

Ontologie a OWL

Prof. Ing. Vojtěch Svátek, Dr.

Zimní semestr 2018

<http://nb.vse.cz/~svatek/rzzw.html>

Témata

- Ontologie v informatice
- Jazyk OWL
- Vybrané aktuální problémy ontologického inženýrství a projekty KIZI (*informativně*)

Témata

- **Ontologie v informatice**
- Jazyk OWL
- Vybrané aktuální problémy ontologického inženýrství a projekty KIZI (*informativně*)

Pojem ontologie

- Ve filosofii:
 - „nauka o bytí“
 - „univerzální soustava znalostí o světě“
(tak, jak je, nezávisle na usuzování o něm...)
 - Kategorie „jsoucen“ (viz např. Aristoteles)
- V informatice:
 - „soustava znalostí o světě“
 - bez nároku na ucelenost
(mnoho dílčích ontologií pro různé domény)
 - často účelově vzniklý artefakt zohledňující způsob použití v informačním systému / aplikaci

Definice ontologie (v inf.)

- (T. Gruber, 1993) „explicitní specifikace konceptualizace“
- (W. Borst, 1997) „*formální* specifikace *sdílené* konceptualizace“
- (B. Swartout, 1997) „hierachicky strukturovaná množina *termínů* popisujících určitou věcnou oblast...“
- a mnoho dalších

Typická struktura ontologie

- **Termíny**, resp. **pojmy**, které se používají k popisu nějaké reality – jim odpovídají reálné objekty
 - Rozlišují se obecné pojmy – **třídy**, a jejich **instance**
- **Hierarchické** uspořádání pojmů z hlediska obecnosti (nadtrída - podtrída)
- I **nehierarchické** vazby – „pojmenované“ vztahy, ve kterých se mohou objekty reálného světa vystupovat

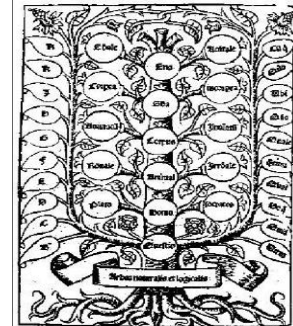
Dva pohledy na ontologii

Aristoteles: Definitio
per **genus proximum**
et **differentia specifica**



Porfyriův strom:
myslíci vs. materiální
živé vs. neživé
racionální vs. neracionální
...

Doktorand je **student**
který úspěšně ukončil magisterskou úroveň
studia a věnuje se určitému výzkumnému tématu
pod vedením kvalifikovaného školitele



Dva pohledy na ontologii

Aristoteles: Definitio
per **genus proximum**
et **differentia specifica**



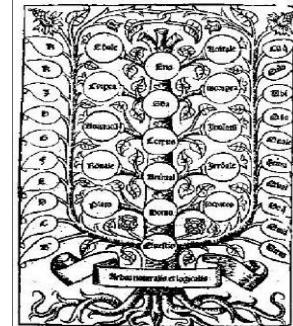
Modulární
systém
propojených
definic

Doktorand je **student**
který úspěšně ukončil magisterskou úroveň
studia a věnuje se určitému výzkumnému tématu
pod vedením kvalifikovaného školitele

Porfyriův strom:
myslí vs. materiální
živé vs. neživé
racionální vs. neracionální

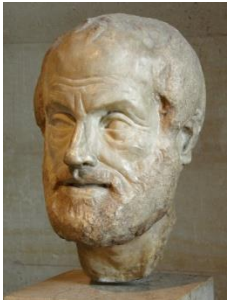
...

Systematická
taxonomie



Dva pohledy na ontologii

Aristoteles: Definitio
per **genus proximum**
et **differentia specifica**



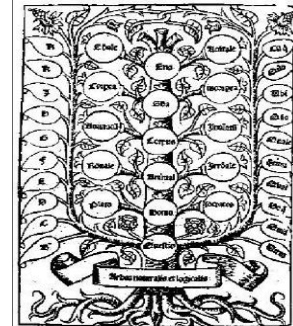
Modulární
systém
propojených
definic

Jedno z využití
deskripční logiky...

Porfyriův strom:
myslíci vs. materiální
živé vs. neživé
racionální vs. neracionální
...

Systematická
taxonomie

Doktorand je **student**
který úspěšně ukončil magisterskou úroveň
studia a věnuje se určitému výzkumnému tématu
pod vedením kvalifikovaného školitele



Jazyky pro reprezentaci ontologií

- Historicky mnoho, vzájemně si konkurujících
- V současnosti dominuje *OWL*, případně *RDFS* jako jeho (cca) podmnožina
- Syntakticky společný základ (rodina W3C standardů), ale významně odlišné *kontexty a způsoby použití*

Využívání ontologií v OWL (1)

Rozsáhlé **doménové** ontologie

- nadále vznikají zejména v **biomedicíně** (viz <http://bioportal.bioontology.org/>), např. ICD (Mezinárodní klasifikace nemocí)
- unifikovaná struktura generovaná automaticky pomocí šablon
- důraz na využití pro strojové odvozování v *deskripční logice* (DL) viz 5FI430 (Ing. Vacura)
- příklad ze SNOMED-OWL (v Manch. syntaxi)

```
Appendicectomy equivalentTo  
Surgical_Procedure and  
method some Excision) and  
procedure-site some Appendix_structure)
```

