

DEFINÍCIE ONTOLÓGIE

Ontológia vo filozofii

Slovo „ontológia“ pochádza z gréčtiny – „ontos“ = bytie, „logos“ = náuka, a teda znamená „náuku o bytí ako takom, nezávislom od jeho špecifických druhov“. Ontológia je náuka o existencii objektov individuálnych pojmov chápaných apriórnu interpretáciou jedinca.

Pojem „ontológia“ je známy z dejín filozofie, kde sa okrem iného chápe aj ako systematický záznam existencie. Najjednoduchšia definícia ontológie je táto: ontológia je slovo, ktoré odpovedá na otázku: čo je tu? (Quine, 1977)

Výkladový slovník Webster's (<http://www.m-w.com/home.htm>) definuje ontológiu ako:

- vedu o alebo štúdium bytia;
- špecifické odvetvie metafyziky súvisiace s pôvodom bytia a vzťahmi v ňom (bytím v jeho celku).

Ontológia bola už od Aristotelových čias súčasťou metafyziky. Termín „metafyzika“ vznikol v 1. stor. pr. n. l. ako označenie pre časť Aristotelovho filozofického dedičstva a znamená doslovne „to, čo nasleduje po fyzike“. Sám Aristoteles ju pokladal za najdôležitejšiu časť svojho filozofického učenia „prvou filozofiou“, ktorá skúma najvyššie, zmyslovým orgánom neprístupné, len rozumom postihnuteľné princípy všetkého jestvujúceho, záväzné pre všetky vedy.

Ontológia v informatike

Ontológia je spôsob reprezentácie znalostí o svete alebo jeho časti. Je to **dátový model, ktorý reprezentuje množinu pojmov a vzťahy medzi nimi**. Z explicitne (výslovne) vyjadrených znalostí zaznamenaných v ontológii možno vyvodzovať implicitné dôsledky a súvislosti zahrnuté v ich obsahu (Frank, 2007).

Ontologie sa vo všeobecnosti zapisujú ako **množina definícií formálneho slovníka** (Gruber, 1993). Ontológia reprezentuje jednotky (entity), myšlienky, udalosti s ich vlastnosťami a vzťahmi vzhľadom na systém kategórií.

WWW obsahuje v súčasnosti obrovské množstvo celosvetovo produkováných informácií. Charakterizuje ho fenomén relatívne nízkeho pomeru producentov obsahu k jeho konzumentom. Táto okolnosť sťažuje vyhľadávanie. Jestvujúce metódy klasifikácie jednotlivých informácií je tzv. značenie (*tagovanie*). Pri ňom sa však stráca informácia o význame (sémantike) tagu a klasifikácia je často veľmi subjektívna.

Ontológia obsahuje prostriedky na rozlišovanie sémanticky odlišných a naopak združovanie sémanticky blízkych termínov. Vďaka formálnemu zápisu rozumejú ontológiám aj stroje, čo otvára ďalšie možnosti spracovania dát.

„Pojem „ontológia“ bol ešte zhruba pred 20 rokmi používaný výhradne filozofmi, v súčasnosti sa však vyskytuje napríklad aj v oficiálnych dokumentoch vydávanými

Európskou úniou pre oblasť informatiky a používajú sa databázoví a softveroví špecialisti, ale aj odborníci v oblasti priestorových a časopriestorových informačných systémov (Ivánová).

V nasledujúcom texte sa bude vyskytovať pojem „realita“, resp. „reálny svet“. Vzhľadom na to, že používanie tohto pojmu nie je vždy jednoznačné, treba spresniť význam pojmu realita, resp. reálny svet relevantný pre nasledujúci text. Realitou sa tu chápe jediný časopriestor neobsahujúci žiadne paralelné svety. Každý opis reality je len jeden z mnohých možných a subjektívnych opisov. Jediným dokonalým opisom reálneho sveta je realita sama.

Zásadným problémom pri abstrakcii reálneho sveta sa javí nasledujúci fakt: ako možno vôbec hovoriť o realite bez ontológie samotnej reality? Problém ako definovať pravidlá na definovanie pravidiel je meta-meta-meta-...-meta aktivita. Diskusia na túto tému je možné nájsť v textoch mnohých filozofov (napr. Eco, Tondl, Wittgenstein, Occam, Quine,..).

Všeobecné definície ontológie

Oblasť umelej inteligencie (UI) a reprezentácie znalostí narába s pojmom ontológia v kontexte explicitnej formálnej špecifikácie spôsobu reprezentácie objektov, pojmov a iných prvkov a vzťahov medzi nimi, ktorých existencia sa predpokladá v určitej oblasti záujmu (Gruber, 1993). Pre systémy UI platí, že to, čo existuje, dá sa reprezentovať. To znamená, že množina objektov, ktorá tvorí oblasť záujmu a vzťahy medzi nimi sú zaznamenané v slovníku (katalógu) reprezentácie, ktorým znalostný systém reprezentuje poznanie o realite. Katalóg reprezentácie obsahuje objekty oblasti záujmu (ich názvy, triedy a operácie) a vzťahy medzi nimi opísané laikom chápaným textom*, ako aj formálne definície, ktoré obmedzujú ich interpretáciu. Jednoducho povedané, ontológia v UI je vyjadrenie logickej teórie systému.

Podľa Uitermarka (2001) je ontológia štruktúrovaná, obmedzená množina jednoznačne definovaných pojmov (Mars, 1995, van der Vet and Mars, 1998).

Táto definícia sa skladá zo štyroch častí definujúcich ontológiu:

- ontológia je množina pojmov a nie termínov,
- pojmy musia byť **jednoznačne** definované,
- množina pojmov je **obmedzená** a má
- **štruktúru**, čiže ontológia obsahuje aj vzťahy medzi pojmi.

