

Prehľad súčasných terminológií a ontológií v biológii a medicíne

F. Freitas, S. Schulz

Elect. J. Commun. Inf. Innov. Health (Rio de Janeiro, 3, 2009, č. 1, s. 7 – 18.

Práca poskytuje prehľad o stav term terminológií a ontológií používaných v biológii a medicíne. Bez nároku na úplnosť sa opisujú niektoré významnejšie zdroje, ktoré upútavajú v súčasnosti pozornosť priemyslu a vedeckých kruhov. Po úvodnom opise rámca sa porovnávajú systémy z hľadiska prvkov ich architektúry, ich výpovednej schopnosti, rozsahu a analyzuje sa charakter ich entít, ktoré vyjadrujú. Konkrétne išlo o prehodnotenie Medzinárodnej klasifikácie chorôb (International Classification of Diseases, ICD), Kľúčových slov medicíny (Medical Subject Headings, MeSH), génovej ontológie (gene Ontology, GO), Systematizovanej nomenklatúry medicíny – klinických termínov (Systematized Nomenclature of Medicine – Clinical Terms, SNOMED CT), Všeobecnej architektúry jazykov, encyklopédií a nomenklatúr (Generalized Architecture for Languages, Encyklopaedias and Nomenclatures – openGaALEN), Výstavbového modelu anatómie (Foundational Model of Anatomy, FMA), Systém zjednoteného lekárskeho jazyka (Unified Medical Language System, UMLS) a Otvorených biomedicínskych ontológií (Open Biomedical Ontologies, OBO).

Narastajúca dostupnosť ohromného množstva zdrojov biomedicínskych údajov a znalostí zaťažujú vedeckých pracovníkov i praktických lekárov úlohou spracúvať terabajty sémantických obsahov, ktoré už svojou povahou navzájom súvisia a treba ich zoskupovať a manipulovať s nimi. Záplava dát používaná na riešenie zložitých úloh vyžaduje stále sofistikovanejšie metódy inteligentného spracovania informácií a znalostí, zvyšujúce interoperabilitu obsahu vo veľkých úložiskách podporovaných rozličnými typmi automatizovaného logického spracovania (reasoning). S touto výzvou stále viac prichádzajú biológovia, klinickí a verejnozdravotnícki vedeckí pracovníci, zdravotnícki ekonómovia, ale aj klinickí praktici. Praktickým výsledkom toho snahy je rastúci počet sémantických referenčných systémov charakterizovaných často ako slovníky, tezaury, terminológie a ontológie (Rubin, 2007).

Súčasný vývoj v manažmente biomedicínskych znalostí má v podstate dva korene:

- tvorba indexujúcich slovníkových a klasifikačných systémov, ako je Medzinárodná klasifikácia chorôb a *Index Medicus*, ktorý sa datuje od 19. storočia a o ktorý je predmetom záujmu na jednej strane verejné zdravotníctvo a epidemiológia a na druhej strane knihovnícka veda
- výskum podpory lekárskeho rozhodovania a expertných systémov, ktorý sa začal v 70. rokoch minulého storočia, ktorý stimuluje výskum v oblasti umelej inteligencie, inšpirovaný ideou vytvoriť počítačové prostriedky založené na vedomostiach s cieľom asistovať pri zložitom procese lekárskeho rozhodovania.

Motivovaním víziou sémantického webu sa stal termín „ontológia“ jedným z najmodnejších termínov počítačovej vedy. Ontológovia propagujú presný opis domén do podrobností a používanie opisy v mnohých typoch aplikácií, počnúc spracúvaním prirodzeným jazykom po systémy logického uvažovania a podpory rozhodovania. Mnohé aplikačné oblasti v súčasnosti ťažia z výhod ontológií, ale najzreteľnejšie sa ich úžitok prejavuje vo vedách o živote, pretože len málo vedeckých oblastí zaznamenáva taký výrazný a rýchlo rastúci objem termínov, pojmov a definícií.

Ontológie

Termín „ontológia“ sa stal veľmi populárny od polovice 90. rokov, neexistuje však, žiaľ, jeho všeobecne akceptovaná definícia (Kuzniersky, 2006). Od 70. rokov sa používal pre odbor

všeobecná metafyzika v tradičnom zmysle Aristotelovej „prvej filozofie“, ako vedy o bytí (*being qua being* = súcno ako také). Často vystupuje komplementárne k pojmu epistemológia (náuka o poznaní, gnozeológia).

V počítačovej vede prevažuje definícia ontológie ako explicitná (výslovná) špecifikácia konceptualizácie (Gruber, 1995). Konceptualizáciou sa tu rozumie abstrakt, zjednodušený pohľad na svet, ktorý chceme reprezentovať na určitý účel, napr. vyvodenie záverov, utvorenie automatickej klasifikácie ap. Konceptualizácia často zahŕňa *pojmy* (zvané aj triedy alebo typy, napr. srdce), *individua* ako príklady („instances“, príklady, typy, druhy) pojmov (napr. individuum Fido je príkladom psa), *binárne vzťahy* medzi pojmami a individuami (napr. pes je stavovec), *obmedzenia založené na logike* (všetky druhy bylinožravcov sa živia len rastlinnou stravou, kým všetky druhy mäsožravcov sa živia niektorými druhmi živočíchov) a *axiómy* (výroky, ktoré sú v danej oblasti vždy pravdivé, napr. každý typ žijúcej osoby má určitý typ srdca). Ako spájadlo, ktoré dáva do súvislosti tieto entity, slúžia zrejme ontologické vzťahy (relácie). Predstavujú rozličné aspekty vzájomných súvislostí jednotlivých pojmov. Najdôležitejším a najpoužívanejším typom vzťahu je *podtrieda* (subclass, napr. srdce je podtriedou orgánu, pretože všetky druhy srdc sú druhom orgánu, s určitými špeciálnymi charakteristikami, ktoré odlišujú jednotlivé typy srdc), a *partonomické vzťahy* (každý typ srdcovej komory je časťou nejakého srdca). Jestvujú však iné definície ontológie, napr. „ontológia reprezentácia predmetnej oblasti, pozostávajúca zo zoznamu termínov, vzťahov medzi nimi a axióm, ktoré sú vždy v danej oblasti platné“ (Antoniou a Harmelen, 2004), alebo „ontológia je reprezentačný artefakt, ktorého reprezentačné jednotky sú určené na označenie tried alebo univerzálií (všeobecných pojmov) v realite a ich vzájomných vzťahov (Smith, 2005).

Predstava ontológie sa často špecializuje a označuje ako „formálna ontológia“ (Guarino, 1998). To znamená, že obsah ontológie sa opisuje pomocou matematickej logiky, ktorej rozumejú počítačové systémy schopné z nich vyvodzovať logické závery. Môže tiež podporiť vyvodzovanie autonómnych zistení zo zaznamenaných dát, ako aj opätovné využitie a výmenu znalosti.

Uplatnenie ontológií v počítačovej vede ako hlavnej tendencie sa rozšírilo aj na iné oblasti poznania: Motivovaní víziou sémantického webu (Berners-Lee, 2001) sa mnohé skupiny z akademických kruhov a priemyslu na celom svete začali zaujímať o ontológie a úmerne tomu narastali počet prostriedkov, štandardov a používateľov. V niektorých oblastiach, najmä v medicíne a biológii sa vyvinulo skutočne veľa úsilia utvoriť štandardné ontológie.

Terminológia a ontológia

Medicína je špeciálne charakterizovaná množstvom tzv. terminológií, najlepšie opísateľných ako jazykovo orientované artefakty, ktoré uvádzajú do vzájomnej súvislosti rozličné zmysly a významy jazykových jednotiek (entít). Terminologie sa vo všeobecnosti tvoria s cieľom slúžiť presne definovanému účelu, ako je vyhľadávanie dokumentov, anotácia (krátka charakteristika) zdroja, záznamy štatistiky chorobnosti a úmrtnosti alebo fakturácia zdravotníckych služieb. Biomedicínske terminologie nepoužívajú formálne a presne definované opisy; definujú skôr termíny (ak vôbec) výrazmi ľudskej reči a vyjadrujú asociácie medzi termínmi pomocou neformálnych vzťahov blízkych ľudskej reči. Slová a viacslovné termíny sú základnými stavebnými blokmi terminológií, ktoré ich organizujú vo všeobecnosti do hierarchií umožňujúce vyjadriť ich vzťahy vo forme synonymie (slová toho istého významu), hyperonymie (slová širšieho významu), hyponymie (slová užšieho významu). Hoci terminologie sa dajú úspešne použiť na reprezentáciu abstraktných významov, napr. pri spracovaní prirodzeného jazyka alebo pri anotácii zdrojov (napr. literárnych abstraktov, výsledkov experimentov), nie sú presné a dostatočne výrečné pre hlbšie znalostné aplikácie.

Kým jeden prípad použitia (use case) vyžaduje znalosť o tom *ako* a *čím* sa líšia niektoré termíny od iných, iné prípady použitia vyžadujú presnejšie vzťahy medzi termínmi, napr. že každý typ normálneho hornej končatiny (= arm) má určitý typ predlaktia ako svoju časť. Na splnenie týchto

požiadaviek nie je jazykovo orientovaný (language-centered) zdroj dostatočne zrozumiteľný. Lepšie sa tu hodí zdroj orientovaný na realitu (reality-centered resource), v ktorom sa entity (objekty, vlastnosti, procesy ap.) nachádzajú za určitých okolností vo vzájomnom vzťahu, takže sa dajú zachytiť detaily týchto vzťahov, za akých okolností tieto vzťahy uplatňujú a ako sa tieto vzťahy dajú presne interpretovať (napr. či vzťah „časť niečoho“ medzi časťami tela a telom jestvuje aj potom, keď sa časť tela ako sú obličky, odstráni). To je prípad, kde sa uplatňujú ontológie. Ontológie sa vyjadrujú vo formuláciách založených na logike, ktorá poskytuje (meta)definície tried (pojmov), vzťahov, príkladov a axióm. Ontológie môže preto reprezentovať doménu vo forme, v ktorej je počítač schopný spracovať definície namiesto používania len termínov alebo sémantických identifikátorov. Systém tak môže kontrolovať, či niektoré interpretácie sú správne alebo nie, keď je niektorý výrok pravdivý podľa niektorej ontológie, medzi inými podobnými úlohami. Ontológie môžu obsahovať aj rozličné dimenzie, ktoré má doména zahŕňať: napr. v organizme stupeň pravosti (autenticity) orgánov (či sa predpokladajú nejaké funkcie organizmu obyčajne prítomné alebo nie), štádium vývoja (napr. zárodok proti dospelému jedincovi), miesto organizmu alebo organickej látky v biologickej taxonómii (napr. mucha proti myši) alebo granularita, pomocou ktorej sa opisuje biologická štruktúra (makroskopická proti mikroskopickej), aby sme spomenuli aspoň niektoré (Schulz, 2004).