SNOMED-OWL

- SNOMED, Systematic NOmenclature of MEDicine, je kategorizační systém vytvářený asociací amerických patologů
- První verze převodu do OWL v r. 2007 obsahovala cca 400 tisíc konceptů
- Systém není uzavřený - je možno libovolně vytvářet nové koncepty, aniž by musely obdržet vyčleněné jméno a ID (post-koordinace)
 - Excision and (procedure-site some (kidney and (laterality some left)))
- To vede k nutnosti využití odvozovacích systémů

Scénář využití SNOMED-OWL

- (Dle J. Daviese, British Telecom + projekt RIDE)
- Lékař vloží klinická data do zdravotnické aplikace a přiřadí je konceptu, který pro tento účel *nově vytvoří*, např. „alergie způsobená mandlemi“
- Aplikace informuje o novém konceptu *terminologickou službu*
- Terminologická služba kontaktuje *odvozovací nástroj* (nad DL), a ten automaticky zařadí nový koncept do *taxonomie* existujících
- Terminologická služba informuje zdravotnickou aplikaci o změně taxonomie, ve které je nyní nový koncept podřazený konceptu „alergie na ořechy“

Využívání ontologií v OWL (2)

2. Ontologie pro lepší přístup ke konkrétnímu datovému zdroji - tzv. **OBDA**
 - kompaktní, využívané aplikací jako celek, využívání přinejmenším zčásti na základě tzv. *předpokladu uzavřeného světa (CWA)*
 - spíše nadstavba relačního datového schématu než skutečná „ontologie sémantického webu“

Využívání ontologií v OWL (3)

3. Využití ontologií pomocí **pravidlových systémů**

- Např. některé systémy business rules, deklarativně definující „politiky“ v informačních systémech

Využívání ontologií v OWL (4)

4. Datové slovníky pro **linked data**

- zajišťují minimální vyjádření „významu“ dat v infrastruktuře bez vymezeného cílového určení
- provázané mezi sebou, kombinují se, **přepoužívají** se pro mnoho různých datasetů
- zpravidla využívají jen malou část vyjadřovací síly OWL
 - bývá to RDFS a „něco málo navíc“...

Témata

- Ontologie v informatice
- Jazyk OWL
- Vybrané aktuální problémy ontologického inženýrství a projekty KIZI (*informativně*)

Jazyk OWL

- První verze doporučení W3C z r. 2004
- Aktuálně platná verze OWL 2 z r. 2009
 - Pro úvodní seznámení vhodný OWL 2 Primer <http://www.w3.org/TR/owl2-primer/>
- Několik syntaxí
 - „funkční“ syntaxe – normativní
 - zápis v podobě trojic RDF (v XML, Turtle npod.) – málo přehledné, ale umožňuje zpracovávat nástroji pro RDF
 - Manchesterská syntaxe – nepokrývá celý jazyk, ale je lidsky nejčitelnější
 - syntaxe OWL přímo v XML

„Množinové“ predikáty v RDFS (rekapitulace)

- `rdfs:subClassOf`, `rdfs:subPropertyOf`, `rdfs:domain`, `rdfs:range`

IF `ex:Professor` **`rdfs:subClassOf`** `ex:Teacher`.
AND `ex:Petr` `rdf:type` `ex:Professor`.
THEN `ex:Petr` `rdf:type` `ex:Teacher`.

(třída – třída)

IF `ex:isHeadOf` **`rdfs:subPropertyOf`** `ex:isEmployedBy`.
AND `ex:Vilem` `ex:isHeadOf` `ex:KIZI`.
THEN `ex:Vilem` `ex:isEmployedBy` `ex:KIZI`.

(vlastnost – vlastnost)

IF `ex:isEmployedBy` **`rdfs:domain`** `ex:Person`.
AND `ex:Vilem` `ex:isEmployedBy` `ex:KIZI`.
THEN `ex:Vilem` `rdf:type` `ex:Person`.

(vlastnost – třída)

IF `ex:isEmployedBy` **`rdfs:range`** `ex:Organization`.
AND `ex:Vilem` `ex:isEmployedBy` `ex:KIZI`.
THEN `ex:KIZI` `rdf:type` `ex:Organization`.

(vlastnost – třída)

OWL jako nadstavba RDFS

- Nejvýznamnějším rozšířením je možnost používat nejen pojmenované třídy, ale i **anonymní třídy** definované logickým výrazem
 - Konjunkce/disjunkce tříd, třídy definované restrikcí nad vlastností (viz příklady ze SNOMED-OWL), výčtové třídy...
 - Lze do sebe vnořovat – skládat z jednodušších pojmů složitější
- Vlastnosti se explicitně rozlišují jako **objektové** (hodnotou je instance nějaké třídy) a **datové** (hodnotou je literál)
 - Vlastnostem lze přiřadit **charakteristiky**, jako je funkčnost, symetrie, vzájemná inverznost (pro dvě vlastnosti) apod.
- Možnost formulovat negativní tvrzení

Profily OWL

- <https://www.w3.org/TR/owl2-profiles/> specifikuje 3 profily („podjazyky“) OWL pro specifické použití
 - OWL 2 EL: vhodný pro ontologie s mnoha třídami, které je třeba automaticky podřazovat (zvl. biomedicína)
 - OWL 2 QL: vhodný pro využití ontologie při obohaceném dotazování do RDBMS (OBDA)
 - OWL 2 RL: vhodný pro implementaci odvození pomocí pravidlových systémů (např. business rules)
- Další rozšíření jsou dosud nestandardizovaná, zejména **OWL LD**