Keďže jestvuje len jediná realita, existuje len jedna ontológia – objektívna – tá, pomocou ktorej vnímajú všetci ľudia tú jedinou realitu, v ktorej žijú. Niektoré autori rozlišujú „subjektívnu“ a „objektívnu“ ontológiu, objektívnu realitu označujú veľkým písmenom a subjektívnu malým písmenom na začiatku).

Väčšina ontológií je definovaná pomocou slovníkov, resp. katalógov, ktoré vysvetľujú používané termíny. Tento spôsob definície objektov sa niekedy nedá realizovať. Ak sa však ontológia zadefinuje ako množina pojmov pridaním formálnych pravidiel definujúcich obmedzenia v používaní týchto pojmov, stáva sa definícia oblasti záujmu (t. j. ontológia) jasnejšou a jej realizácia je jednoduchšia.

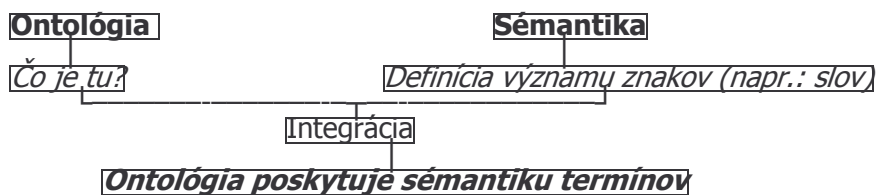
V mnohých prípadoch nemožno definovať ontológiu pre všetky druhy použitia bázy údajov. V tejto situácii je riešením, ak je proces tvorby ontológie štandardizovaný. Štandardizácia prebieha na rôznych úrovniach: medzinárodnej, národnej, odbornej a aplikačnej. Jednotlivé úrovne definície, treba zladit' podľa charakteru vzťahov medzi jednotlivými úrovňami. Ontológiu možno reprezentovať pomocou taxonómie, stromov, katalógov, slovníkov, tezaurov, axióm alebo teorém.

Pojmy, ktoré nie sú obsiahnuté v ontológii nemožno použiť! Väčšina problémov pri budovaní báz údajov, prípadne ich integrácii vzniká práve z dôvodu neakceptovania príslušnej ontológie (napríklad aj pomocou katalógu objektov) ako jedinej novej reprezentácie reality, ktorú báza údajov reprezentuje.

Princípy ontológie majú pri budovaní informačných systémov kľúčovú úlohu, najmä pri integrovaní báz údajov, budovaní užívateľského rozhrania informačných systémov a stanovovaní hodnoty údajov.

Ontológia a sémantika

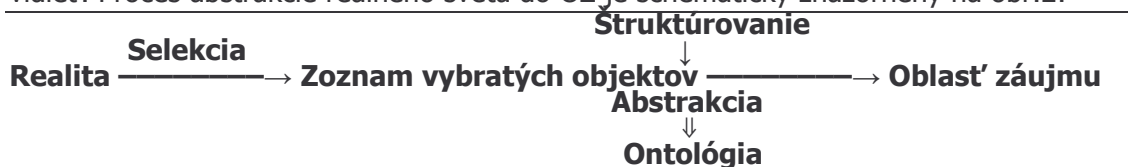
Pri používaní báz údajov je kľúčové poznať význam týchto báz (význam objektov nachádzajúcich sa v bázach údajov). Z pohľadu kvality údajov je sémantická presnosť, resp. správnosť bázy údajov jeden z rozhodujúcich parametrov pri interpretácii údajov. Vzájomný vzťah ontológie a sémantiky je znázornený na obr. 1 (Ivánová).



Obr. 1 Vzťah ontológie a sémantiky

Abstrakcia objektov reálneho sveta – tvorba oblasti záujmu pre bázu údajov

Oblasť záujmu (OZ, z angl. Universe of Discourse) je podľa normy STN EN ISO definovaná ako pohľad na reálny alebo hypotetický svet, ktorý obsahuje všetko, čo skúmajúceho zaujíma. OZ je abstraktným modelom databázy obsahujúcim výlučne objekty záujmu. Tak ako je počúvanie subjektívne, keď každý počuje čo chce, aj pohľad na realitu je ovplyvnený pomyselným zorným poľom, ktoré prepúšťa iba to čo chceme vidieť. Proces abstrakcie reálneho sveta do OZ je schematicky znázornený na obr.2:



Obr. 2 Výber objektov oblasti záujmu

Abstrakcia reálneho sveta do OZ pozostáva z nasledujúcich krokov:

- **selekcia** objektov oblasti záujmu,
- **abstrakcia** vybraných objektov ich opisom, definovaním atribútov a vzťahov medzi objektmi,

- **štruktúrovanie** objektov a vzťahov medzi nimi.

Špecifikácia oblasti záujmu.

Špecifikácia oblasti záujmu závisí od aplikácie a rozumie sa ňou tvorba konceptuálnej schémy a opis obsahu bázy údajov. Tvorba konceptuálnej schémy je analogická tvorbe gramatiky jazyka a opis obsahu bázy je porovnateľný so slovníkom pre konkrétny jazyk. Proces špecifikácie možno rozdeliť na dva základné kroky:

- **morfologická** špecifikácia – definuje tvar a absolútnu, resp. relatívnu polohu objektov;
- **sémantická** špecifikácia – definuje opis objektov a ich atribútov, vzťahov medzi nimi a ich operácie, resp. funkcie. Sémantická špecifikácia objektov bázy údajov obsahuje aj opis ich kvantitatívnych a kvalitatívnych charakteristík.

V procese špecifikácie OZ možno uvažovať napr. aj o špecifikácii vstupu, ktorá definuje proces zberu objektov reálneho sveta do bázy údajov. Súčasťou špecifikácie OZ je aj množina parametrov kvality, ktorú producent očakáva po splnení podmienok pre utvorenie predmetnej bázy údajov.

