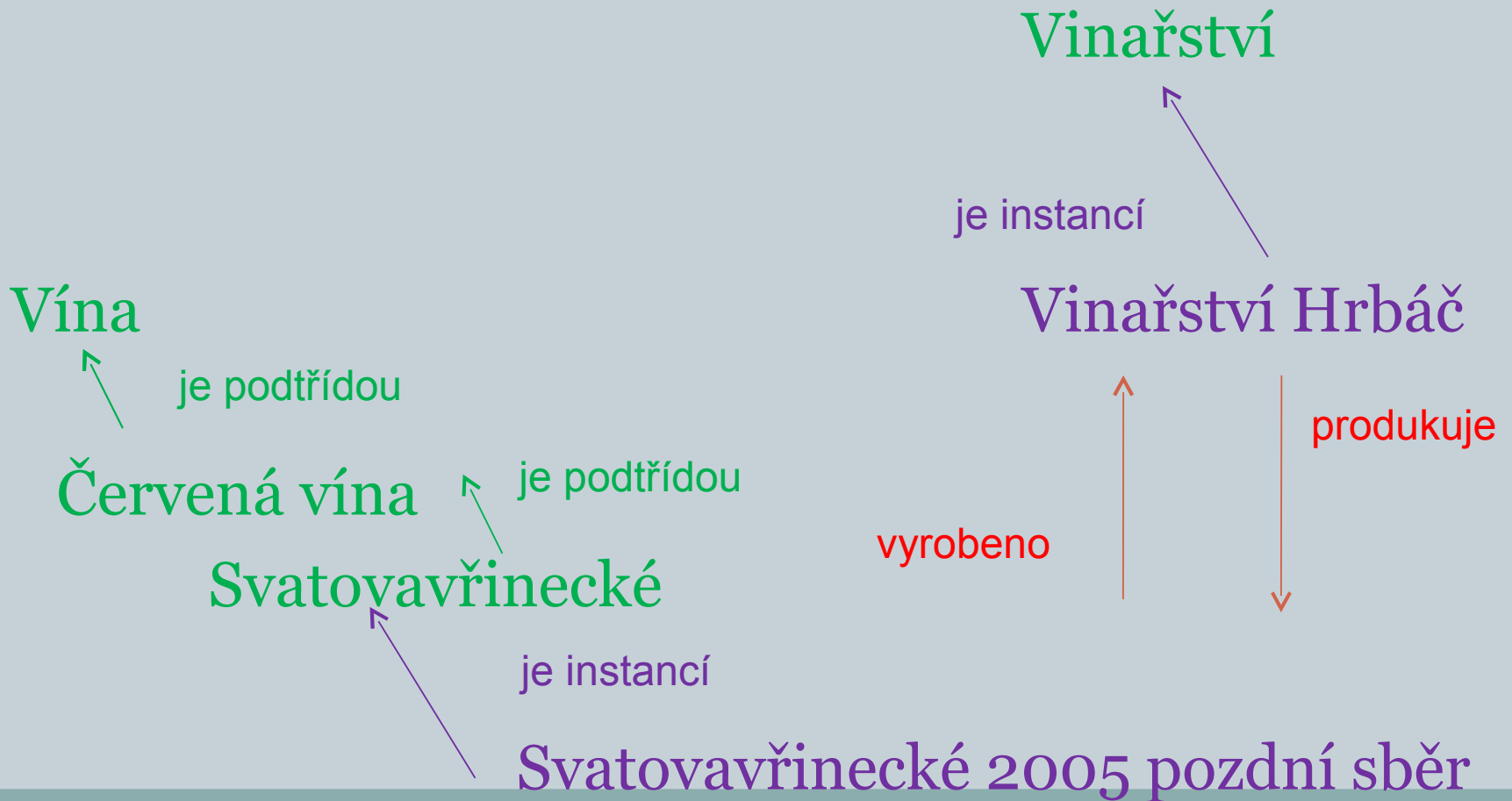
Svatovavřínecké - **Svatovavřínecké 2005**

pozdní sběr - instance a

současně hodnota slotu

produkce - pro daný

slot je třída **VÍNA** rozsahem



Tvorba individuí



- definování individua vyžaduje:
 - výběr třídy
 - tvorbu individuální instance
 - vyplnění hodnot slotů