OWL LD

- Profil OWL vycházející z potřeb návrhářů slovníků linked data
 - Empirické šetření provedené v r. 2011, viz <https://arxiv.org/abs/1202.0984>
- Podmnožina OWL RL, obsahující výhradně konstrukce vyjádřitelné pomocí **jediné** RDF trojice (tj. bez „temných trojic“)
- <http://semanticweb.org/OWLLD/>

Konstrukty OWL LD

(nad rámec RDFS)

- Ne/identita individuí
 - owl:sameAs, owl:differentFrom
- Individuální charakteristiky vlastností
 - owl:FunctionalProperty,
owl:InverseFunctionalProperty,
owl:IrreflexiveProperty, owl:SymmetricProperty,
owl:AsymmetricProperty, owl:TransitiveProperty
- Vzájemné vztahy tříd / vlastností
 - owl:equivalentClass, owl:disjointWith,
owl:equivalentProperty, owl:propertyDisjointWith,
owl:inverseOf

Využití některých konstruktů

- owl:sameAs
 - explicitně udává identitu objektů
- owl:InverseFunctionalProperty
 - „vlastnost klíče“, identitu objektů z ní lze nepřímo odvodit
- owl:TransitiveProperty
 - odvození v řetězci výskytů vztahu
- owl:equivalentClass / owl:equivalentProperty /
 - zjednodušený zápis dvou subsumpcí (rdfs:subClassOf / rdfs:subPropertyOf); typicky se používá pro zajištění kompatibility se starými verzemi tříd/vlastností
- owl:disjointWith / owl:propertyDisjointWith
 - kontrola neslučitelných tříd/vlastností

Témata

- Ontologie v informatice
- Jazyk OWL
- Vybrané aktuální problémy ontologického inženýrství a projekty KIZI (*informativně*)

Vybrané aktuální problémy

- KIZI se systematicky zabývá tvorbou a využíváním ontologií cca od r. 2000
- Podrobně viz např.
<https://kizi.vse.cz/english/research-groups/semantic-web-and-ontological-engineering/research-focus/>
- Aktuálně řešené problémy: zejména
 - Fokusovaná kategorizační síla ontologií
 - Rozklad webové znalostní báze na ontologické schéma, znalostní graf a kódovníky
 - Modely ontologického pozadí (viz 5FI430, Ing. Vacura)

Fokusovaná kategorizační síla

- Je dána výchozí třída – **fokusová třída** – o které již víme, že je objekt její instancí
- Výstupem kategorizace je jiná třída, která je sémantickou **specializací** fokusové třídy
- Teze: Ontologií vhodnou pro přepoužití je taková, která nabízí **mnoho možností kategorizace** pro fokusové třídy objektů ve vytvářeném datasetu, tj. má pro něj vysokou **fokusovanou kategorizační sílu**

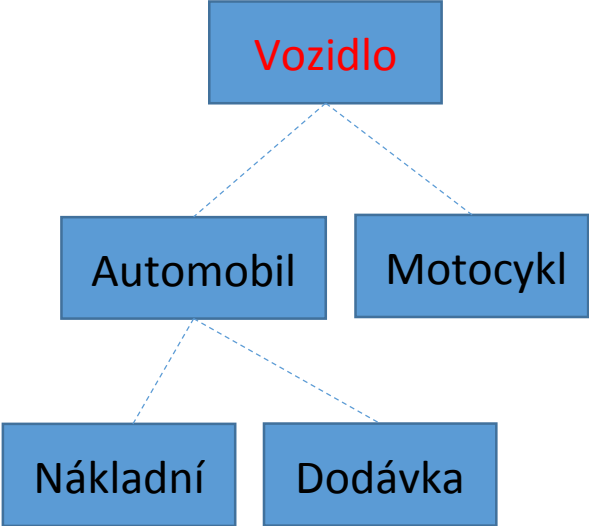
Ilustrační příklad

- Majitel autobazaru chce na webu sdílet data o prodávaných vozech
- Pro popis dat chce přepoužít jednu nebo malý počet ontologií

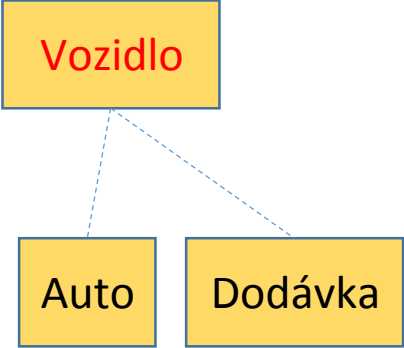
Ilustrační příklad

- Majitel autobazaru chce na webu sdílet data o prodávaných vozech
- Pro popis dat chce přepoužít jednu nebo malý počet ontologií
- Kterou z dostupných ontologií, obsahujících požadovanou fokusovou třídu (nazvanou **Vozidlo**), pro přepoužití vybrat?