V klasickom prístupe k terminológii však jestvuje zvyšujúci sa zmätok týkajúci sa princípov moderného dizajnu ontológie, ontologických jazykov z oblasti počítačovej vedy a vznikajúce disciplíny aplikovanej ontológie z oblasti analytickej filozofie.

Treba poukázať na široký rozsah veľmi heterogénnych artefaktov, pre ktoré zatiaľ prekleňujúci termín chýba (často používaný termín „biomedicínske slovníky“ je zavádzajúci, pretože príliš zdôrazňuje jazykový aspekt). Navrhne sa preto používať pre „biomedicínske terminológie a ontológie“ akronym BMTO. V ďalšej stati sa podrobne vysvetľujú BMTO, stať 3 je venovaná základom a úsiliu integrovať mnohé z týchto systémov, v stati 4 sa diskutuje o niektorých dôležitých témach BMTO a stať 5 je vyhradená pre voľné otázky a výzvy k integrácii BMTO.

Dôležité príklady biomedicínskych terminológií a ontológií (BMTO)

Opisná schéma

V oblasti biomedicíny sa vyvinulo veľa úsilia na vývoj sémantických štandardov, ako sú lekárske terminológie, ontológie a kódovacie systémy. V tejto stati sa analyzuje skupina BMTO, ktoré odrážajú širokú rozmanitosť tohto žánru. Zameriame sa na ICD, MeSH, GO, SNOMED CT, openGALEN, FMA, UMLS a OBO. Opíšu a porovnajú sa tieto systémy podľa ich identifikujúcich spoločných charakteristík a rozdielov a prediskutuje miesto, ktoré reprezentujú a akú architektúru používajú. Nakoniec sa uvedú tieto prvky architektúry, s ktorými sa budeme stretávať:

- **Uzly:** nodes, primárne identifikátory významu
- **Spojenia:** links, väzby medzi uzlami
- **Kódy:** codes, alfanumerické identifikátory pre uzly alebo spojenia
- **Hierarchie:** hierarchies, sieť spojení, ktorá tvorí čiastkové poradie, a tým definuje stromy alebo orientované (directed) grafy
- **Vlastnosti:** attributes, znázornené ako ďalšie opisy uzlov a spojení
- **Axiómy:** axioms, výroky vyjadrené v logike, ktoré sú vždy v oblasti pravdivé
Systém sa opisuje ďalej pomocou týchto vlastností:
- **Účel:** purpose, prečo sa utvoril a kde sa použil
- **Rámec:** scope, znalosť domény, ktorú reprezentuje
- **Poznámka:** reference, čo znamenajú uzly a spojenia

Medzinárodná klasifikácia chorôb (International Classification of Diseases, ICD)

Terminologická štandardizácia v medicíne má dlhú históriu. ICD bola utvorená r. 1880 (WHO, 2008) na základe London Bills of Mortality, ktorý rozoznával 200 príčin smrti a poskytoval kódy pre všetky choroby známe v tom čase. Dlho bol ICD jediným zdrojom lekárskej terminológie. Jeho súčasné (10.) vydanie spravuje Svetová zdravotnícka organizácia (WHO) a preložila sa do 42 jazykov. ICD-10 poskytuje asi 13 000 tried na kódovanie chorôb a príčin návštevy lekára. Pôvodne bola utvorená na epidemiologické účely, v súčasnosti predstavuje najviac používaný systém kódujúci choroby a používa sa globálne ako spoločný základ zdravotníckej štatistiky. V mnohých krajinách sa ICD používa aj ako základ skupín súvisiacich s diagnózou (Diagnosis Related Groups, DRG) na fakturáciu. Skupiny pacientov DRG, ktorí sú klinicky podobní a dá sa preto očakávať, že využívajú tie isté zdravotné zdroje.

ICD má jednoduchú ale efektívnu architektúru. Delí sa na 22 kapitol (infekcie, nádory, krvné choroby, endokrinné choroby atď.), jeho uzly znamenajú triedy chorôb a príbuzné problémy. To znamená, že jednotlivé choroby spadajú do kategórie, ktorá má jednotný kód, napr. myopia sa dá kódovať pomocou H52.1. Triedy ICD sú hierarchicky usporiadané do 5 úrovní. Vzťahy tvoriace hierarchiu sú vzťahy podtriedy *is-a* (JE) a vyjadrujú, že každý člen triedy je aj členom akejkoľvek rodičovskej triedy. ICD axiomaticky predpokladá, že triedy súrodencov sa neprekrývajú. To zabezpečuje, že ani jedna trieda nemá viac ako jednu triedu rodičov a že presne jedna terminálna trieda pre každú jednotku, ktorá má byť klasifikovaná, preto sa charakterizuje ako „klasifikácia“. Jednoduchým dôvodom toho je zabrániť, aby sa jedna choroba rátala dvakrát. S cieľom zabrániť medzerám sa utvorili zvyškové (reziduálne) kategórie („inde nezaradené“). Prídavným atribútom tried ICD je inklúzia a exklúzia výrokov a v jednej kapitole aj voľná textová definícia podobná glosáru. Výroky inklúzie zahŕňa zoznam špecifickejších chorôb ako sú obsiahnuté v tej istej triede, kým triedy s výrokom exklúzie segreguje (vylučuje) určité stavy z triedy, čím ich zaraďuje do inej triedy.

Rozsah ICD rozširuje oblasť chorôb a zahŕňa aj úrazy a vonkajšie príčiny zdravotných problémov, znakov a príznakov, ako aj akékoľvek stavy, ktoré zdôvodňujú stretnutie so zdravotníckymi profesionálmi. Obr. 1 znázorňuje výňatok ICD týkajúci sa určitého typu očnej poruchy, ktorá podtriedou trojmiestnej kategórie H52. Treba si všimnúť exklúziu H52.1 a inklúziu H52.5. Prvá sa musí kódovať v odlišnej vetve, kým druhá označuje špecifickejšiu poruchu, pre ktorú nie sú k dispozícii osobitné kódy. Takisto si treba všimnúť, že H52.6 tvorí komplement k H52.0–H52.5 a že H52.7 korešponduje s H52 a vyjadruje to, že kódu chýbajú detaily umožňujúce použitie špecifickejšieho kódu.

H52 Poruchy refrakcie a akomodácie
H52.0 Ďalekozrakosť
H52.1 Krátkozrakosť
 Vylučuje: degeneratívnu krátkozrakosť
H52.2 Tupozrakosť
H52.3 Anizometropia a nerovnozrakosť
H52.4 Starecké videnie
H52.5 Poruchy akomodácie
 Vnútná oftalmoplégia (kompletná)(celková)
 Paréza }
 Spazmus } akomodácie
H52.6 Iné poruchy refrakcie
H52.7 Poruchy refrakcie, nešpecifikované

Obr. 1. Výňatok z Medzinárodnej klasifikácie chorôb, verzia 10 (ICD-10)

Kľúčové slová medicíny (Medical Subject Headings, MeSH)

MeSH (Nelson, 2007; MeSH, 2008) vydaný a spravovaný Národnou lekárskou knižnicou USA (U. S. National Library of Medicine, NLM) pozostáva z riadeného slovníka používaného na indexovanie obsahu dokumentov súvisiacich so zdravím a abstrakty z databázy MEDLINE obsahujúcej všetky abstrakty z literatúry z oblasti vied o živote s takmer 20 miliónmi citácií (Nelson, 2007, PubMed). MeSH je dostupný v 41 jazykoch.

MeSH sa delí na svojej najvyššej úrovni na 16 vetiev (medzi inými na anatómiu, organizmy a choroby). Uzly MeSH sa nazývajú „heslá“ („headings“) a označujú štandardizovaný význam skupiny lekárskeho termínov. Na rozdiel od stromovej hierarchie ICD sú heslá MeSH umiestnené v mnohopočetných hierarchiách. Hierarchické usporiadanie sa zakladá na princípe, že všetky dokumenty indexované príslušným heslom sú relevantné aj pre akýkoľvek rodičovský deskriptor (lexikálnu jednotku používanú pri spracovaní a vyhľadávaní informácií, predpísané kľúčové slovo). Toto neformálne spojenie je charakterizované aj názvom „širší/užší“. Tak je heslo MeSH *leišmanióza* časťou hierarchie *Parazitárne choroby* a hierarchie *Choroby kože a spojivového tkaniva* (obr. 2). Dokumenty o leišmanióze sa tak nájdu pri zadaní tohto hesla v MEDLINE v parazitických chorobách, ako aj chorobách kože. Heslá MeSH majú tak okrem jednoznačného identifikátora aj tzv. číslo stromu pre každú hierarchickú súvislosť.

Heslá sú ďalej špecifikované textovou definíciou, tzv. „scope note“. Prídavnými atribútmi sú vstupné termíny (entry terms – synonymá alebo špecifickejšie termíny) a prípustné kvalifikátory (allowable qualifiers), ako je prevencia, terapia a iné v prípade chorôb, resp. patogenity v prípade organizmov.