Tvorba bázy údajov na základe špecifikácie oblasti záujmu.

Po abstrakcii reálneho sveta, výbere objektov do OZ a jej špecifikácii možno pristúpiť k samotnej tvorbe bázy údajov. Proces tvorby bázy údajov pozostáva z nasledujúcich krokov (Aalders, 2000):

- definícia objektov, ich atribútov a vzťahov medzi nimi;
- definícia referenčného systému, používaných jednotiek, rozlíšiteľnosti;
- definícia logických vzťahov v báze údajov – topológie, atribútových obmedzení ap.
- definícia kvality údajov a sémantickej kvality bázy údajov;
- definícia času – používanej časovej sústavy;
- aktualizácia bázy údajov – frekvencia, metódy ap.
- opis pôvodu bázy údajov; možnosti aplikácie – prístup, prípady použitia ap.

Štandardy ISO pre príslušné informácie špecifikujú ontológiu na úrovni definície objektov, ich atribútov a vzťahov medzi nimi. Takisto umožňuje dodatočnú definíciu operácií objektov. Všetky definície by mali byť opísané prirodzeným jazykom a tezaurus (v prípade priestorovej bázy údajov ide napr. o katalóg objektov s definíciami týchto objektov) by mal obsahovať indikáciu povinných, podmienených, resp. nepovinných častí. ISO predpisuje opis ontológie použitím jazyka UML.

Pri tvorbe ontológie interoperabilnej bázy údajov treba splniť tieto podmienky (Aalders, 2002):

opisujú sa iba tie objekty a ich atribúty, ktoré sa nachádzajú v OZ;

- fyzikálne charakteristiky objektov možno opísať pomocou ich atribútov;

- ontológia je aplikačne nezávislá;
- ontológia je flexibilná;
- klasifikácia objektov je platná pre všetky druhy použitia bázy údajov;
- zachovávajú sa vzťahy medzi objektmi.

Konceptuálny model takto vybudovanej bázy údajov musí mať jednoznačne definovaný rámec komunikovateľný aj laikom.

Kvalita bázy údajov

Koncepcia hodnotenia kvality vytvára priestor pre tvorcov aj používateľov bázy údajov pri jej ďalšom používaní. Tvorca hodnotí kvalitu z hľadiska dodržania vopred stanovených kvalitatívnych parametrov v špecifikácii bázy údajov. Používateľ sa na základe opisu kvality údajov môže rozhodnúť pre použitie bázy na svoj konkrétny účel.

Pri modelovaní kvality údajov je nevyhnutné definovanie procesu tvorby bázy údajov od abstrakcie objektov reálneho sveta, cez špecifikáciu oblasti záujmu až po naplnenie bázy údajov. Celý tento proces je v základoch ovplyvnený nasledujúcimi dvoma pravidlami (Uitermark, 2001):

- pravidlami **abstrakcie** – ktoré definujú proces abstrakcie reálneho sveta do bázy údajov (t. j. určujú objekty nachádzajúce sa v oblasti záujmu a žiadne iné);
- pravidlami **modelovania** – na základe týchto pravidiel prebieha naplnenie bázy údajov. Tieto pravidlá obsahujú pravidlá kategorizácie, zjednodušenia, agregácie a reprezentácie objektov reálneho sveta. Taktiež zahrňajú pravidlá opisovania vzťahov medzi objektmi. Pre každý účel tvorby bázy údajov sú definované iné pravidlá naplnenia.

Pre jednoduchosť možno vyššie uvedené rozdelenie zhrnúť: pravidlá abstrakcie opisujú proces selekcie a abstrakcie objektov reálneho sveta do OZ a pravidlá modelovania definujú proces transformácie objektov reálneho sveta na objekty bázy údajov (obr.3).



Obr. 3. Pravidlá abstrakcie reálneho sveta do bázy priestorových údajov

Pri posudzovaní kvality procesu tvorby bázy údajov možno rozlíšiť dva základne druhy kvality:

- kvalitu konceptuálneho modelu – je definovaná sémantickou kvalitou bázy údajov;
- kvalitu operačného modelu – je definovaná súborom parametrov kvality údajov ako sú polohová presnosť, tematická presnosť, časová presnosť, úplnosť, logická konzistentnosť ap.

Sémantická kvalita údajov

Slovo „sémantický“ pochádza z gréčtiny, kde „semanticos“ znamená „významný“. Vo všeobecnosti sémantika znamená vzťahy medzi znakmi a symbolmi a tým čo reprezentujú.

V kontexte modelovania kvality informácií rozumieme pod pojmom sémantika (sémantický) používanie množiny slov integrovaných špecifickou gramatikou, ktorá umožňuje modelovanie reálneho sveta. Sémantickou kvalitou sa rozumie kvalita procesu tvorby OZ a definuje sa ňou správnosť definície údajov, inými slovami – kvalita ontológie. Proces tvorby OZ je z dôvodu narastajúcej potreby výmeny priestorových údajov nevyhnutné štandardizovať. Je známych a používaných niekoľko štandardov, napr. štandardy ISO, OGC, DIGEST, a iné.

Sémantická presnosť

Sémantická presnosť definuje sémantickú vzdialenosť medzi objektmi v báze údajov a tými istými objektmi pozorovanými v reálnom svete uvážením zvoleného referenčného modelu. Vzhľadom na definíciu sémantickej kvality, sémantická presnosť hovorí o relevancii (význame) objektov OZ.

Zámer **tvorca bázy údajov** pri dokumentácii parametrov kvality bázy údajov z hľadiska sémantickej presnosti má podľa Salgého (1995) dve roviny:

- snaha priblížiť používateľom významový obsah bázy;
- snaha zverejniť sémantickú interpretáciu ponúkanej bázy údajov.