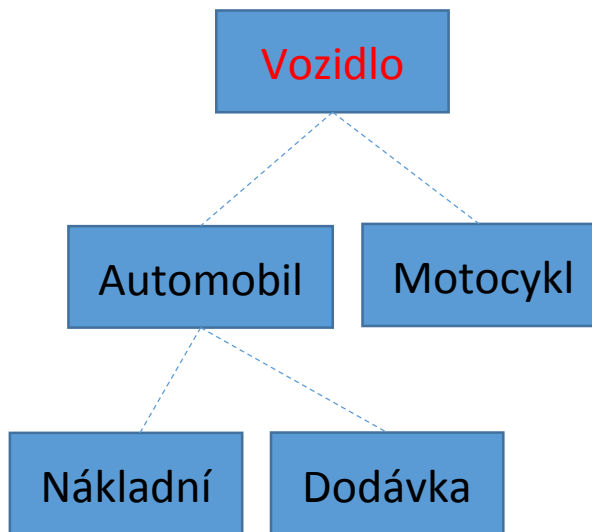
Ontologie A



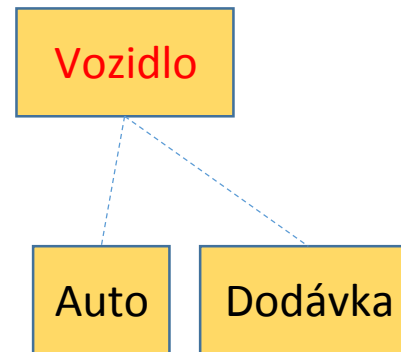
Ontologie B



Ontologie A



Ontologie B



Kategorie vozidel dle A:

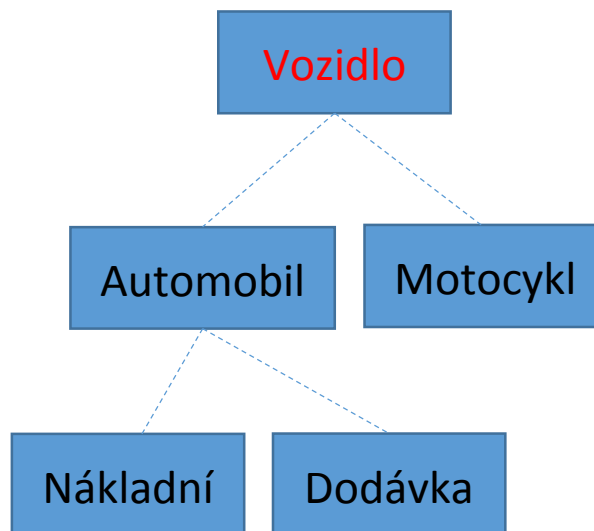
Automobil

Motocykl

Nákladní

Dodávka

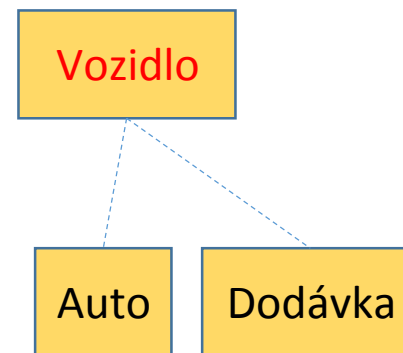
Ontologie A



Kategorie vozidel dle A:

Automobil
Motocykl
Nákladní
Dodávka

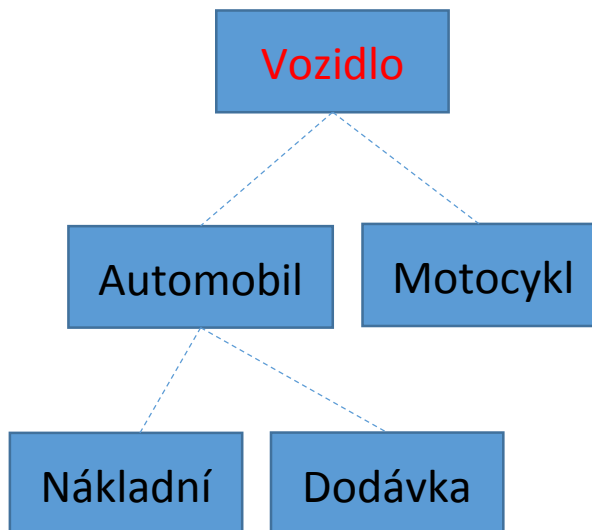
Ontologie B



Kategorie vozidel dle B:

Auto
Dodávka

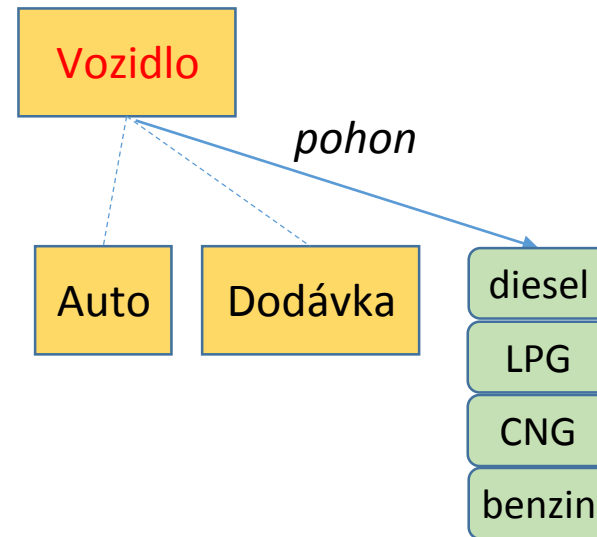
Ontologie A



Kategorie vozidel dle A:

Automobil
Motocykl
Nákladní
Dodávka

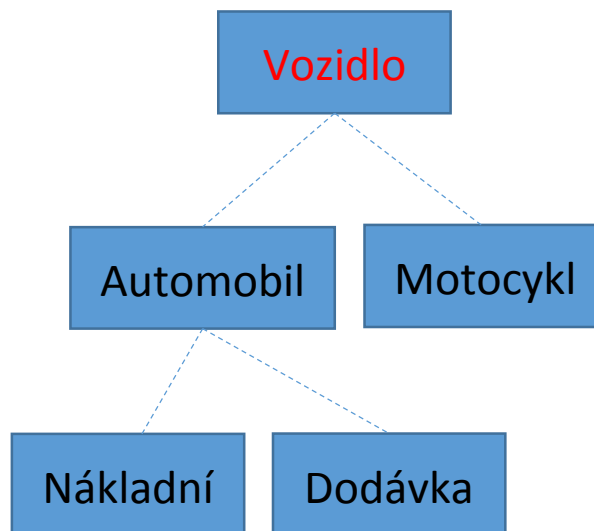
Ontologie B



Kategorie vozidel dle B:

Auto
Dodávka

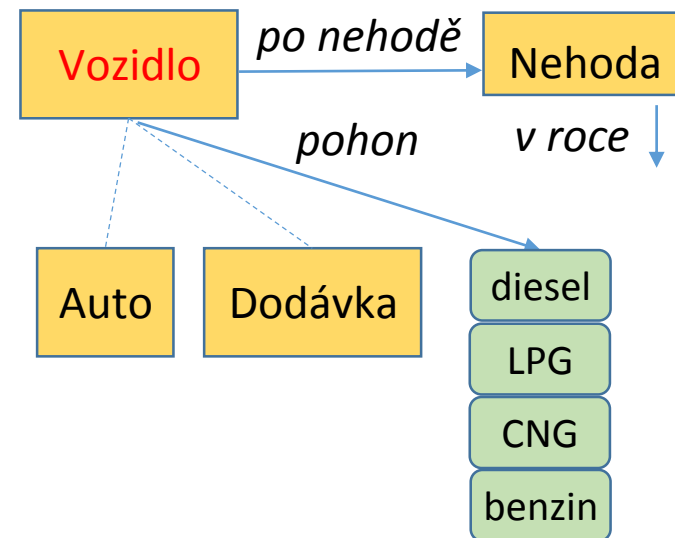
Ontologie A



Kategorie vozidel dle A:

Automobil
Motocykl
Nákladní
Dodávka

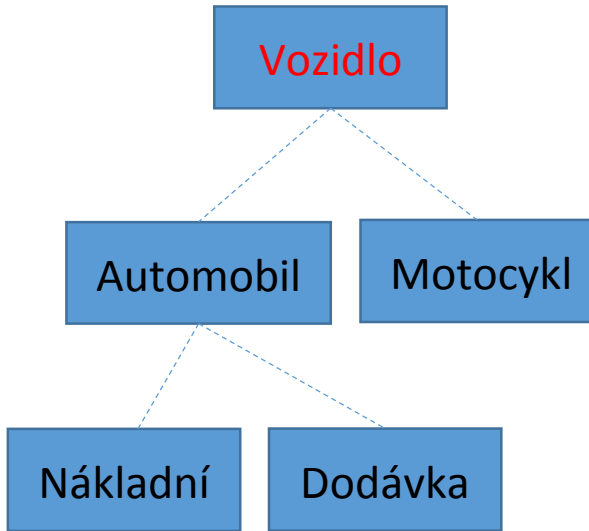
Ontologie B



Kategorie vozidel dle B:

Auto
Dodávka

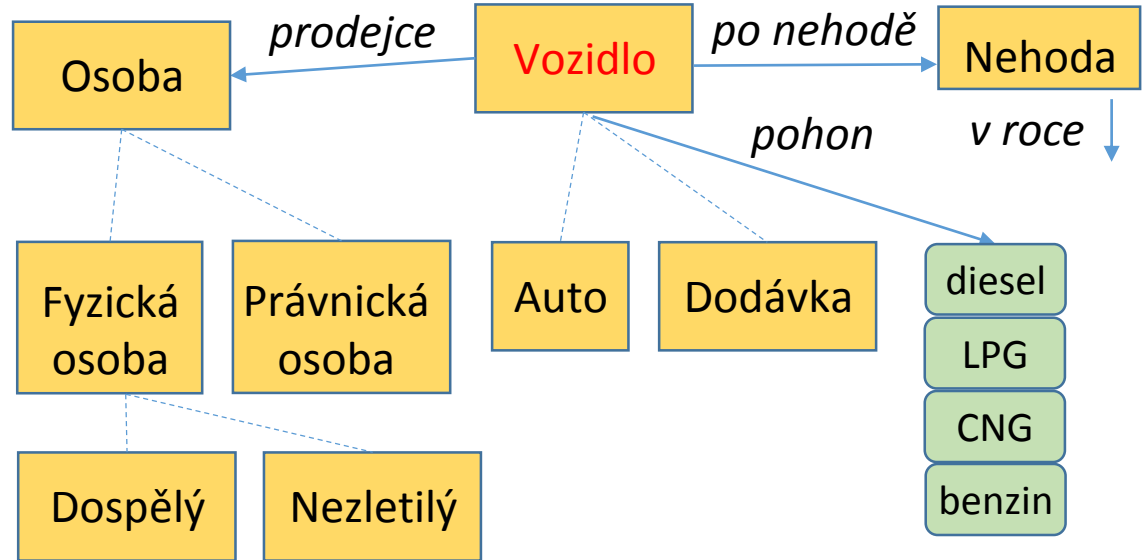
Ontologie A



Kategorie vozidel dle A:

Automobil
Motocykl
Nákladní
Dodávka

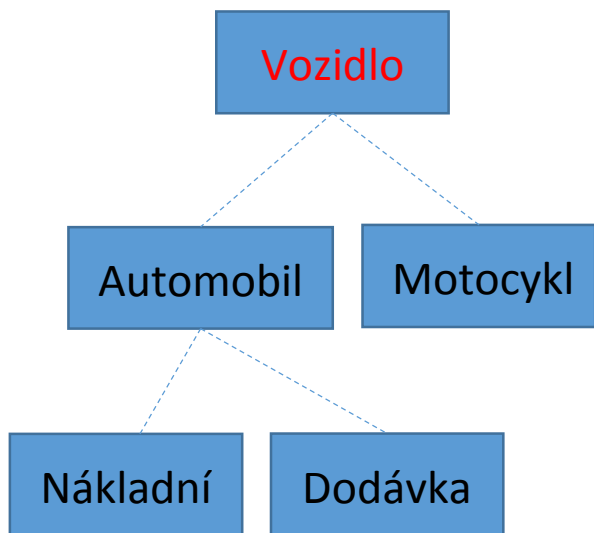
Ontologie B



Kategorie vozidel dle B:

Auto
Dodávka

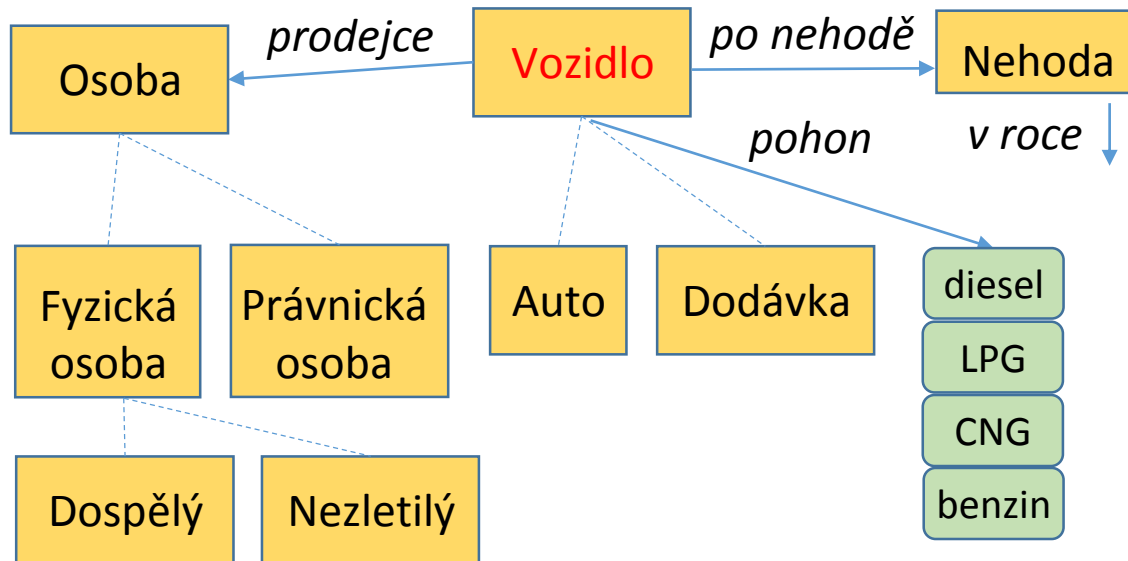
Ontologie A



Kategorie vozidel dle A:

Automobil
Motocykl
Nákladní
Dodávka

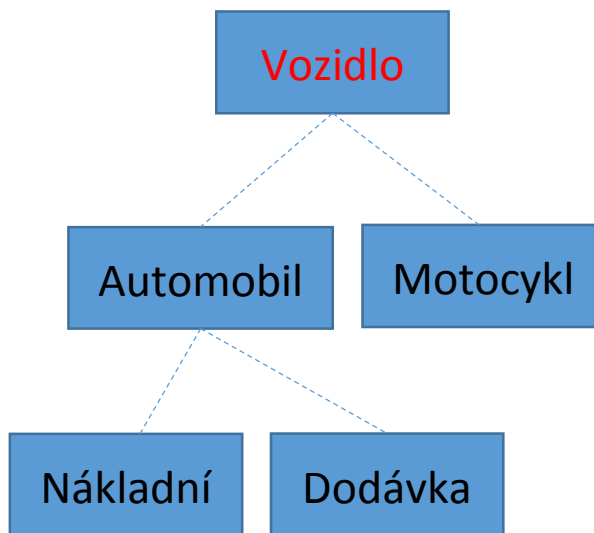
Ontologie B



Kategorie vozidel dle B:

Auto
Dodávka
po nehodě *
pohon = diesel
prodejce je fyzická osoba
prodejce je právnícká osoba