Heslo MeSH (MeSH Headings)	Leišmanióza
Číslo stromu (Tree Number)	C03.752.700.500.508
Číslo stromu (Tree Number)	C.03.858.560
Číslo stromu (Tree Number)	C17.800.838,775.560
Anotácia (Annotation)	Infekcia prvokom; generický alebo nešpecifikovaný; preferuj špecifikovaný; americká leišmanióza je LEISHMANIASIS, AMERICAN, pozri LEISHMANIASIS, CUTANEOUS; tegumentárna leišmanióza = LEISHMANIASIS, CUTANEOUS
Definícia (Scope Note)	Choroba vyvolaná akýmkoľvek z rôznych druhov prvokov rodu LEISHMANIA. Jestvujú štyri hlavné klinické typy tejto infekcie: kožná (starosvetská a novosvetská)(LEISHMANIASIS, CUTANEOUS), difúzna kožná (LEISHMANIASIS, DIFFUSE CUTANEOUS), mukokutánná (LEISHMANIASIS, MUCOCUTANEOUS) a viscerálna (LEISHMANIASIS, VISCERAL)
Prípustné kvalifikátory (Allowable Qualifiers)	BL CF CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI IM ME MI MO NU PA PC PP PS PX RA RH RI RT SU TH TM UR US VE VI
Dátum zadania (Date of Entry)	19990101
Jednoznačný identifikátor (Unique ID)	D007896
Parazitická choroba [C03] Infekcia prvokmi [C03.752] Infekcia <i>Sarcomastigophora</i> [C03.752.700.500] Leišmanióza [C03.752.700.500.508]	Choroby kože a spojivového tkaniva [C17] Choroby kože [C17.800] Infekčné choroby kože [C17.800.838] Parazit. choroby kože [C17.800.838.775] Leišmanióza [17.800.838.775.560]

Obr. 2. Vstupné heslo MeSH „Leišmanióza“. Tabuľka poskytuje definíciu a atribúty. Dva „stromy“ v ktorých sa toto heslo nachádza, sú znázornené na spodku. Veľkými písmenami sú označené pôvodné heslá.

Génová ontológia (Gene Ontology, GO)

GO (GO, 2008) je spravovaná konzorciom (Gene Ontology Consortium), ktoré ju pôvodne utvorilo s cieľom podporiť výmenu anotácií genómových údajov v troch modelových organizmoch (drozofila, kvasnice, myš). Odvtedy sa jej rozsah rozšíril tak, že v súčasnosti zahŕňa všetky celú biológiu nezávisle od charakteristík špecifického organizmu. Na rozdiel od jej názvu GO nie je ontológiou génov. Namiesto toho poskytuje sémantické identifikátory, ktoré štandardizujú opis údajov o génoch a génových produktoch (napr. proteínoch) v troch dimenziách: **1.** v ktorej sa bunkový oddiel génu exprimuje (napr. mitochondria); **2.** s ktorými funkciami proteín súvisí (napr. signalizácia); **3.** na ktorých biologických procesoch proteín participuje (napr. mitóza). GO je tak schopné podporiť poskytovanie informácií naprieč databázami udržiavanými členmi konzorcia, a tým uľahčiť prístup k nimi získaným poznatkom.

Podobne ako MeSH sa GO delí na jeho najvyššej úrovni nesúvisiace vetvy. Tri vetvy *Zložky buniek*, *Biologické procesy* a *Molekulová funkcia* vymedzujú ich rozsah. Každá vetva pozostáva z mnohopočetnej hierarchie, celkovo z 24 500 uzlov, zvaných *termíny GO*. Aj keď na prvý pohľad sa architektúra GO podobá MeSH, jestvujú medzi nimi zásadné rozdiely, ktoré oprávňujú jeho kvalifikáciu ako ontológiu. Predovšetkým jeho uzly sú viac ako sémantické deskriptory. Na rozdiel od hesiel MeSH, termíny GO predstavujú triedy reálnych jednotiek (entít). Napr. (abstraktná) trieda Bunkové jadro má všetok (materiál) bunkových jadier na svete ako členov. Termíny GO sú charakterizované identifikátormi, tzv. prístupovými číslami (accession numbers) a majú ako prídavné atribúty synonymá a definície. Ďalším rozdielom proti MeSH je sémantická explicitnosť spojení. Namiesto „širší/užší“ poskytuje GO dva presne označené vzťahy: „je“ (is-a) a „časť niečoho“ („part-of“). Prvý znamená, že každá jednotka (entita), ktorá je členom jednej triedy je aj členom všetkých rodičovských „je“ (is-a) tried, práve tak ako v ICD. „Časť niečoho“ („part-of“) sa má interpretovať v tom zmysle, že každá jednotka, ktorá je členom jednej triedy je časťou niektorej jednotky a člena všetkých jej tried „časti niečoho“ („part-of“). Obr. 3 znázorňuje vstup z GO týkajúceho sa triedy *Bunka*.

GO:0005623

(PO)GO:0044464: časť bunky

(I)GO:0009334: komplex 3-fenylpropionátdioxygenázy

(I)GO:0020007: Apikálny komplex

(P) GO:0020032 : bazálny prstenec apikálneho komplexu

(P) GO:0020010 : konoid

(P) GO:0033289 : intrakonoidový mikrotubulus

(P) GO:0020009 : mikronéma

(P) GO:0070074 : mononéma

(P) GO:0020031 : polárny prstenec apikálneho komplexu

(P) GO:0020008 : roptria

(P) GO:0020025 : subpelikulárny mikrotubulus

Bunka

Informácia o termíne

Prístup (Accession) GO:0005623

Ontológia: zložka bunky

Synonymum: nijaké

Definícia: Základná štruktúrna a funkčná jednotka všetkých organizmov. Zahŕňa plazmatickú membránu a akékoľvek enkapsulované štruktúry, ako je stena a obal bunky

Zdroj: GOC:go_kuratori

Obr. 3. Vstup triedy Bunka v génovej ontológii (GO). (I) znamená hierarchiu „je“ („is-a“), (P) znamená hierarchiu „časť niečoho“ („part-of“)

SNOMED-CT

Systematizovaná nomenklatúra medicíny – klinické termíny (Systematized Nomenclature of Medicine–Clinical Terms, SNOMED-CT)(Spackman, 2004; IHTSDO, 2008) je vyčerpávajúca terminológia utvorená na pokrytie celého záznamu o pacientovi. Nadväzuje aj na štruktúry tela, procedúry a závažné aspekty týkajúce sa zdravia vrátane sociálnych súvislostí. SNOMED CT je výsledkom zlúčenia britských Readových kódov (UK Clinical Terms version 3) a SNOMED RT (referenčná terminológia)(Spackman, 2007), druhý z nich bol utvorený z niekoľkých generácií prekursorových verzií (Cornet, 2008). Od apríla 2007 vlastní, spravuje a distribuuje Medzinárodná organizácia pre vývoj štandardov zdravotníckej terminológie (International Health Terminology Standards Development Organization, IHTSDO), neziskovej spoločnosti, ktorá sídli v Dánsku. Produkty a servis SNOMED CT sú voľne prístupné pre výskumných pracovníkov, ich použitie na klinické kódovanie a iné komerčné použitie je obmedzené jeho licenciami (v súčasnosti desať krajín a niektoré spoločnosti). SNOMED CT je oficiálne dostupný v angličtine a španielčine, kým na iných prekladoch (napr. holandskom, dánskom a švédskom) sa v súčasnosti pracuje.

Z hľadiska štruktúry poskytuje SNOMED CT mnohopočetné hierarchie „je“ („is-a“), ktoré obsahujú asi 310 000 uzlov. Uzly SNOMED CT, označované ako *pojmy* (*concepts*), označujúcich zväčša triedy jednotlivých jednotiek (entít), ako sú choroby, procedúry, laboratórne výsledky, lieky ap., ale aj špeciality, ako sú geografické jednotky, hoci jestvujú dosiaľ isté spory o niektorých pojmoch, napr. o pojme *bolesť hrudníka* (Chest Paine) je predmet sám (t. j. že ide o bolesť v hrudníku daného pacienta) alebo jeho údaj v zdravotnom zázname (t. j. zápis „bolesť hrudníka“). Pojmy SNOMED CT sú jednoznačne identifikované numerickými kľúčmi spolu s ich úplne určenými názvami. Väčšina pojmov SNOMED CT zahŕňa rozličné synonymá (zvané opisy, „*descriptions*“) a len v niekoľkých málo prípadoch aj voľné textové definície. Prídavné atribúty sú kvalifikátory SNOMED, ktoré poskytujú voliteľné spresnenie pojmov, napr. *lateralita* (*Laterality*) pre anatómiu a *stupeň závažnosti* (*Severity*) pre choroby.

Súčasný pojem

Úplne určený názov: **Cholecystektómia (procedúra)**

Identifikátor pojmu: **38102005**

Definujúci vzťah:

„Je“ **Excízia žlčových ciest (procedúra)**

„Je“ **Operácia žlčníka (procedúra)**

Skupina 1:

Metóda (atribút): **Excízia – výkon (kvalifikujúca hodnota)**

Miesto procedúry – priame (atribút): **Štruktúra žlčníka (štruktúra tela)**

Tento pojem je úplne špecifikovaný

Kvalifikátory:

Prístup (atribút): **Hodnoty chirurgického prístupu (kvalifikujúca hodnota)**

Priorita (atribút): **Priority (kvalifikujúce hodnoty)**

Opisy (synonymá):

Preferované: **Cholecystektómia**

Synonymá: **Excízia žlčníka, Žlčníková excízia, Odstránenie žlčníka**

Rodičia:

Excízia žlčových ciest (procedúra)

Operácia žlčníka (procedúra)

Deti:

Cholecystektómia a revízia žlčových ciest (procedúra)
Cholecystektómia a peroperačný cholecystogram (procedúra)
Excízia lézie žlčníka (procedúra)
Laparoskopická cholecystektómia (procedúra)
Parciálna cholecystektómia (procedúra)
Totálna cholecystektómia a excízia okolitého tkaniva (procedúra)

Obr. 4. Definícia cholecystektómie v SNOMED CT. Pojem je úplne určený, t. j. kombinácia Metóda–Výkon excízie s miestom procedúry–štruktúra žlčníka je dostatočná podmienka pre žlčník

SNOMED CT pokrýva aj 50 typov spojení, zvaných *spojovacie pojmy (linkage concepts)*. Používajú sa v prípade najdôležitejších rozlišovacích kritérií SNOMED CT, menovite v jazyku bohatom na reprezentáciu ontológie kompatibilnom so sémantickým štandardom web OWL-DL (opisná logika)(Bechhofer a spol., 2004). Opisná logika umožňuje definíciu nových tried pomocou existujúcich tried a vzťahov. Ako vyplýva z obr. 4. *Cholecystektómia* je úplne definovaná ako nová trieda za použitia existujúcich tried *Excízia* a *Žlčník*, spolu so spojeniami (vzťahmi) *Metóda* a *Miesto procedúry*. To znamená, že každá a všetky procedúry excízie na niektorom žlčníku je cholecystektómia a vice versa.