Prvú rovinu si možno predstaviť ako dokumentáciu špecifikácie bázy údajov. Sémantická interpretácia bázy údajov nie je nič iné ako dokumentácia sémantickej presnosti a výsledkov jej testovania. Koncept sémantickej presnosti zahŕňa úplnosť, konzistentnosť, aktuálnosť a atribútovú presnosť objektov. Jednotlivé parametre sémantickej presnosti sa skúmajú ako podobné parametre kvality údajov.

Používateľ bázy údajov sa pri posudzovaní sémantickej presnosti riadi podobným princípom ako jej tvorca – v prvom rade je v jeho záujme pochopiť obsah ponúkanej bázy údajov. Zvyčajne sa táto snaha uskutočňuje porovnaním modelu používateľskej aplikácie a modelu tvorca pri tvorbe bázy údajov. Inými slovami, porovnávajú sa pojmy jednotlivých objektov, napr. čo rozumie tvorca pod pojmom „nádor“ a čo pod tým istým pojmom rozumie používateľ. Zvyčajne potom ako používateľ ponúkanej bázy porozumie sa rozhoduje, či táto vyhovuje jeho požiadavkám – či je z jeho pohľadu vhodná na použitie. Používateľ sa rozhoduje na základe súboru parametrov kvality ponúkanej bázy údajov.

Hodnotenie sémantickej presnosti

Podľa vyššie uvedeného (kap. 3.2.1) pojem sémantickej presnosti zahŕňa úplnosť, konzistentnosť, atribútovú presnosť a aktuálnosť (Salgé, 1995). Hodnotenie sémantickej presnosti preto pozostáva z hodnotenia vymenovaných parametrov:

- **úplnosť** – porovnaním bázy údajov s realitou pre daný moment. Tento parameter možno vyjadriť pomocou prebytku alebo vynechania objektov. Z pohľadu sémantickej presnosti sa hodnotí existencia objektov určitej triedy, ich atribútov resp. vzťahov medzi nimi.
- **konzistentnosť** – hodnotenie platnosti sémantických obmedzení, ktoré definuje špecifikácia (napr. trieda objektov C vznikne zlúčením tried A a B) alebo

vyplývajú z reality (napr.: tepny sú cievy, ktoré vedú krv zo srdca na perifériu). Parameter konzistentnosti je vyjadrený pomerom počtu neplatných testovaných prvkov k počtu všetkých testovaných prvkov v báze priestorových údajov.

- **atribútová presnosť** – vyjadruje správnosť priradenia atribútov k objektom bázy údajov.
- **aktuálnosť** – je mierou zmeny objektov bázy údajov za daný čas. Je vyjadrená pomerom vyjadrujúcim zmenu objektov bázy voči úrovni aktualizácie bázy údajov.

Výsledná sémantická presnosť je súhrnom vyššie uvedených parametrov dokumentovaných v časti kvality v metaúdajovom katalógu.

Proces tvorby bázy údajov teda zahŕňa definíciu objektov, abstrakciu reálneho sveta do bázy údajov – tvorby oblasti záujmu a zhodnotenie celého procesu.

Interoperabilita báz údajov je reálna iba v prípade, ak sú tieto vzájomne „sémanticky transparentné“, čiže význam objektov v oboch bázach je jasný. Pravidlá abstrakcie definujú vzťah ontológie a reálneho sveta. Pojmy v ontológii a pravidlá modelovania slúžia na utvorenie referenčného modelu oblasti záujmu. Pravidlá modelovania opisujú vzťahy medzi ontológiou a jej možnými aplikáciami. Sémantiku bázy údajov potom definujú vzťahy medzi pojmami ontológie a pojmami aplikačnej ontológie.

V procese budovania bázy údajov sa majú vytvoriť štandardy pre všetky kroky budovania báz údajov – definíciu objektov (čiže ontológiu bázy údajov) nevynímajúc. Štandardizovaným procesom tvorby bázy údajov možno zabezpečiť homogenitu logických vzťahov v procese jej tvorby a dokumentácie výsledku. Tvorba a používanie štandardov patria k najvýznamnejším prvkom tvorby ontológií.

Ontológia je tu definovaná aj ako opis entít problémovej oblasti, ich vlastností a vzťahov medzi nimi. Ontológia je formálna, explicitná špecifikácia vzájomne vymieňaných pojmov.

Príklad definovania vzťahov

Zápis ontológie – UML, objektový jazyk/reprezentácia, XML, RDF, DAML+OIL, OWL, formálny text