Ukázka definice instance pro třídu Víno

The screenshot displays the Protégé 3.4 interface for defining an instance of the class 'Zweigeltrebe'. The main window is titled 'Vino Protégé 3.4' and shows a menu bar (File, Edit, Project, Code, Window, Collaboration, Tools, Help) and a toolbar. The 'Instances' tab is active, showing the 'INSTANCE EDITOR' for the instance 'Zweigeltrebe pozdní sběr 2006'. The editor contains several slots for defining the instance's properties:

- Name:** Zweigeltrebe pozdní sběr 2006
- Vůně:** švestky
- Výrobce:** Vinařství Hrbáč
- Jídlo:** rybí maso, těstoviny
- Barva:** červená
- Jakost:** pozdní-sběr
- Obsah Cukru:** suché
- Obsah Tříslovi:** nízký
- Plnost:** plné

The 'INSTANCE BROWSER' on the left shows the class hierarchy, with 'Zweigeltrebe' selected. The 'Types' panel at the bottom left shows the class 'Zweigeltrebe'.

Individium nebo třída?



- Rozhodnutí se odvíjí od konkrétní aplikace ontologie.
- Jednotlivá individua jsou nejspecifičtější pojmy reprezentované ve znalostní bázi.
 - ? individuum Rulandské šedé či Rulandské šedé výběr z hroznů 2000 košer (Vinařství Chrámce)?
(kombinace s jídlem ovlivňují vlastnosti jednotlivých ročníků, nejen odrůda)
 - třídy - abstraktnější, nemění se tak často - stabilní
 - individua - dynamický charakter, konkrétnější
- Jestliže se dají pojmy hierarchizovat
→ třídy

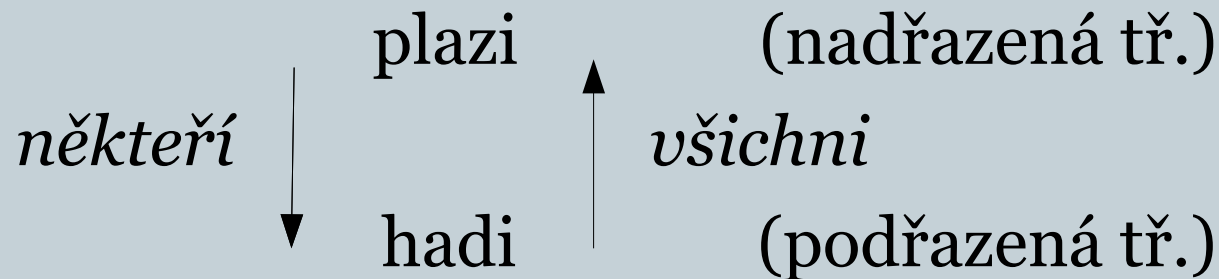


- Tvorba hierarchie a definování vlastností pojmů jsou spolu úzce spojené.
 - V praxi se oba kroky střídají – definuje se několik pojmů v hierarchii a následně se popíše vlastnosti těchto pojmů atd.
- Výchozí verze ontologie – prochází **hodnocením** a **laděním** – prozkoušení v dané aplikaci, užití metod řešení problému, konzultace s experty.
 - Může nastat revize původní verze.

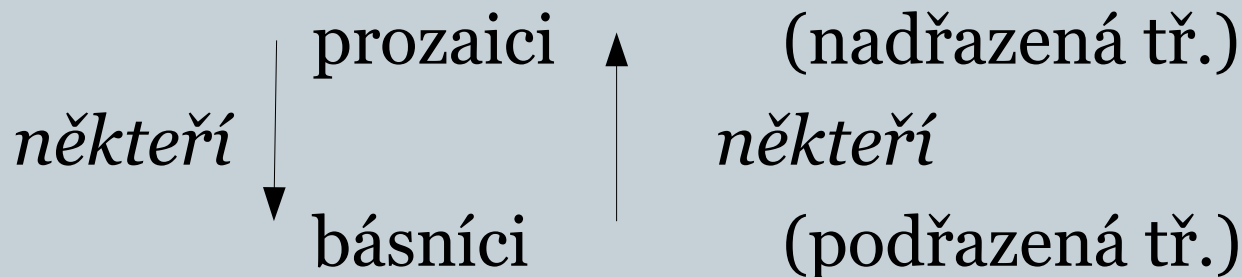
Definice tříd a hierarchie - vybrané problémy



- Kontrola vztahu - nadřazenosti ↔ podřazenosti
- generický (rodo-druhový) vztah



správné
určení vztahu



chybné určení
vztahu

Definice tříd a hierarchie - vybrané problémy



- **Udržování hierarchie** - Souvisí s rozvojem dané domény a změnami v pojmovém systému.
- **Třídy a jejich označení** - Třídy reprezentují pojmy a ne termíny, ty označují dané pojmy.
 - Označení třídy se může měnit v souvislosti se zvolenou terminologií. (Rulandské bílé x Burgundské bílé)
 - Synonyma - Systémy umožňují vytvořit seznam synonym, pokud ne - uvedení v dokumentaci ke třídě.