Tvorba zložitých výrazov založených na pojmoch SNOMED riadiaca sa formánou syntaxou a sémantikou sa nazýva koordinácia. Tú možno vykonať v momente kódovania (prekoordinácia) alebo predtým, zavedením nového pojmu do terminológie (poskoordinácia) (Chen, 2005).

openGALEN

the Generalized Architecture for Languages, Encyclopedias and Nomenclatures – generalizovaná architektúra pre jazyky, encyklopédie a nomenklatúry poskytuje otvorený zdroj klinickej ontológie, ktorá bola vyvinutá v 90. rokoch ako výsledok série európskych projektov (GALEN)(Rector, 2003). Je určená pre klinické aplikácie a obsahuje asi 25 000 uzlov (pojmov) a 26 typov spojení (vzťahov). Pojmy openGALEN sú tiež usporiadané v mnohopočetných hierarchiách „je“ („is-a“). Využíva jazyk opisnej logiky zvaný GRAIL (GALEN Representation and Integration Language), ktorý dovoľuje definíciu tried podobne ako SNOMED CT, ale poskytuje bohatšiu syntax, ako to vidno na obr. 5, kt. znázorňuje fixáciu zlomeniny krčka ľavej stehrovej kosti. Model GALEN sa pozostáva z týchto položiek:

- Vysoká úroveň ontológie, ktorá poskytuje celkový rámec kategorizácie
- Spoločný referenčný model (common reference model, CORE) obsahujúci znova použiteľné definície z anatómie, chorôb, chirurgických procedúr, príznakov atď.
- Podrobné rozšírenia pre špeciálne subdomény, ako je chirurgia.

Jej účel je podobný ako SNOMED CT, ale nikdy nedosahuje jeho rozsah a granularitu. OpenGALEN možno však pokladať za pioniera v používaní formálnej logiky v biomedicínskych terminológiách. Jej najdôležitejšou oblasťou použitia bol vývoj francúzskej procedúry klasifikácie CCAM (Trombert-Pavot, 2000).

openGALEN: „Otvorená fixácia zlomeniny krčka ľavej stehrovej kosti“

(„Chirurgický proces“, ktorý JeCharakterizovanýNAJMA (výkon ktorý jeSpojením(„ChirurgickejFixácie“ ktorá pôsobíŠpecifickyNa (PatologickúŠtruktúruTela ktorá < zahŕňa Kost’	HLAVNÉ: fixácia PÔSOBÍ_NA zlomeninu MÁ_LOKALIZÁCIU krčiek dlhej kosti JE_ČASŤOU stehrovej kosti MÁ_LATERALITU ľavú
--	--

máCharakteristickýSúvisiaciProces ProcesZlomeniny
máŠpecifickúLokalizáciu (Krčiek ktorý
jeŠpecifickýTuhýOddiel (Femur ktorý
má ĽavoPravýVolič ľavúVol'bu))>))))

MÁ_PRÍSTUP otvorený

Obr. 5. Podrobný zápis definujúci typ fixácie zlomeniny pomocou OpenGALEN. Val'vo: reprezentácia podobná opisnej logike (syntax GRAIL). Vpravo: syntax blízka používateľovi vypracovaná kvôli uľahčeniu definície chirurgických pojmov

Výstavbový model anatómie (Foundational Model od Anatomy, FMA)

FMA (2008) je biomedicínska ontológia, ktorá poskytuje deklarované poznatky o makroskopickej štruktúre ľudského tela. Pôvodne bola vyvinutá na opis anatomických obrazov na didaktické účely. Podobne ako pri GO sú uzly FMA usporiadané do dvoch hierarchií, Taxonómia Anatómie, čo je monohierarchia typu „je“ („is-a“), a multihierarchickú sieť časť–celok (Part-Whole Network), ktorá využíva ako riadiaci vzťah „časť niečoho“ („part-of“). Prídavným atribútom je identifikátor, synonymá a prídavné vzťahy (napr. *má-rozmer*, *má-hmotnosť*, *prilieha_k* atď.). FMA sa reprezentuje rámcovým formalizmom, ktorý podmieňuje jeho menej rigidné ontologické premisy, preto sa dá preložiť do opisnej logiky len neúplne.

Uzly FMA sa nazývajú triedy alebo typy, čo zdôrazňuje skôr jej súvislosť s entitami reálneho sveta ako s významom slov. FMA však výslovne vyhlasuje, že jeho triedy pokrývajú kánonické anatomické jednotky podobne, ako anatomické atlasy, čo má za následok opis ideálneho ľudského tela bez nedostatkov alebo anatomických zmien a znetvorenín. To má za následok nezrovnalosti, ako je to s axiómou FMA, podľa ktorej „*Dolný gastrointestinálny trakt má-časť apendix*“. Je to v rozpore s častými klinickými situáciami.

Obr. 6 znázorňuje triedu *Pravá dolná nosová mušľa*, konštatujúc, že ide o časť *Lebky*, ktorá je zasa časťou *Kostry* atď. Iný zápis ju definuje ako podtyp *Dolnej nosovej mušle*, čo je Kostný orgán, a ten je podtypom mnohých iných tried vrátane najvšeobecnejšej triedy *Anatomickej jednotky*.

FMA: Pravá dolná nosová mušľa	
Svalovokostrový systém	Anatomická jednotka
Kostrový systém	Fyzikálna anatomická jednotka
Apendikulárny kostrový systém	Anatomický materiál jednotky
Osový kostrový systém	Anatomická štruktúra
Kostra (in vivo)	Orgán
Lebka	Dutinový orgán
Viscerokránium	Orgán s dutinovými orgánovými časťami
Pravá dolná nosová mušľa	Kostný orgán
	Nepravideľná kosť
	Dolná nosová mušľa
	Pravá dolná nosová mušľa

Obr. 6. Definícia pravej dolnej nosovej mušle podľa výstavbového modelu anatómie

Snahy zhromažďovať rozličné zdroje biomedicínskych poznatkov

Zdôvodnenie

Značné úsilie sa vyvinulo s cieľom na jednej strane zoradiť početné a značne sa prekrývajúce terminológie a ontológie, ale na druhej strane aj s cieľom zabrániť anarchickej proliferácii BMTO stanovením princípov pre koordinovaný vývoj interoperabilných zdrojov. V ďalšom sa opisuje Zjednotený systém lekárskeho jazyka (Unified Medical Language System, UMLS) a Otvorené

biologické ontológie (Open Biological Ontologies Foundry, OBO). Kým UMLS je príkladom prvej stratégie, OBO vyjadruje druhý prístup.

Tezaurus UMLS

Najbohatší zdroj biomedicínskych terminológií, tezaurów, klasifikačných systémov a ontológií sa konštituoval v metatezauze UMLS (Nelson, 2006; UMLS, 2008), ktorého začiatky siahajú do r. 1986, keď ho začala tvoriť Národná lekárska knižnica USA s cieľom integrovať informácie z množstva rôznorodých terminologických zdrojov. V súčasnosti UMLS pokrýva vyše 2 milióny názvov pre asi 1 milión biomedicínskych pojmov z vyše 120 BMTO, ako aj 12 miliónov vzťahov medzi týmito pojmami (Bodenreider, 2004). Okrem *openGALEN* sú všetky opísané systémy zahrnuté v metatezauze UMLS, spolu s mnohými inými, pokrývajúcimi organizmy, lieky, chemikálie, zariadenia, procedúry atď.

Okrem uľahčenia transparentného prístupu k zdrojom (poskytnutím prvotných nespracovaných súborov a online servisov) spočíva hlavný prínosom metatezauru UMLS v týchto dvoch zdrojoch:

- každý uzol zdroja BMTO je retrospektívne mapovaný na pojem metatezauru, z ktorých každý má jedinečný identifikátor, zvaný jednoznačný identifikátor pojmu (*Concept Unique Identifier, CUI*). Tieto mapovania sa pravidelne aktualizujú manuálnym spôsobom. Umožňujú premostenie medzi rozličnými zdrojmi BMTO. Následkom toho sa spojenia medzi zdrojovými uzlami mapujú na spojenia medzi CUI, zvané *sémantické vzťahy*. Aplikácie, ktoré ich využívajú znamenajú prínos úre spojenia pojmov z oboch smerov;
- každý pojem metatezauru je kategorizovaný aspoň jedným *sémantickým* typom *sémantickej siete* UMLS (UMLS SDeMantic Network), premostením pojmov nad biomedicínskou doménou (McCray, 2003). Kostru *sémantickej siete* tvorí strom 135 *sémantických* typov, spojených vzťahmi „je“ („is-a“). Sieť obsahuje navyše hierarchiu 53 asociatívnych vzťahov (napr. *location_of, treats*), ktoré sa používajú na tvorbu 612 tripletov (napr. *tkanivá, diagnostická procedúra* atď.), z ktorých sa dá odvodiť 6 252 prídavných tripletov. Tieto triplety sa interpretujú ako obmedzenia vzťahov doménou/rozsahom.

Otvorené biomedicínske ontológie (Open Biomedical Ontologies Foundry, OBO)

OBO sa utvorila r. 2003 (OBO, 2008) ako platforma ako on-line knižnica, verejná doména biomedicínskych ontológií. Na jej báze vyvinula iniciatíva OBO Foundry sadu vzájomne vymieňaných princípov regulujúcich vývoj biomedicínskych ontológií (Smith, 2007). OBO Foundry zahŕňa niekoľko anatomických ontológií (vrátane FMA), génovú ontológiu, ako aj špecializované ontológie biochémie (ChEBI), fenotypov (PATO), sekvencií (SO) a vyšetrovacích metód (OBI). V súčasnosti predstavuje zoznam kandidátov OBO Foundry vyše 50 ontológií.

OBO Foundry propaguje dva reprezentačné jazyky. Okrem OWL-DL jestvuje patentovaný formát (OBO-EDIT 2009), v ktorom je kódovaná väčšina OBO ontológií.