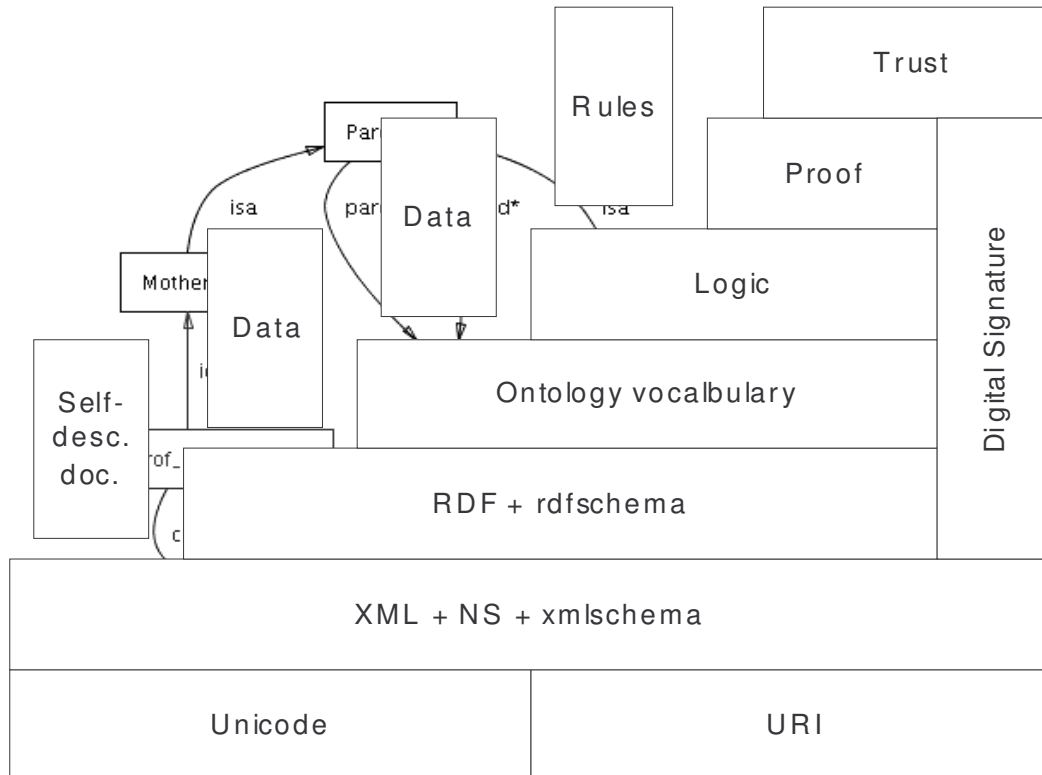
Základný stavebný prvok v **RDF** je atribút, objekt zapísaný v trojici formálne zapisovaný ako A(O,V). Teda objekt O má atribút A s hodnotou V.

child(mother_of_mary_john, John)

RDF Schéma znamená ďalší krok k bohatšej reprezentácii formálneho vyjadrenia a zavádza základné ontologické modelovacie prvky do prostredia webu. Pomocou RDFS sme schopní uvažovať napr. o triedach, podmnožinách vlastností, doménových a rozsahových obmedzeniach vlastností.

OIL – rozšírenie RDFS o kvantifikáciu vlastností, implementačné typy a výlučný výber alternatív, jeho axiomatizáciou vznikol **DAML+OIL**

Rozšírením posledného na web vznikol **OWL**



Architektúra vízie sémantického webu

Jazyk OWL poskytuje tri podmnožiny jazykov:

OWL Lite podporuje predovšetkým používateľov, ktorí potrebujú zatriedovanie do hierarchií a jednoduché možnosti hraničných podmienok.

OWL DL podporuje používateľov, ktorí požadujú maximálnu schopnosť vyjadrenia bez straty výpočtovej úplnosti (garantuje sa, že všetky vyvodenia budú spočítané) a rozhodnuteľnosti (všetky výpočty skončia v konečnom čase) deduktívneho systému.

OWL Full má byť určený pre používateľov, ktorí požadujú maximálnu schopnosť vyjadrenia a voľnosť zápisu RDF, čo má za následok, že nemožno stanoviť žiadne výpočtové garancie.

V tomto kontexte používame ontológie na:

- Potenciálne univerzálnu reprezentáciu – konverziu znalostí
- Slovník, taxonómia, tezaurus
- Deduktívne služby – vyhľadávanie „podobných“ a súvisiacich informácií, sémantický web
- Podporu práce so „všeobecnými“ informáciami

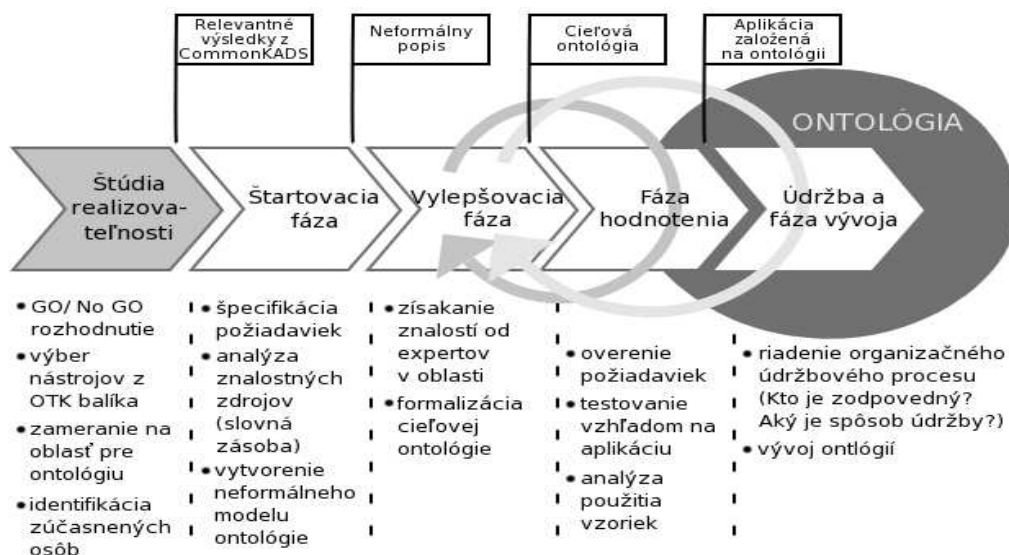
Sémantický web je vízia ďalšej generácie web, v ktorej sémantická pridaná hodnota (markup) bude mať za následok väčšiu dostupnosť zdrojov web k automatickým

procesom. Pri vývoji sémantického webu má kľúčovú úlohu opisná logika (Description Logics, DLs), kde jeho úlohou je poskytnúť formálnu podporu a automatizovaných služieb usudzovania (automated reasoning services) pre jazyky ontológie sémantického webu, ako je DAML + OIL. Opisná logika SHOQ(D), ktorá bola určená na poskytovanie takýchto služieb, sa dá rozšíriť pomocou n -tých datatypových predikátov a kvalifikovaných obmedzení n -tými datatypovými predikátmi pre danú SHOQ(D n) a prezentujú algoritmus na rozhodovanie dostatočnosti (satisfiability) pojmov SHOQ(D n), spolu s dôkazom jeho spoľahlivosti a úplnosti. Práca bola motivovaná požiadavkami na n -té datatypové predikáty a kvalifikovaný počet obmedzení s n -tými predikátmi vo vzťahu k vlastnostiam „reálnemu svetu“ v ontológiách a aplikáciách sémantického webu.