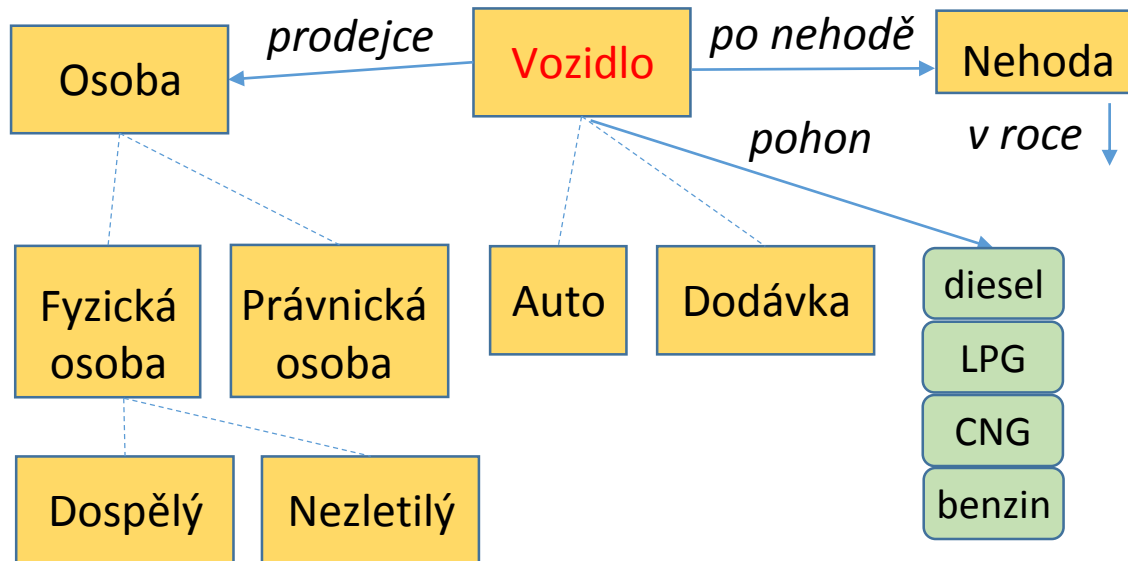
Ontologie A



Kategorie vozidel dle A:

Automobil
Motocykl
Nákladní
Dodávka

Ontologie B

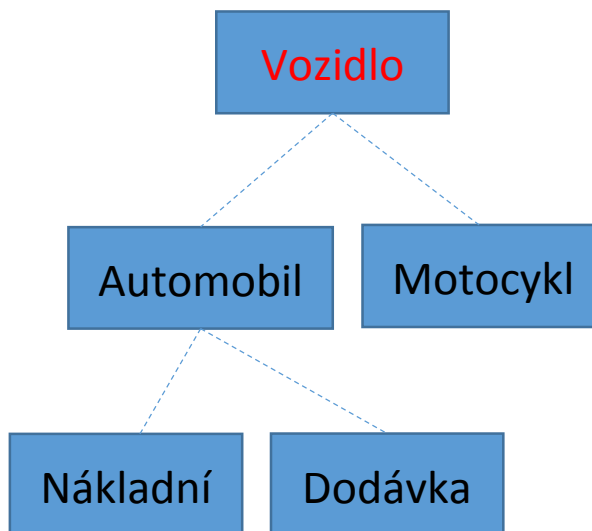


Kategorie vozidel dle B:

Auto
Dodávka
po nehodě *
pohon = diesel
prodejce je fyzická osoba
prodejce je právnícká osoba



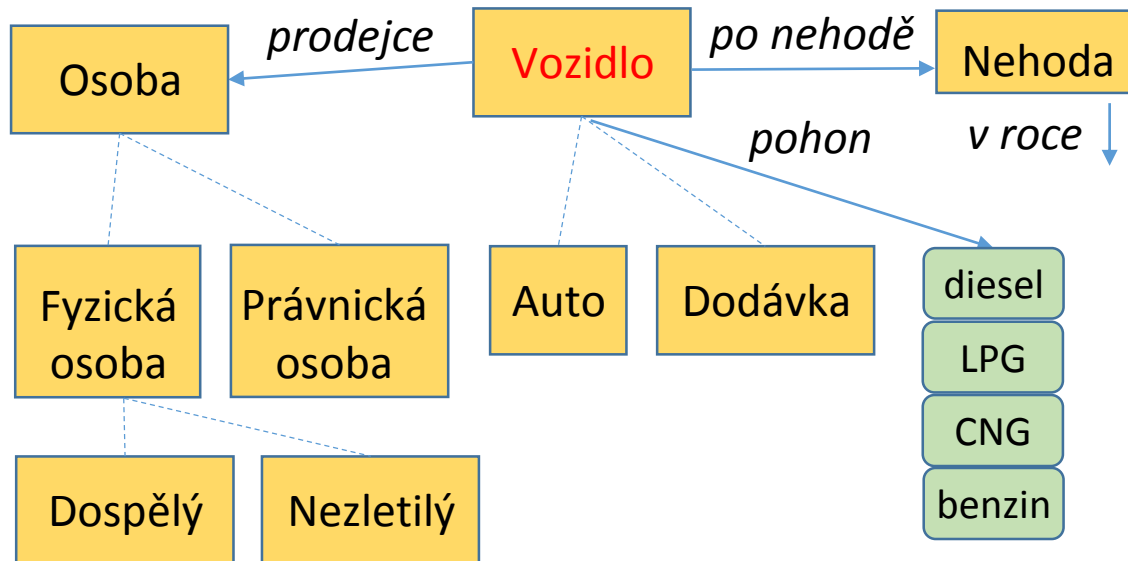
Ontologie A



Kategorie vozidel dle A:

Automobil
Motocykl
Nákladní
Dodávka

Ontologie B



Kategorie vozidel dle B:

Auto
Dodávka

po nehodě *

pohon = diesel

prodejce je fyzická osoba

prodejce je právnícká osoba

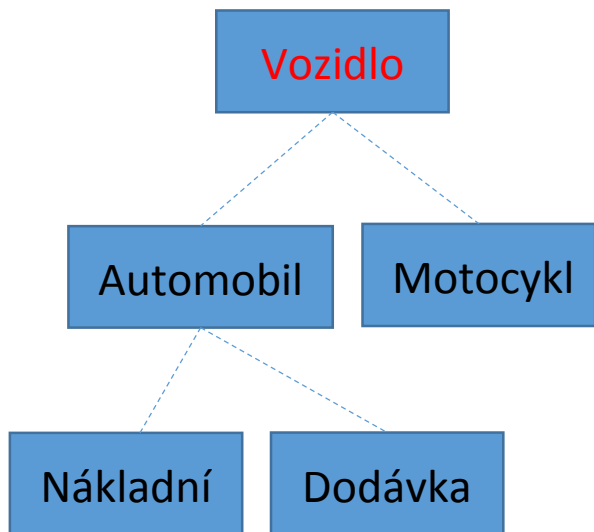


pohon *

prodejce *

prodejce je nezletilý

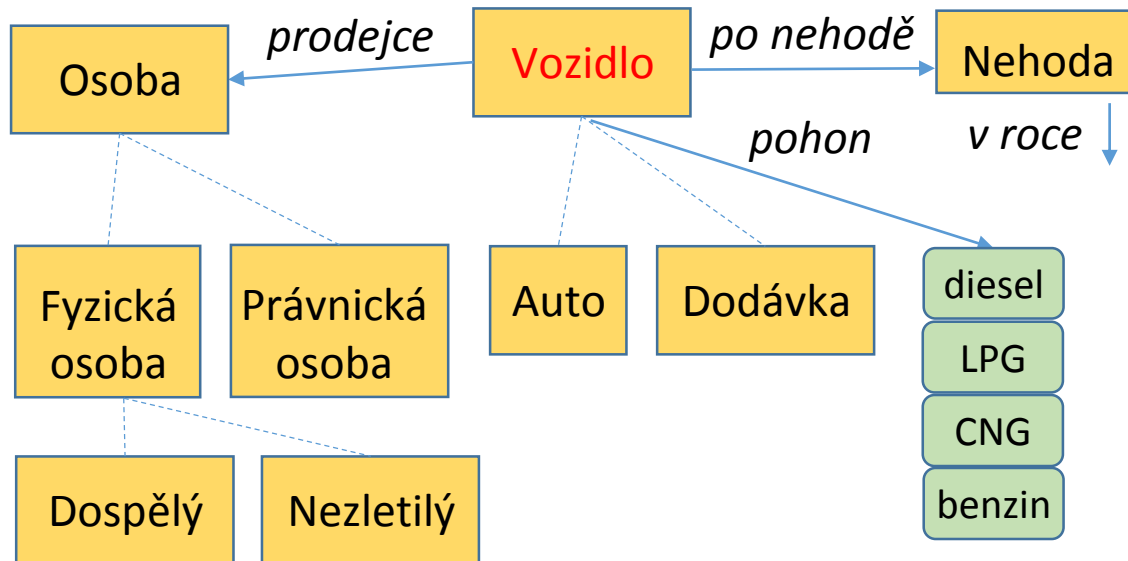
Ontologie A



Kategorie vozidel dle A:

Automobil
Motocykl
Nákladní
Dodávka

Ontologie B



Kategorie vozidel dle B:

Auto
Dodávka

po nehodě *

pohon = diesel

prodejce je fyzická osoba

prodejce je právnícká osoba

Kategorie mohou být i odvozené

pohon *

prodejce *

prodejce je nezletilý

Ale ne všechny jsou stejně užitečné!

Projekt FCatPoWO

- „Focused Categorization Power of Web Ontologies“
- Projekt Grantové agentury ČR (2018-2020)
 - Velmi prestižní: jde o první standardní projekt GAČR získaný „informatickou větví“ FIS od r. 2012
- Rozpočet cca 2,7 mil. Kč
- Objem práce cca 46 člověkoměsíců
- Stálý tým: Svátek, Zamazal, Dudáš, Vacura, Ivánek
- Zapojení doktorandů (Nguyen, Vataščinová), ale potenciálně i studentů magisterského stupně studia (Ize projevit zájem **kdykoliv**)



Návrh projektu IGA (2019-2020)

- Problém nalezení optimálního rozkladu („partitioning“) souboru znalostí pro publikování na webu na
 - schéma (ontologii)
 - třídy (pojmy) uspořádané do hierarchie, případně opatřené logickými definicemi
 - kódovníky
 - ploché, ev. hierarchické; kódy mohou odpovídat jak jednotlivým objektům, tak i pojmům; typicky SKOS
 - znalostní grafy
 - síť vztahů mezi instancemi (reprezentujícími jak jednotlivé objekty, tak i pojmy)
 - příkladem jsou DBpedia, WikiData, Google KG, ...
- Tým: Nguyen, Vataščinová, Svátek, Zamazal, Vacura, (prozatím) 2 magisterští studenti
 - poslední možnost vyjádřit zájem o zapojení: do 5.12. do večera

Relevantní seminář KEG

6. 12. 2018

16:00 - 17:30

473 NB

Václav Jirkovský, Ondřej Šebek, ČVUT Praha:

Utilization of Semantic Web technologies within Industrial Automation Domain

The talk will provide an overview of outcomes which were achieved during the long-term collaboration with Rockwell Automation and Ford Motor Company. The interest of these companies in Semantic Web technologies emphasizes their importance for the future industrial systems. The following outcomes will be covered: Semantic Big Data Historian, Plug&Play concept for cyber-physical systems, diagnostics, and ontology learning for facilitating data model integration.