Tak ako v génovej ontológii, uzly OBO ontológie označujú triedy jednotiek v reálnom svete. Spojenia medzi týmito triedami sa interpretujú ako existenčne kvalifikované spojenia; napr. *A časť B* znamená, že každý typ *A* je časťou nejakého typu *B* (nie však opačne, *vice-versa*). Hlavné vzťahy OBO (*je časťou integrálnou časťou niečoho, vlastnou časťou, umiestnený v, prilieha k niečomu, je transformáciou niečoho, odvodzuje sa z, predchádza čomu, zúčastňuje sa na, pôsobí na, je druhom niečoho*) poskytuje s konzistentnými a jednoznačnými formálnymi definíciami (Smith, 2005).

Diskusia

Opísala sa vzorka BMTO, ktorá pomenúva predmet niektorou jeho časťou (*pars pro toto*) a predstavuje rozmanité *sémantické štandardy* v biológii a medicíne. Cieľom bolo poskytnúť

čitateľovi prehľad podstatných snáh zameraných na opis termínov a jednotiek (entít), ktoré označujú, s cieľom uľahčiť vyhľadávanie a spracovanie inteligentných údajov a poznatkov vo všeobecných a špecifických aplikáciách. Navyše sa prezentovali úsilia podľa ich expresívnosti v narastajúcom poradí. Jedným aspektom súvisiacim priamo s expresivitou je odstupňovanie (scaling) a rozsah (coverage), keď sa má BMTO kódované expresívnym formalizmom použiť v obmedzenejšej doméne, kým pre neformálne terminológie nie je toto obmedzenie relevantné.

Aj keď sa zdá v teórii jednoduché rozlíšiť terminológie od formálnych ontológií, v praxi je rozdiel medzi nimi menej jasný. Hlavnou ideou je, že terminológie súvisia len viac s termínmi organizujúcimi doménu (pretože ohromné množstvo termínov je v jadre každej podoblasti biomedicíny) – kým ontológie poskytujú presnejší opis, čo je dôsledkom používania formálnej logiky a koľko je len možnej nezávislosti od ľudskej reči. Typickým príkladom toho je SNOMED CT. Jeho predchodcovia mali svoje korene v kompozičnej štandardizovanej nomenklatúre (SNOMED Int.) a klinickom kódujúcom systéme (NHS Clinical Terms Version 3), ale jeho súčasná úprava sa riadi čím ďalej tým viac ontologickými princípmi. Na rozdiel od toho, BMTO, ako je ICD a MeSH možno pokladať za viac osvedčené ako dôležité a globálne užitočné prípady použitia jestvujúce už desaťročia. Najdlhšiu históriu a najväčšie rozšírenie zaznamenalo ICD v dôsledku jeho jednoduchej architektúry a včasnej potreby pre štatistiky zdravia alebo choroby. Schválené WHO a národnými inštitúciami jeho cieľ postupne zahrnul klinickú epidemiológiu, manažment zdravia, zabezpečenie kvality a fakturáciu mnohých krajín vrátane Brazílie. Na druhej strane MeSH má zložitú multihierarchickú štruktúru prispôbenú potrebám vyhľadávania v biomedicínskych textových fondoch.

Trend, ktorý sa dá jasne pozorovať, je rastúce osvojovanie jazykov sémantického webu a formalizmu, najmä v ontologickom jazyku OWL a jeho podskupine OWL-DL, z ktorých posledná je sa upravila pre potreby strojového spracovania. Hlavnou prednosťou používania deduktívnych mechanizmov, ako sú mechanizmy opisnej logiky je ich schopnosť kontrolovať zahrnutie axiém obsiahnutých v ontológiách, s cieľom podporiť vyhľadávanie poznatkov, vyjadriť sémantickú ekvivalenciu syntakticky rozdielnych výrazov a podporiť jednoznačnosť výpovedí prirodzeného jazyka. Hoci súčasne dostupné klasifikátory zápasia s problémami odstupňovania s expresívnejšími (a tým zaujímavejšími) formalizmami, fakt, že jestvujú štandardy, ako je opisná logika a OWL je výhodný pre aplikácie vyžadujúce hlbokú znalosť malého počtu podoblastí. Dokazuje to skutočnosť, že mnohé BMTO vyvinuli snahu prejsť zo svojho pôvodného formátu na opisnú logiku: SNOMED bol v minulosti čistou terminológiou; FMA sa už čiastočne premenil z rámcov na OWL a aj OBO ontológie vykazujú tendenciu používať OWL-DL, hoci v minulosti bol vyvinutý patentovaný formát a dodnes sa široko používa. Je zaujímavé, že openGALEN bol vyvinutý spočiatku na používanie ako jazyk založený na logike, podobnej jazyku DL. Môže byť preto skvelé mať najprv axiomatizované významné množstvo lekárskeho termínov, a mať v tom čase lekcie ponaučenia z biomedicínskeho ontologického inžinierstva by bolo veľmi užitočné.

Už v 80. rokoch sa identifikovalo množstvo BMTO opisujúcich sčasti sa prekrývajúce domény pre podobné alebo rozdielne prípady použitia založené na rozličných formalizmoch, filozofiách a (nevyslovených) predpokladoch. Odvtedy sa venovalo veľa úsilia do metatezauru UMLS, pomocou ktorého sa ročne krížovo mapuje a kategorizuje narastajúci počet heterogénnych zdrojov. Treba však konštatovať, že jestvujú dve obmedzenia. Po prvé, že mapovanie nemôže byť expresívnejšie ako najmenej expresívny zdroj BMTO a po druhé, že užitočnosť UMLS pre praktické aplikácie obmedzuje skutočnosť, že mnohé zo zdrojov sú predmetom jednotlivých licencií.

Na rozdiel od toho sú zdroje OBO úplne verejnou doménou a sú dostupné každému. To aspoň sčasti vysvetľuje ich úspech a vysokú úroveň biologickej odbornosti pri jej tvorbe a udržovaní.

Na obr. 7 sú zhrnuté hlavné charakteristiky opísaných BMTO a nahromadené poznatky o ich rozsahu, pokrývajúcej oblasti, objeme, formalizme a používaní.

Názov	Rozsah	Formalizmus	Počet uzlov	Aplikácie	URL
ICD	Choroby	Klasifikácia, striktné „je“	Asi 13 000 tried	Zdravotná štatistika, epidemiológia, fakturácia	www.who.int/classifications/apps/icd
MeSH	Medicína, ošetrovatelstvo, stomatológia, veterinárna medicína, systémy zdravotnej starostlivosti, predklinické disciplíny	Terminológia, sémantické siete	24 767 termínov (2008)	Indexovanie článkov z 4800 biomedicínskych časopisov z databázy MEDLINE/PubMed	www.pubmed.gov
SNOMED	Všetko kódované v elektronickom zdravotnom zázname	Opisná logika	311 000 pojmov (2008)	Informácie o anamnéze pacienta, chorobách, liečení a laboratórnych výsledkoch	www.ihtsdo.org
GO	Súčasti buniek, funkcie molekúl, biologické procesy	OBO/OWL	24 500 termínov (2008)	Výskum génov, proteínov	www.geneontology.org
GALEN	Anatómia, chirurgické výkony, choroby, zdravotná starostlivosť	Jazyk GRAIL podobný opisnej logike	Vyššie 10 000	Elektronický záznam o zdravotnej starostlivosti, rozhrania klinických používateľov, systémy podpory rozhodovania, systémy prístupu k znalostiam, spracovanie prirodzeného jazyka	www.opengalen.org
FMA	Obsah anatómie	Rámce a (sčasti) OWL	75 000 tried	Výučba, biomedicínsky výskum	http://sig.biostr.washington.edu/projects/fm/AboutFM.html
OBO	Bioinformatika a molekulová biológia	OBO/OWL/OBO_XML/RDF	60 ontológií	Používa sa ako úložisko a unifikovaná schéma na interoperujúce biomedicínske projekty	www.obofoundry.org
UMLS	Biomedicínske pojmy a pojmy súvisiace so zdravím	Sémantické siete	Vyššie 1 milión pojmov		http://www.nlm.nih.gov/research/umls/

Obr. 7. BMTO, OBO, UMLS a niektoré ich hlavné charakteristiky

Otvorené otázky a výzvy

V súčasnosti nastupuje nová éra biomedicínskej informatiky. Okrem algoritmov používaných vo výskume génov, za najaktuálnejšiu tému sa pokladajú ontológie. Jestvuje už aktívna komunita výskumných pracovníkov a používateľov profitujúcich zo sémantickej interoperability vďaka ontológiám, keďže ontológie sa čím ďalej viac používajú pri anotácii výskumných dát v molekulevej biológii a genomike. Vychádzajúce opakovane použiteľné slovníky vykazujú svoju užitočnosť pri opise biomedicínskych dát pri ich rozličných aplikáciách. Presné zachytávanie biologických poznatkov počítačovými prostriedkami umožňujú tvorbu systémov schopných riešiť robustné požiadavky nastoľované biológmi, medicínskymi výskumnými pracovníkmi a praktickými lekármi: ľahký prístup k textom a databázam obsahujúcim podrobné údaje, informácie a výroky; využívanie systémov umožňujúcich jasné a kompletne úsudky a podporu zrýchľujúcu rozhodovacie procesy pri veľkom rozsahu prípadov použitia (use cases) ap. Niektoré náročné úlohy treba však ešte vyriešiť, kým sa táto oblasť stane zrelá na použitie.