Ontologie v porovnaní s objektovými modelmi (UML) kladú väčší dôraz na presnejšie postihnutie sémantiky reálneho sveta, a naopak menší na efektivitu spracovania dát. Ich strojové využívanie má charakter odvodzovania formúl v presne vymedzenom logickom kalkule. Pre objektovo-orientovaného vývojára je tvorba ontológií spojená s prekvapivými črtami. Článok se okrem všeobecného porovnania obidvoch typov modelov zameriava na oblasť sémantického webu, v koncepcii ktorého majú ontologie kľúčovú úlohu, a pokúša sa vymedziť vzájomné postavenie obidvoch typov modelov v tomto kontexte.

Približne od 70. rokov sa v širokom priestore informatiky začali kryštalizovať komunity vychádzajúce z tej istej koncepcie – zachytenia štruktúry reálneho sveta pomocou tried usporiadaných do hierarchie, ktorým možno priradiť objekty alebo inštancie.

Životný cyklus ontológie



Používané sú nasledujúce metodológie:

- Kostrová metodológia – podniky
- KACTUS – znovupoužitie
- Methontology – metaúroveň

Podpora priamo z RDF a ontologických editorov – Protégé, OntoEdit, OilEd, DAML UML Enhanced Tool, KAON OI modeler, Ontolingua, ...

Základné pojmy ontologického inžinierstva

danosť → entita, ktorá existuje bez nášho pričinenia.

entita – (z g *ens* súcno), základný pojem scholastiky. Entita je → súcno nejakej veci (objektu, udalosti, procesu), jej súcnosť, to 'že' na rozdiel od 'čo'. Je to hocičo, o čom možno mať informácie, o čom možno niečo vedieť, čo možno predpokladať, domnievať sa, že má nejaké vlastnosti, čo možno opísať, čo môže byť predmetom našich úvah. Je to niečo existujúce, jestvujúce, nejaká vec, bytosť, bližšie neurčený útvar skutočnosti/vedomia, nie je napríklad určené ani to, či je naozaj skutočný, predpokladaný, domnelý a pod. Entitou je napr. významový útvar, významový proces, stránka materiálnej skutočnosti, ľudské individuum, Slnko, vzťah, proces, bylina, anjel ap. Entity môžu existovať naším pričinením alebo bez neho. Entity existujúce bez nášho pričinenia sú danosti. Najdôležitejším typom entít typom vo filozofii sú filozofémy (štruktúrne prvky filozofie, napr. tvrdenia, učenia, teórie). Na entitách študujeme najmä: **1.** momentálny prierez ich štruktúrou, **2.** genetickú a dynamickú dominantu ich štruktúry; **3.** nevyhnutné vonkajšie súvislosti ich štruktúry. Entita je súcno nejakej veci, jej súcnosť, to 'že' na rozdiel od 'čo'. Pojem entity je základný pojem scholastiky.

Pretože bytie vystupuje ako základ (logos) súcna (on), nazýva Heidegger rozlišovanie bytia a súcna ontologickou diferenciou. Súcno nie je totožné so zmyslovo vnímateľným (ako sa domnievali pozitivistí) a netvorí len osobitnú oblasť vedľa iných, postavených jej na roveň, ako to hlása pluralizmus a filozofia hodnôt.

ontológia vo filozofii – náuka o bytí; univerzálna sústava znalostí opisujúca objekty, javy a zákonitosti sveta „tak ako je“ (maximálne nezávisle od ľudského usudzovania o ňom)(Svátek, 2002).

ontológia v informatike – predmet praktického ontologického inžinierstva, t. j. ako „informačný artefakt“, opisuje to, čo „existuje“ a dá sa teda reprezentovať v informačnom, resp. znalostnom systéme. Podľa jedného z „duchovných otcov“ ontológií T. Grubera (1993): ontológia je explicitná špecifikácia konceptualizácie. W. Borst modifikoval túto definíciu ako „...formálnu špecifikáciu zdieľanej konceptualizácie. V Gruberovej definícii sa požaduje len to, že konceptualizácia musí byť špecifikovaná *explicitne*, t. j. nielen „skrytá“ v hlave svojho autora. V Borstovej definícii vystupuje už požiadavka na formalizáciu (použitie jazyka s presne definovanou syntaxou (príp. aj sémantikou), ako aj vzájomná vymieňateľnosť („zdieľanosť“) – ontológia nie je individuálnou záležitosťou, ale je výsledkom konsenzu určitej záujmovej skupiny ľudí. Tieto dve dodatočné požiadavky sa však už striktno nedodržiajú, napr. znázornenie štruktúry tried a relácií pomocou diagramu sa samo osebe niekedy pokladá za ontológiu a práve tak sa za ontológiu pokladajú aj novo vzniknuté modely, ktoré dosiaľ neprešli kolektívnou diskusiou, najmä pokiaľ syntakticky vyhovujú formálnemu jazyku.

Ontológiu možno chápať aj ako *znalostné modely*, teda abstraktné opisy (určitej časti) znalostného systému, ktoré sú relatívne nezávislé od finálnej reprezentácie a implementácie znalostí.

Ontológie umožňujú:

- podporu porozumenia medzi ľuďmi (napr. expertmi a znalosťnými inžiniermi)
- podporu komunikácie (interoperability) medzi počítačovými systémami
- uľahčenie návrhu znalostne orientovaných aplikácií.