Definice tříd a hierarchie - vybrané problémy



Souřadné třídy (tj. stejné úrovně)

- Musí být na stejné úrovni obecnosti.
 - např. třída BÍLÁ VÍNA vers. RULANDSKÉ BÍLÉ (**vztah nadřazenosti - podřazenosti**), třída BÍLÁ VÍNA, ČERVENÁ VÍNA (**souřadné třídy**)
 - Výjimka - základní třídy (tj. nejvyšší úrovně) - reprezentují základní kategorie domény.

Počet souřadných tříd



Pokud pouze 1 přímá podtřída - může být modelování dané oblasti neúplné či nepřesné.

Ryzlink

Ryzlink rýnský

Veltlínské zelené
Sauvignon

Sylvánské zelené

Ryzlink rýnský

Veltlínské zelené
Sauvignon

Sylvánské zelené

Počet souřadných tříd



Pokud více jak deset podtříd - je vhodné vytvořit obecnější třídu.

Zweigeltrebe

Frankovka

Svatovavřinecké

Rulandské bílé

Ryzlink rýnský

Ryzlink vlašský

Veltlínské zelené

Sauvignon

Sylvánské zelené

Neronet

bílá vína

Rulandské bílé

Ryzlink rýnský

Ryzlink vlašský

Veltlínské zelené

Sauvignon

Sylvánské zelené

červená vína

Zweigeltrebe

Třída nebo slot?



Určitá vlastnost/vztah - klasifikační kritérium či slot?

- Barva vína, zralost/jakost vína, - klasifikační kritérium/princip dělení či sloty?
 - Závislost na důležitosti pojmu v dané doméně, tematickém rozsahu a zaměření dané ontologie.
 - Jestliže pojmy s různými hodnotami vlastností jsou omezeními pro jiné sloty v různých třídách - potom uvedeny jako třídy, např. odlišné kombinace bílého a červeného vína s jídlem, barva vína ovlivňuje jiné vlastnosti vína.
 - Na základě odlišnosti lze odvodit rodo-druhový vztah - tvorba třídy.

Tři klíčové aspekty adekvátnosti ontologického obsahu



(Svátek, 2007, s. 15)

- **Přesnost:** Obsah ontologie by měl co nejvěrohodněji odpovídat situacím v reálném světě.
- **Srozumitelnost:** Ontologie by měla být co nejlépe srozumitelná lidem, kteří se na jejím vývoji nepodíleli.
- **Způsobilost k odvozování:** Nad ontologií by mělo být možné netriviálně strojově odvozovat. Proto je třeba využít možností formálního aparátu jazyka.

Povinná literatura



- Svátek, V., Vacura, M. 2007. **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/dkono7final.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



- Denny, M. 2004. **Ontology Tools Survey, Revisited** [online]. July 14, 2004 [cit. 2009-03-27]. Dostupné z WWW: <http://www.xml.com/pub/a/2004/07/14/onto.html>
- Denny, M. 2002. **Ontology Building : A Survey of Editing Tools** [online]. November 06, 2002 [cit. 2009-03-27]. Dostupné z WWW: <http://www.xml.com/pub/a/2002/11/06/ontologies.html>
- Husáková, M. 2009. **Znalostní technologie I.** [online]. Datum poslední aktualizace: 27. 10. 2009 [cit. 2011-01-27]. Dostupné z WWW: http://lide.uhk.cz/fim/ucitel/fshusam2/lekarnicky/zt1/zt1_index.html

Doporučená literatura



- Noy, N. F., McGuinness, D. L. 2009. *Ontology Development 101 : A Guide to Creating Your First Ontology* [online]. [cit. 2009-03-27]. Dostupné z WWW: <http://www.ksl.stanford.edu/people/dlm/papers/ontology101/>