Predovšetkým ide o problém modelovania. Subtílné aspekty, ktoré treba opísať v biomedicínskych ontológiách obyčajne vyžadujú ontológie na najvyššej úrovni a metódy hodnotenia ontológií (Guarino, 2000), kým sa uplatnia; inak môže ich používanie zlyhať. Názorný príklad možno nájsť vo vzťahoch medzi hlavnými triedami *fyzikálnych objektov a množstvom látky*. Známa ontológia WordNet (Miller, 1995) používaná vo výskume informatiky, najmä v oblasti spracovania prirodzeného jazyka, konštatuje, že *fyzikálny objekt „je“ množstvo látky*. Na druhej strane Panglos, veľká ontológia používaná najmä pri preklade jazykov, opisuje tieto dve triedy opačným spôsobom, a to tak, že *množstvo látky je supertriedou fyzikálneho objektu*. Guarino a Welty (2000) vyhlasujú, že obidve interpretácie sú nesprávne: Každý druh *fyzikálneho objektu sa skladá z jedného alebo viacerých druhov množstva látky*. Zatiaľ nejestvuje vzťah supertriedy, ktorý by bol zjavný pri analýze metavlastností, ako je jednota, rigidita alebo identita. Takéto nepresnosti sa vyskytujú aj v oblasti biomedicíny. V staršej verzii génovej ontológie sa vyskytovala axióma *axón je časťou bunky*. Pri dôkladnejšom výskume mala táto definícia za následok nejednoznačnosť a nedostatočnú špecifikáciu, pretože jestvujú bunky bez axónov a axóny bez buniek, aspoň v laboratóriu (Schulz, 2004). Tieto dva príklady vyzdvihujú potrebu klásť väčší dôraz na formálnosť a sémantické bohatstvo v biomedicínskych ontológiách.

Iná kľúčová otázka, ktorá sa týka aj prvého príkladu, je integrácia. Pretože počet biomedicínskych ontológií narastá, mnohé aplikácie vyžadujú používanie viacerých ako jednej ontológie, čo má za následok rad závažných dôsledkov. Nepochybne nejde len o problém biomedicíny; hlavnou prekážkou opätovného využitia poznatkov v hlavnom prúde počítačovej vedy vyplýva z heterogenity poznatkov. Veda je už prirodzene rozmanitá svojimi charakteristikami: formou, vyjadrovaním, formalizmami reprezentácie, jazykom, syntaxou, obsahmi, významom, princípmi modelovania, praxou a štandardami, hľadiskami, perspektívami, využívaním, granularitou, terminológiou, premisami, nehovoriac o tom, že niektoré ich spojenia, môžu byť ťažko riešiteľné počítačovými prostriedkami. Hoci ontológie (v prísnom zmysle, výroky o tom, čo je vždy pravdivé a jednoznačne akceptované) pokrývajú len jasne vymedzený segment toho, čo sa bežne rozumie pod reprezentáciou poznatkov, tieto odlišnosti budú vždy ovplyvňovať rozhodovanie o konečnom dizajne a nastoľovať otázky o aplikáciách ontológie. Riešenie heterogenity sa stalo predmetom opakovaného a podnetného výskumu používania ontológie a na druhej strane aj dobrým zdrojom využitia ontológie, napr. pri integrácii informácií z heterogénnych ontológií, ako je vyhľadávanie hotelov, ktorých opis je jasne uvedený v mnohých systémov.

Granularita je osobitným problémom, ktorý má značný vplyv na integráciu biomedicínskych ontológií (Schulz, 2009). Očakáva sa spojenie lekárskeho a biologického výskumu ontológií na úrovni buniek, anatómie, liekov ap. Tieto komunity môžu vyžadovať rozličnú granularitu alebo dokonca rozličné názory na tú istú ontológiu. Ďalšou výzvou súvisiacou s integráciou je ako naložiť s existujúcimi

biomedicínskymi ontológiami, ktoré obsahujú prekrývajúce sa informácie, za predpokladu rozdielnych názorov na určité subdomény oblasti alebo pokrývajúce rozličné domény.

Umožneniu integrácie ontológie sa venuje vo výskume značná pozornosť. Ich stručný opis je zhrnutý v práci Freitas a spol. (2007) a prezentovaný podrobnejšie v práci Suckenschmidta a spol. 2000.

Pri aktuálnej aplikácii biomedicínskych ontológií je nepochybne jedným z kľúčových problémov spracovanie textu. Veľmi populárny prípad použitia je automatické zadanie termínov MeSH pri vyhľadávaní v PubMed. Iným prípadom je automatická extrakcia informácie týkajúcej sa jednotlivých génov alebo proteínov z vedeckých textov. Elektronický záznam a platforma používateľa predstavuje tiež široké pole pre spracovanie textu a poznatkov. Pri vysporadúvaní sa s touto otázkou sú systémy odkázané na systémy extrakcie informácie a dolovania textu (Muslea, 1999; Ananiadou, 2006). Mnoho otázok však ostáva nezodpovedaných a výzvu pre budúci výskum predstavuje kombinácia metodológií vysoko kvalitnej analýzy textu a vysoko expresívnych a dokonale štandardizovaných ontológií.

Linteratúra

AALDERS, H.J.G.L., (2002): The Standardisation of Object Definition: Internationally and The Dutch Case. In: There is more than geometry, VBK Editorial Management b.v., Liemeer, NL, ISBN 90-806917-1-2

AALDERS, H.J.G.L., (2000): The universe of discourse in GIS. Material for GIS Data Quality Course, GIS-section, Faculty of Geodesy, TU DELFT, The Netherlands

FREITAS, F., SCHULZ, S. (2009): Survey of current terminologies and ontologies in biology and medicine. *Elect. J. Commun. Inf. Innov. Health*. Rio de Janeiro 3, č.1, s. 7 – 18.

GRUBER, T., (1993), What is an Ontology? Knowledge Systems Laboratory, Computer Systems Dept., Stanford University, Stanford, CA94305, USA

RUCH, P., GOBEILL, J., LOVIS, CH., GEISSBUHLER, A. (2008): Automatic medical encoding with SNOMED categories. *BMC Med Inform Decis Mak*. 2008; 8(Suppl 1): S6.

SALGÉ, F., (1995): Semantic Accuracy, in *Elements of Spatial Data Quality*, Elsevier Science Ltd Oxford, UK, ISBN 0-08-042432-5, p.139 – 151.

STN EN ISO 19115:2005 Geografická informácia – Referenčný model

UITERMARK, H. T., 2001: *Ontology-Based Geographic Data Set Integration*, Deventer, NL, ISBN 90-365-1617-X

WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY, <http://www.m-w.com/home.htm> - 5.9.2005, 12:30

1. Ehrler F, Jimeno A, Geissbühler A, Ruch P. Data-poor Categorization and Passage Retrieval for Gene Ontology Annotation in Swiss-Prot. *BMC Bioinformatics*. 2005;**6**:s23. doi: 10.1186/1471-2105-6-S1-S23. [PubMed]
2. Gobeill J, Tbahriti I, Ehrler F, Mottaz A, Veuthey A, Ruch P. Gene Ontology density estimation and discourse analysis for automatic GeneRif extraction. *BMC Bioinformatics*. 2008;**9** [PubMed]
3. Aronson A, Demner-Fushman D, Humphrey S, Lin J, Liu H, Ruch P, Ruiz M, Smith L, Tanabe L, Wilbur J. Fusion of Knowledge-intensive and Statistical Approaches for Retrieving and Annotating Textual Genomics Documents. *TREC 2005*. 2006.

4. Shaw W, Wood J, Wood R, Tibbo H. The Cystic Fibrosis Database: Content and Research Opportunities. *LSIR*. 1991;**13**:347–366.
5. Marti Hearst's pages. <http://www.sims.berkeley.edu/~hearst/irbook/>
6. Larkey L, Croft W. *SIGIR*. ACM Press, New York, US; 1996. Combining classifiers in text categorization; pp. 289–297.
7. Manber U, Wu S. GLIMPSE: A Tool to Search Through Entire File Systems. *Proceedings of the USENIX Winter 1994 Technical Conference, San Francisco CA USA*. 1994. pp. 23–32.
8. Ruch P. Using contextual spelling correction to improve retrieval effectiveness in degraded text collections. *COLING 2002*. 2002.
9. Aronson A, Bodenreider O, Chang H, Humphrey S, Mork J, Nelson S, Rindfleisch T, Wilbur W. The Indexing Initiative. A Report to the Board of Scientific Counselors of the Lister Hill National Center for Biomedical Communications. *Tech rep, NLM*. 1999.
10. Ruch P, Baud R, Geissbühler A. Learning-Free Text Categorization. *LNAI 2780*. 2003. pp. 199–208.
11. Ruch P. Automatic Assignment of Biomedical Categories: Toward a Generic Approach. *Bioinformatics*. 2006;**6**
12. Hull D. Stemming Algorithms: A Case Study for Detailed Evaluation. *Journal of the American Society of Information Science*. 1996;**47**:70–84. doi: 10.1002/(SICI)1097-4571(199601)47:1<70::AID-ASIS7>3.0.CO;2-#.
13. Baud R, Nystrom M, Borin L, Evans R, Schulz S, Zweigenbaum P. Interchanging Lexical Information for a Multilingual Dictionary. *AMIA Symposium Proceedings*. 2005.
14. Ruch P. Query Translation by Text Categorization. *COLING 2004*. 2004.
15. Singhal A, Buckley C, Mitra M. Pivoted document length normalization. *ACM-SIGIR*. 1996. pp. 21–29.
16. Ruch P, Ehrler F, Abdou S, Savoy J. Report on the TREC2005 Experiment: Genomics Track. *TREC 2005*. 2006.
17. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc*. 2004;**11**:392–402. doi: 10.1197/jamia.M1552. [PubMed]
18. Lussier Y, Shagina L, Friedman C. Automating SNOMED coding using medical language understanding: a feasibility study. *J Am Med Inform Assoc (Symposium Suppl)*. 2001. pp. 418–22.
19. Funk M, Reid C. Indexing consistency in MEDLINE. *Bull Med Libr Assoc*. 1983;**71**:176–83. [PubMed]
20. de Bruijn L, Hasman A, Arends J. Automatic SNOMED classification – a corpus based method. *Yearbook of Medical Informatics*. 1999.
21. Despont-Gros C, Mueller H, Lovis C. Evaluating user interactions with clinical information systems: a model based on human-computer interaction models. *J Biomed Inform*. 2005;**38**:244–255. doi: 10.1016/j.jbi.2004.12.004. [PubMed]
22. Bowman S. Coordinating SNOMED-CT and ICD-10. *J AHIMA*. 2005:60–1. [PubMed]

23. Rassinoux A, Baud R, Ruch P, Trombert-Paviot B, Rodrigues J. Model-based Semantic Dictionaries for Medical Language Understanding. *J Am Med Inform Assoc (Symp Suppl)*. 1999. pp. 122–6.
24. Rodrigues J, Rector A, Zanstra P, R RB, Innes K, Rogers J, Rassinoux A, Schulz S, Paviot BT, ten Napel H, Clavel L, Haring E van der, Mateus C. An Ontology driven collaborative development for biomedical terminologies: from the French CCAM to the Australian ICHI coding system. *Stud Health Technol Inform*. 2006;**124**:863–8. [PubMed]

Niektoré dôležité pojmy v ontológii

axióma – základná veta systému, z ktorej sa pomocou pravidla odlúčenia dokazujú ostatné vety či →teorémy daného systému. Pojmy zavedené v axiómách sa nazývajú základné alebo primitívne pojmy. O takto zavedených pojmoch sa hovorí, že sú definované implicitne. Všetky ďalšie pojmy sa definujú explicitne pomocou axióm a základných pojmov.

deskriptívna logika →opisná logika.