Za hlavné oblasti využitia ontológií sa pokladajú:

- Znalostný manažment v organizácii. Pre efektívne fungovanie zdravotníckych zariadení treba zabezpečiť, aby sa informácie a znalosti (interného, ako aj externého pôvodu) nestrácali, a včas sa dostávali k pracovníkom, ktoré ich môžu využiť. Pomocou ontológií možno zachytiť vecnú podstatu znalostí, a tým zabezpečiť ich konzistenciu a uľahčiť ich vyhľadávanie.
- Elektronická komunikácia pacient–lekár, zdravotníckych zariadení, kontakt poskytovateľov a používateľov zdravotníckych služieb. Ontológia môže uľahčiť vyhľadanie zdravotníckych služieb, liekov a pomôcok ap.
- Spracovanie prirodzeného jazyka – terminologické ontológie môžu napomôcť napr. pri preklade alebo automatickej sumarizácii textov.
- Inteligentná integrácia informácií – ontológia môže slúžiť ako zastrešenie dátových schém distribuovaných zdrojov (štruktúrovaných alebo semištruktúrovaných databáz), príp. „tabulárnych“ webových stránok) na vysokej úrovni abstrakcie
- Pojmové vyhľadávanie informácií ako vylepšenie jestvujúcich internetových vyhľadávačov.
- Sémantické webové portály konštruované poloautomaticky na základe metadát od poskytovateľov informácie.
- Inteligentné výučbové systémy.

Typy ontológií

Ontológia je súbor heterogénnych informačných artefaktov. Dajú sa rozdeliť podľa historických paradigiem, miery a predmetu formalizácie.

Podľa *historických paradigiem* sa rozlišujú:

- A. Terminologická (lexikálna) ontológia**, ktorá sa dá stotožniť s pokročilými tezaurami používanými v knihovníctve a ďalších odboroch orientovaných na textové zdroje. Charakterizuje ich ústredná úloha termínov, ktoré už nie sú ďalej formálne definované. Používané relácie majú zväčša taxonomický charakter (vymedzenie všeobecnejšieho a špeciálnejšieho termínu), okrem toho býva vyjadrená synonymia, meronymia (vzťah termínov označujúcich celok a jeho časť) a ďalšie relácie všeobecného charakteru. Najznámejšia terminologická ontológia je

WordNet; z nej sa odvodil napr. *Sensus* alebo viacjazykový variant *EuroWornNet*.

- B. Informačné ontológie** sú rozvinutím databázových konceptuálnych schém. Majú úlohu nadstavby nad primárnymi (štruktúrovanými, napr. relačne - databázovými) zdrojmi, pre ktoré zabezpečujú pojmovú abstrakciu potrebnú na pojmové dotazovanie, jednak vyššiu úroveň kontroly integrity ako bežné nástroje.
- C. Znalostné ontológie** nadväzujú na výskum v oblasti reprezentácie znalostí v rámci umelej inteligencie. Ontológie sa tu chápu dôsledne ako logické teórie, a ich väzba na reálne objekty (inštanície) je v porovnaní s informačnými ontológiami relatívne voľná. Triedy a vzťahy sú systematicky definované prostredníctvom formálneho jazyka.

Z hľadiska formalizácie, ktorá je do istej miery definičnou vlastnosťou ontológií, využívajú sa aj neformálne ontológie. Ide napr. o glosáre, v ktorých sú jednotlivé pojmy vysvetlené prirodzeným jazykom.

Podľa predmetu formalizácie sa rozoznávajú:

- D. Doménové ontológie** – sú najčastejšie. Jej predmetom je vždy určitá špecifická vecná oblasť, vymedzená širšie (celá oblasť medicíny) alebo užšie (problematika určitej choroby). Príkladom doménových ontológií je lekárska *On9*.
- E. Generické ontológie** – zachytávajú všeobecné zákonitosti, ktoré platia naprieč vecným oblastiam, napr. problematiky času, vzájomnej polohy objektov (topológia), skladby objektu z častí (mereológia) ap. Niekedy sa vyčleňujú tzv. ontológie vyššej úrovne („upper-level“), ktoré sa usilujú zachytiť najvšeobecnejšie pojmy a vzťahy, ako základ taxonomickej štruktúry každej ďalšej (napr. doménovej ontológie).
- F. Úlohové ontológie** – označujú generické modely znalostných úloh a metód ich riešení. Na rozdiel od ostatných ontológií, ktoré zachytávajú znalosti o svete („tak, ako je“), sú zamerané na procesy odvodzovania. K úlohám takýchto znalostných modelov patrí napr. diagnostika, hodnotenie, konfigurácia alebo plánovanie.

Štruktúra ontológií

Základom znalostných ontológií sú *triedy*, ktoré označujú množiny konkrétnych objektov. Termínu trieda niekedy zodpovedá termín *pojmem*, príp. *kategória* a úzko súvisí aj s termínom *rámec* ako základným konštruktom mnohých systémov umelej inteligencie.

Na rozdiel od tried v objektovo-orientovaných modeloch a jazykoch nezahŕňajú „ontologické“ triedy procedurálne metódy. Ich interpretácia je skôr odvodená z pojmu vzťah, v tom zmysle, že trieda zodpovedá *unárnej (jednotkovej) relácii* na danej doméne objektov. Triedy, pre ktoré sú *špecifikované podmienky nutnosti i postačiteľnosti* (príslušnosti individua) sa označujú ako *definované*, ostatné triedy (pri ktorých sú špecifikované len nutné podmienky alebo ani tie) ako *primitívne*.

Na množine tried býva definovaná hierarchia (taxonómia). Napriek tomu, že „filozofický“ smer niekedy zdôrazňuje požiadavku stromovej štruktúry, v praxi všetky hlavné ontologické jazyky podporujú viacnásobnú dedičnosť, a táto možnosť sa v praxi hojne využíva.