DAML = DARPA Agent Markup Language – “markup” jazyk pre semantický web.

DARPA = Defense Advanced Research Projects Agency; markup language = značkovací jazyk, sada poznámok k textu, ktorý opisuje, ako má byť text štruktúrovaný, rozložený a formátovaný.

entita – hovor. položka, s ktorou možno manipulovať ako s celkom; člen určitej kategórie alebo druhu. Pôvodne a vo filozofii znamenala súcno nejakej veci (na rozdiel od jej podstaty), niečo existujúce. V odborných textoch sa používa na označenie skutočného alebo predpokladaného bližšie neurčeného útvaru, objektu, predmetu, subjektu, javu, vlastnosti, jednotky (amerikanizmus).

FMA – Foundational Model of Anatomy, výstavbový model anatómie.

formalizácia – spresnenie obsahu poznania tým, že skúmaným objektom, javom, procesom v danej oblasti reality sa priradujú materiálne konštrukcie, ktoré majú relatívne stály charakter, a preto umožňujú vyjadriť a zafixovať podstatné a zákonité vlastnosti skúmaných objektov.; prostredníctvom formalizácie sa zafixuje forma. Metóda logickej formalizácie vznikla so vznikom formálnej logiky.

formalizmus – smer v matematike, ktorý sa usiluje dospieť k riešeniu otázok týkajúcich sa teoretických základov vedy prostredníctvom formálnych axiomatických konštrukcií. Na rozdiel od →intuicionizmu hľadá východisko z krízy základov matematiky v axiomatickej metóde.

formálna logika – veda, ktorá skúma myšlienkové akty (pojmy, sudy, úsudky, dôkazy), ich logickú štruktúru alebo formu. V 2. pol. 19. stor. vznikla →matematická (symbolická) logika, ktorá rozpracovala logické teórie matematických úsudkov a dôkazov.

Zákony formálnej logiky sú: **1.** zákon totožnosti – myšlienka musí byť jednoznačná; **2.** zákon protirečenia – myšlienka nesmie obsahovať protirečenia; **3.** zákon vylúčeného tretieho – na tú istú, správne postavenú a správne pochopenú otázku, je neprípustné odpovedať neurčito – z dvoch protirečiacich si súdov je nevyhnutne jeden pravdivý a druhý lživý, tretej možnosti niet, „A“ je buď „B“ alebo nie „B“; **4.** zákon dostatočného dôvodu – každá myšlienka je pravdivá len vtedy, keď je odôvodnená, keď vyplýva z inej správnej myšlienky, ktorá jej slúži ako dôvod.

GALEN – Generalized Architecture for Languages, Encyclopedias and Nomenclatures, Generalizovaná architektúra pre jazyky, encyklopédie a numenklatúry

GO – Gene Ontology, génová ontológia

ICD – International Classification of Diseases, Medzinárodná klasifikácia chorôb

interakcia ontológií – vzájomné pôsobenie medzi „statickými“ doménovými znalosťami a postupmi usudzovania; nevyhnutný je kompromis medzi použiteľnosťou a znovapoužiteľnosťou ontológie (angl. „*usability–reusability tradeoff*“); čím je ontológia nezávislejšia od konkrétnej aplikácie, tým menej efektívne je jej priame využívanie.

intuicionizmus – smer v matematike, podľa ktorého sa matematické myslenie zakladá na racionálnej intuícii (schopnosti bezprostredne postihovať pravdu); matematické konštrukcie vyžadujú inú formu logiky než aristotelovskú – tzv. intuicionistickú logiku – bez zákona vylúčenia tretieho.

jazyk – systém znakov ľubovoľnej fyzikálnej podstaty, ktorý v procese ľudskej činnosti plní poznávaciu a komunikatívnu (dorozumievaciu) funkciu. Jazyk môže byť prirodzený a umelý.

Objektový jazyk – skúmaný jazyk, na rozdiel od neho → *metajazyk* je jazyk v ktorom sa formuluje nejaká → *metateória* (teória, ktorej predmetom skúmania je nejaká iná teória).

Prirodzený jazyk je jazyk každodenného života, forma vyjadrenia myšlienok a prostriedok spoločenského styku medzi ľuďmi.

Umelý jazyk je jazyk utvorený ľuďmi pre nejaké špeciálne účely (jazyk matematickej symboliky, jazyk fyzikálnych teórií, rozličné formy signalizácie atď.). Jedine vďaka jazyku môže existovať abstraktné myslenie. Existencia jazyka je nevyhnutnou podmienkou zovšeobecňujúcej činnosti myslenia. Napriek tomu jazyk a myslenie nie sú totožné. Z hľadiska vzniku je jazyk relatívne samostatný, pretože špecifické zákony, ktorú sú odlišné od zákonov myslenia. Preto nejestvuje totožnosť medzi pojmom a slovom, súdom a vetou atď. Jazyk je predmetom skúmania nielen jazykovedy (lingvistiky), ale aj logiky a semiotiky.

JPEG – Joint Photographic/Pictures Experts Group, normotvorná organizačná skupina na komprimovanie obrazových súborov

kalkul – systém pravidiel operovania so znakmi, ktorý rozširuje možnosti obsahového myslenia pri riešení úloh a dokazovaní súdov vyjadriteľných prostriedkami („jazykom“) daného kalkulu. Objekty, ktorými sa v kalkulke narába sú hmotné predmety (čísllice, písmená a iné znaky), ktoré sa v procese, keď sa na ne aplikujú pravidlá kalkulu, prakticky nemenia. Vyjadrenie určitých oblastí poznania v podobe kalkulov budovaných na základe metód vypracovaných modernou logikou, je najdôslednejšou formou formalizácie tejto oblasti poznania.

koncept – (l. *conceptus* –pojem) 1. formulácia, rozumový obraz, všeobecná myšlienka, pojem (→ *konceptualizmus*)

konceptualizácia – „pojmotvorba“, proces tvorby systému pojmov modelujúceho určitú časť sveta. Je to operácia, pomocou ktorej zvukový znak nadobúda význam pojmu, pričom to, čo sa označuje (vyjadruje), môže mať za následok množstvo súdov. Vo výskume je to fáza výskumného procesu, v ktorej sa definujú teoretické pojmy spôsobom, ktorý by umožnil ich následnú operacionalizáciu.

konceptualizmus – smer scholastickej filozofie, ktorého predstaviteľmi sú Abélard, J. Salisbury a i. V spore o univerzáliách konceptualisti, takisto ako nominalisti odmietali učenie realizmu, popierali reálnu existenciu všeobecného nezávisle od jednotlivých vecí. Na rozdiel od nominalistov však uznávali všeobecné pojmy ako osobitné formy poznania skutočnosti. V novoveku sa približoval konceptualizmu Locke.

Locke rozvinul teóriu poznania materialistického empirizmu, ovplyvneného aj Hobbesovým nominalizmom a Descartovým racionalizmom. Odmietol Descartovo učenie o vrodenných ideách a za jediný zdroj všetkých ideí vyhlásil skúsenosť. Idey vznikajú v dôsledku pôsobenia vonkajších vecí na zmyslové orgány (idey pociťovania) alebo v dôsledku pozornosti, zameranej na stav a činnosť duše (idey reflexívne). Prostredníctvom idey pociťovania vnímame vo veciach prvotné alebo druhotné kvality. Idey získané skúsenosťou sú len materiálom poznania, ale ešte nie samo poznanie. Aby sa

materiál ideí stal poznania, musí byť spracovaný činnosťou rozumu, odlišnou od pociťovania, ako aj od reflexie a zakladajúci sa na porovnávaní, zlučovani a abstrahovaní Prostredníctvom tejto činnosti sa jednoduché idey premieňajú na zložité. Za podmienku možností všeobecných poznatkov Locke pokladá jazyk. Poznanie definuje ako vnímanie vzájomnej zhody (alebo nezhody) dvoch ideí a za vierohodné pokladá rozumové poznanie, t. j. zistenie zhody ideí prostredníctvom rozumu. Naopak, skúsenostné poznanie je len pravdepodobné, v ňom sa zhoda ideí zisťuje odvolaním sa na fakty skúsenosti.. O zmysly sa opiera naše presvedčenie o existencii vonkajších predmetov. Tento druh poznania („senzitívny“) kladie Locke vyššie ako vierohodnosť rozumového poznania. Podľa Lockeho našou úlohou nie je poznať všetko, ale len to, čo je dôležité pre naše konanie a praktický život a také poznanie naše schopnosti plne zabezpečujú.

konštrukt – zložka, prvok obsahujúci elementárne informácie.

logická sémantika – časť logiky, ktorá skúma význam jazykových výrazov; časť metalogiky Metalogika je teória, ktorá študuje systémy a pojmy (metateória) súčasnej formálnej logiky. Rozpracúva otázky teórie dôkazu, definovateľnosti pojmov, pravdy vo formalizovaných jazykoch, interpretácie, zmyslu atď. Metalogika sa delí na logickú syntax a logickú sémantiku. Rozvoj metalogiky je spojený s utvorením a štúdiom formalizovaných jazykov.

logika – veda, ktorá skúma zákony, formy a metódy správneho myslenia, ako predpoklad pravdivého poznania. Rozdiel medzi pravdivým a správnym myslením →výroky, ktoré sú logicky správne formulované, nemusia byť pravdivé.

matematická logika – symbolická logika, skúma logické procesy prostredníctvom ich zobrazenia vo formalizovaných jazykoch alebo logických →kalkuloch.