Indivídua, inštancie

Indivídium (angl. „individual“) zodpovedá konkrétnemu objektu reálneho sveta, a je do istej miery „protipólom“ triedy. Termín *inštancia* sa často chápe ako ekvivalentný, asociuje však príslušnosť k určitej triede, čo nemusí byť nutne skutočnosťou – indivídium sa môže do ontológie vložiť „provizórne“ aj bez väzby na triedu.

Treba si však uvedomiť, že ontológia principiálne slúži na opis pojmov (tried) a nie faktov o konkrétnych objektoch. Niektoré jazyky preto indivíduá ako súčasť ontológie interne nepodporujú alebo pojmy a indivíduá zahŕňajú do spoločného nadradeného modelu (*ontology-instance model*, OIM), zložený z ontológie a jej zodpovedajúce bázy inštancie. Takýto prístup sa približuje databázovým schémam (t. j. informačným ontológiám).

Podobne ako v databázových modeloch sú podstatnou zložkou ontológií *vzťahy* („relations“). V tradičných jazykoch môžu byť pomenované vzťahy (ako triedy, ktoré sú ich špeciálnym prípadom) špecifikované pomocou ľubovoľných logických podmienok. V niektorých jazykoch možno im len priradiť preddefinované obmedzenia.

Pre binárne vzťahy (na ktoré sa obmedzujú „odľahčené“ jazyky) sa používa pojem *slot*, vzniknutý v oblasti znalostných systémov založených na rámcoch, príp. *vlastnosť* („property“). Na rozdiel od objektovo-orientovaného prístupu sú však sloty (vzťahy) v ontológii „občanmi prvej kategórie“. Vo vzťahu k predmetu, ktorý je jeho hodnotou, sa niekedy slot označuje ako *rola*. [Např. tvrdenie „hodnotou slotu matka pre predmet *Oskár* je predmet *Alexandra*“ možno vyjadriť ako „predmet *Alexandra* je v role matka voči predmetu *Oskár*“.]

Ako osobitný typ vzťahu sa chápu *funkcie*; v súlade s bežným matematickým chápaním ide o vzťahy, pri ktorých je hodnota n -tého argumentu jednoznačne určená predchádzajúcimi $n-1$. Funkčný slot sa označuje aj ako atribút, predpokladá sa, že je definovaný pre všetky inštancie triedy.

Na rozdiel od vzťahov v predikátovom kalkule 1. rádu (predikát = vlastnosť predmetu) možno v ontológiách priradovať vlastnosti, tzv. *metasloty*. Najčastejším metaslotom je hierarchický vzťah podradeného a nadradeného slotu.

Obmedzenia, ktoré sa vzťahujú na slot bez ohľadu na spôsob jeho použitia sa označujú ako globálne. Často však treba vymedziť hodnotu slotu aplikovaného na konkrétnu triedu zo svojho definičného oboru. Takéto lokálne obmedzenia („slot constraints“) sa označujú ako *fazety*.

Formálna reprezentácia ontológií

V súčasnosti sa používa mnoho formálnych jazykov. Jedným z prvých (dosiaľ „živých“) jazykov je projekt Cyc, ktorého názov je odvodený od slova „enCYClopedia“ začatý r. 1984 pod vedením D. Lenata. Tento projekt sa usiluje o zhromaždenie všeobecných znalostí („common sense“), ktoré by boli v znalostných systémoch komplementárne k expertným znalostiam a zabraňovali absurdnému správaniu. Cyc používa svoj

vlastný jazyk Cycl kombinovaný o. i. s prkami rámcových jazykov. Od r. 2001 je projekt voľne prístupný a ponúka už 6000 pojmov a 60 000 tvrdení o nich.

T. Gruber a spol. zo stanfordskej Knowledge System Laboratory vyvinuli na začiatku 90. rokov jazyk Ontolingua (nadstavba jazyka Knowledge Interchange Format) , ktorý obsahuje konštrukty, ako je definícia tried, vzťahov a funkcií.

Obmedzené možnosti odvodzovania v jazyku Ontolingua motivovali E. Mottu z Open University vo Veľkej Británii k návrhu jazyka, ktorý by výraznejšie podporoval priamy vývoj programových aplikácií bez toho, aby bolo treba model prekladať do iného jazyka. Vyvinul jazyk modelujúci pojmy (Operational Conceptual Modelling Language, OCML), ktorý prepojil s tvorbou jeho interpreteru, implementovaného v prostredí CommonLISP. Tento jazyk sa však príliš nerozšíril mimo Open University.

K ďalším jazykom z 90. rokov patrí aplikačné rozhranie umožňujúce otvorenú komunikáciu medzi rámcovo-orientovanými znalostnými systémami (Open Knowledge Base Connectivity, OKBC).

R. 1999 sa stala východiskovým bodom pre návrh ďalšieho ontologického jazyka: eXtensible Ontology Language (XOL) jednoduchá verzia OKBC-Lite. Jeho motiváciou bola potreba bioinformatickej komunity zdieľať štruktúry znalostí o génovom výskume.

K novým „webovým“ ontologickým jazykom patria: Resource Description Framework (RDF), Ontology Inference Lazer (OIL), DARPA Agent Mark-up Language (DAML) a Web Ontology Language (OWL)

V oblasti databázových, resp. textových informačných systémov výsadné postavenie majú UML a TopicMaps.

Editors ontológií

Napriek tomu, že možno prakticky všetky ontologické jazyky spracúvať bežným textovým editorom (a tie novšie editorom XML), ich rutinné používanie je ťažké bez špecializovaného editora. (Protégé, OntoEdit, OilEd, DAML UML Enhanced Tool, KAON OI modeler, Ontolingua, ...)