MeSH – Medical Subject Headings, kľúčové slová medicíny.

metalogika – teória, ktorá skúma systémy a pojmy (→*metateória*) súčasnej formálnej logiky

metateória – teória, ktorej predmetom skúmania je nejaká iná teória. Študuje systém viet a pojmov danej teórie, zisťuje jej hranice, spôsoby uvádzania nových pojmov a dôkazy jej viet atď., čím umožňuje vybudovať teóriu racionálnejším spôsobom. Metateória sa formuluje v metajazyku. Najviac je rozvinutá metalogika a metamatematika. Hlavnou úlohou metateórie je skúmať podmienky formalizácie vedeckých teórií, ako aj syntaktických a sémantických vlastností formalizovaných jazykov.

nominalizmus – smer vo filozofii, ktorý pokladá všeobecné pojmy iba za mená jednotlivých predmetov. Na rozdiel od realistov nominalisti tvrdili, že reálne jestvujú len jednotlivé veci s ich individuálnymi vlastnosťami (neexistujú choroby, len chorí ľudia). Všeobecné pojmy utvárané naším myslením o týchto veciach, nielen nejestvujú nezávisle od vecí, ale dokonca neodrážajú ich vlastnosti a kvality. Nominalisti nechápali, že všeobecné pojmy odrážajú reálne kvality objektívne jestvujúcich vecí a že jednotlivé veci nie sú oddelené od všeobecného, ale obsahujú ho v sebe.

OBO – Open Biomedical Ontologies Foundry, otvorené biomedicínske ontológie.

OCML – Operational Conceptual Modeling Language, operacionálny jazyk modelujúci pojmy.

operacionalizácia – súhrn všetkých krokov alebo operácií, ktoré treba urobiť, aby sme sa od teoretickej formulácie hypotézy napojili na empirický fakt. Je to prevedenie hypotézy do podoby, s ktorou možno pracovať.

operačný jazyk modelovania pojmov – angl. *Operational Conceptual Modelling Language*, OCML.

ontologické inžinierstvo – odbor informatiky zaoberajúci sa vývojom všeobecných jazykov, metódik a softvérových nástrojov, ako aj konštrukciou samotných ontológií opisujúcich rozličné vecné oblasti a aplikácií, ktoré ich budú používať.

ontologický záväzok – angl. *ontological commitment*, rozhodnutie subjektu používať na vyjadrenie pojmov prvky a štruktúru danej ontológie, ako aj každé zásadné rozhodnutie vykonané pri samotnej konštrukcii ontológie, keď sa vyberá z rôznych variantov konceptualizácie.

ontology learning – učenie ontológiou, automatické využívanie webových zdrojov, ako sú slovníky, súhrny a adresáre (vrátane katalógov typu Open Directory).

opisná logika – deskriptívna logika, DL, skupina reprezentačných jazykov, ktoré sa dajú použiť na reprezentáciu definícií pojmov v nejakej aplikačnej oblasti (známej ako terminologická veda) štruktúrovaným a formálne dobre preštudovaným spôsobom. Opisná logika označuje na jednej strane opisy pojmov používaných na opis domény a na druhej strane na sémantiku založenú na logike, ktorá sa dá utvoriť prevedením do predikátovej logiky prvého rádu. Opisná logika vznikla ako extenzia na rámce a sémantické siete, ktoré neboli vybavené sémantikou založenou na formálnej logike. Názov opisná logika vznikla v 80. rokoch, predtým sa nazývala terminologické systémy, pojmové jazyky. V súčasnosti je opisná logika základom sémantického webu, ktorý sa používa pri utváraní dizajnu ontológií. Na opisnej logike je založený jazyk OWL-DL a OWL-Lite schválené W3C OWL.

OWL – Web Ontology Language, webový ontologický jazyk, ktorý vznikol r. 2002 na základe skúseností s využívaním DAML + OIL. Implementáciu programových nástrojov podporuje vyčlenenie „minimálnej množiny“ konštruktív jazyka, tzv., OWL Lite.

pojmem – súhrn vlastností, ktoré prislúchajú predmetu napr. pojem „strom“ – byť živou rastlinou, kmeň, koruna, koreň.

predikát – logický prísudok, jeden z dvoch termínov súdu, a to ten, ktorý hovorí o predmete (objekte), o ktorom je reč (subjekte). To, čo sa v logickom súde hovorí o predmete, subjekte súdu; vlastnosť predmetu; výraz alebo zmysel výrazu, ktorým sa niečomu (niekomu) niečo pripisuje (predikuje).

prípád použitia – scenár, use case, v softverovom inžinierstve a systémovom inžinierstve správanie sa systému, ako odpovedá na požiadavku zvonka systému. Opisuje „kto“ vie „čo“ má urobiť s daným systémom. Metóda použitia prípadov sa používa na zistenie požiadaviek správania sa systému pomocou podrobného sledovania funkčných požiadaviek systému počas jeho scenára.

RDF – angl. *Resource Description Framework* rámec opisujúci zdroj, špecifikácia W3C pre metadátový model, základ sémantického webu. Zložitejšie konštrukty ontológií (logické výrazy) sa však musia rozložiť do niekoľkých trojíc RDF, pričom každá trojica je ďalej serializovaná do elementov XML.

rozsah ontológie – angl. *hugeness*, vzhľadom na nezvládnuteľné množstvo pojmov, ktoré by sa za určitých okolností mohli chápať ako relevantné pre zvolenú doménu, treba pri tvorbe ontológie systematicky vykonávať selekciu, a zaraďovať najmä tie pojmy, ktoré majú jasnú väzbu na pojmy z „jadra“ ontológie. Ďalším prirodzeným opatrením je rozdelenie ontológie do viacerých nezávislých modulov.

semantic web mining – dolovanie sémantického webu.

SGML – Standard Generalized Markup Language, štandard všeobecného metajazyka, základ jazykov Markup na tvorbu hypertextových jazykov HTML

teoréma – ľubovoľná veta nejakej striktno vybudovanej deduktívnej (napríklad axiomatickej) teórie, ktorá je dokázaná (vyvodená) na základe východiskových postulátov tejto teórie (axióm) a/alebo na základe už dokázaných viet pomocou vyvodzovacích pravidiel prípustných v tejto teórii.

UMLS – Unified Modeling Language System, systém zjednoteného modelovacieho jazyka, grafický jazyk v softverovom inžinierstve na vizualizáciu, špecifikáciu, navrhovanie a dokumentáciu programových systémov; základný modelovací nástroj používaný pri vývoji informačných systémov. UML ponúka štandardný spôsob zápisu návrhov systému vrátane konceptuálnych prvkov ako sú biznis procesy a systémové funkcie, ako aj konkrétnych prvkov, ako sú príkazy programovacieho jazyka, databázové schémy a znova použiteľné programové komponenty. Tento editor umožňuje modelovanie UML diagramov, exportovať diagramy do formátu JPEG, WMF alebo SVG, načítať diagramy zo súboru XMI a i.

Hlavným rozdielom ontológií oproti UML je, že ich triedy v sebe nezahŕňajú procedurálne metódy ako vyjadrenie zodpovednosti za určitú funkciu skúmaného alebo navrhovaného systému. Dôraz sa naopak kladie na logické vzťahy medzi triedami, ktoré môžu mať komplikovanejšiu sémantiku ako je možné v UML graficky zachytiť pomocou asociácií. To súvisí aj so skutočnosťou, že UML je primárne určený na navrhovanie softverových systémov, kým ontológie (najmä doménové) slúžia na vyjadrenie konceptuálnych vzťahov relatívne nezávisle od programových aplikácií.

Ďalším rozdielom je postavenie inštancií. Kým v UML je každá inštancia pevne spojená s triedou, do ktorej náleží, v ontológiách inštancie (individuá) zodpovedajú špecifickým objektom reálneho sveta, sú „občanmi prvej kategórie“ a príslušnosť k triede (často aj k viacerým triedam súčasne) je len ich vlastnosťou, ktorá sa môže v čase prípadne aj meniť.

univerzálie – v stredovekej filozofii všeobecné idey. V spore o univerzálie išlo o to, či sú objektívne, reálne, ale sú to len názvy vecí. Podľa krajného realizmu (Eriugena) univerzálie existujú „pred vecami“, ideálna, podľa umierneného realizmu (Tomáš Akvinský) existujú „vo veciach“, kým podľa konceptualizmu existujú „po veciach“ – sú to len výtvary rozumu a podľa nominalizmu sú to len slová (Occam, Roscellinus).

use case – prípad použitia

výrok – syn. veta, súd; vo formálnej logike veta určitého jazyka, ktorej možno priradiť pravdivostné hodnoty (pravdivá, nepravdivá) alebo modálnosti (pravdepodobná, možná, nemožná, nevyhnutná a i.).

W3C – World Wide Web Consortium, konzorcium www

XML – eXtensible Markup Language, rozšíriteľný značkovací jazyk, ktorý bol vyvinutý a štandardizovaný konzorciom W3C (World Wide Web Consortium) ako pokračovanie jazyka SGML a HTML. Umožňuje jednoduché vytváranie konkrétnych značkovacích jazykov na rôzne účely a široké spektrum rôznych typov údajov.

Jazyk je určený predovšetkým na výmenu údajov medzi aplikáciami a na publikovanie dokumentov. Jazyk umožňuje popísať štruktúru dokumentu z hľadiska vecného obsahu jednotlivých častí a nezaobrá sa sám o sobe vzhľadom dokumentu alebo jeho časti. Prezentácia dokumentu (vzhľadu) sa potom definuje pripojeným štýlom. Ďalšou možnosťou je pomocou rôznych štýlov vykonať transformáciu do iného typu dokumentu alebo do inej štruktúry XML.

Pôvodný jazyk na publikovanie HTML už prestal vyhovovať svojou zložitosťou, ktorá vznikla jeho ustavičným rozširovaním. Jazyk XML nemá žiadne preddefinované značky (tagy, názvy jednotlivých elementov) a jeho syntax je podstatne prísnejšia ako syntax HTML. Hlavnou ideou XML je oddelenie obsahu a dizajnu dát.

XOL – angl. *eXtensible Ontology Language* rozšíriteľný ontologický jazyk.