

SÉMANTICKÁ PRESNOŠŤ – ZANEDBÁVANÝ PARAMETER KVALITY PRIESTOROVÝCH ÚDAJOV

Ing. Ivana Ivánová¹

1 ÚVOD

„Ontológia“ a s ňou úzko súvisiaci pojem „sémantika“ sa v súčasnosti teší zvýšenej pozornosti v komunite databázových a softvérových špecialistov. Ale aj odborníci v oblasti priestorových, resp. časopriestorových informačných systémov cítia dôležitosť diskusie na túto tému. Pojem „ontológia“ bol ešte zhruba pred 20 rokmi používaný výhradne filozofmi, dnes sa však vyskytuje napríklad aj v oficiálnych dokumentoch vydávanými Európskou úniou pre oblasť informatiky.

Podľa Williama z Ockhamu: „*Entita non sunt multiplicande praeter necessitatem*“, vo voľnom preklade: „nové koncepty by sa mali rozlišovať, iba v prípade ak je to nevyhnutné.“ Zároveň však platí, že prílišná generalizácia vnáša do predmetnej oblasti rozpor. Prax na druhej strane ukazuje, že nutnosť používania nových konceptov vždy „akosi“ vyplynie zo situácie a ku generalizácii sa pristupuje často z iných ako vedeckých dôvodov.

V nasledujúcom texte sa bude vyskytovať pojem „realita“, resp. „reálny svet“. Vzhľadom na to, že používanie tohto pojmu nie je vždy jednoznačné, je potrebné upresniť význam pojmu realita, resp. reálny svet relevantný pre nasledujúci text. Realitou sa tu chápe jediný časopriestor neobsahujúci žiadne paralelné svety. Každý popis reality je len jeden z mnohých možných a subjektívnych popisov. Možno nie je zbytočné pripomenúť, že jediným dokonalým popisom reálneho sveta je realita sama.

Zásadným problémom pri abstrakcii reálneho sveta sa javí nasledujúci fakt: ako je možné vôbec hovoriť o realite bez ontológie samotnej reality? Problém ako definovať pravidlá na definovanie pravidiel je meta-meta-meta-...-meta aktivita. Diskusie na túto tému je možné nájsť v textoch mnohých filozofov (napr. Eco, Tondl, Wittgenstein, Occam, Quine,..). Odhliadnuc od toho, že nastolený problém je veľmi príťažlivý sa predsa len (aj z dôvodu obmedzeného priestoru) tento príspevok obmedzí na definíciu ontologických pravidiel z technického pohľadu na kvalitu priestorových údajov.

2 ONTOLÓGIA – DEFINÍCIA

Slovo „ontológia“ pochádza z gréčtiny – „ontos“ = bytie, „logos“ = učenie, a teda znamená „učenie o bytí ako takom, nezávislom od jeho špecifických druhov“. Ontológia je náuka o existencii objektov individuálnych konceptov 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).

Oblasť Umelej inteligencie (UI) a reprezentácie znalostí narába s pojmom ontológia v kontexte explicitnej formálnej špecifikácie spôsobu reprezentácie objektov, konceptov a iných prvkov a vzťahov medzi nimi, ktorých existencia sa predpokladá v určitej oblasti

¹ Katedra geodetických základov, Stavebná fakulta, Slovenská technická univerzita v Bratislave

záujmu (Gruber, 1993). Pre systémy UI platí, že to, čo existuje, môže byť reprezentované. 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 popísané laikom chápaným textom a tiež obsahuje formálne definície, ktoré obmedzujú ich interpretáciu. Jednoducho povedané, ontológia v UI je vyjadrenie logickej teórie systému.

V ďalšom texte sa budem pridržiavať Uitermarkom použíwanej definície ontológie (Uitermark, 2001), ktorá je vyčerpávajúca a overená v oblasti priestorových informačných systémov. Definícia znie nasledovne:

„Ontológia je štruktúrovaná, obmedzená množina jednoznačne definovaných konceptov (Mars, 1995, van der Vet and Mars, 1998).“

Táto definícia sa skladá zo štyroch častí definujúcich ontológiu:

- ontológia je množina **konceptov** a nie pojmov (termínov),
- koncepty musia byť **jednoznačne** definované,
- množina konceptov je **obmedzená** a má
- **štruktúru**, čiže ontológia obsahuje aj vzťahy medzi konceptmi.

Vychádzajúc z poznámky v úvode o existencii jedinej reality pre tento článok je na mieste poznámka aj o tom, že existuje len jedna ontológia – objektívna – tá, pomocou ktorej vnímajú všetci ľudia tú jedinou realitu, v ktorej žijú. Ako vysvetľuje Frank (2005), niektorí autori odlišujú „subjektívnu“ od „objektívnej“ ontológie použitím veľkého písmena pre objektívnu a malého pre subjektívne na začiatku pojmu. Z dôvodu minimalizácie chaosu v texte sa obmedzím na nerozlíšenie ontológií z vyššie uvedeného hľadiska.

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

V mnohých prípadoch nie je možné definovať ontológiu pre všetky druhy použitia bázy priestorových 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, je nutné zladit' podľa charakteru vzťahov medzi jednotlivými úrovňami. Ontológia môže byť reprezentovaná pomocou taxonómie, stromov, katalógov, slovníkov, tezaurov, axióm alebo teorém.

Napriek vyčerpávajúcej definícii ontológie je potrebné zdôrazniť nasledovné: koncepty, ktoré nie sú obsiahnuté ontológiou **NIE JE MOŽNÉ** používať! Väčšina problémov pri budovaní báz priestorových údajov, prípadne ich integrácii vzniká práve z dôvodu neakceptovania príslušnej ontológie (napríklad aj pomocou katalógu objektov) ako jedinej možnej reprezentácie reality, ktorú báza priestorových údajov reprezentuje.

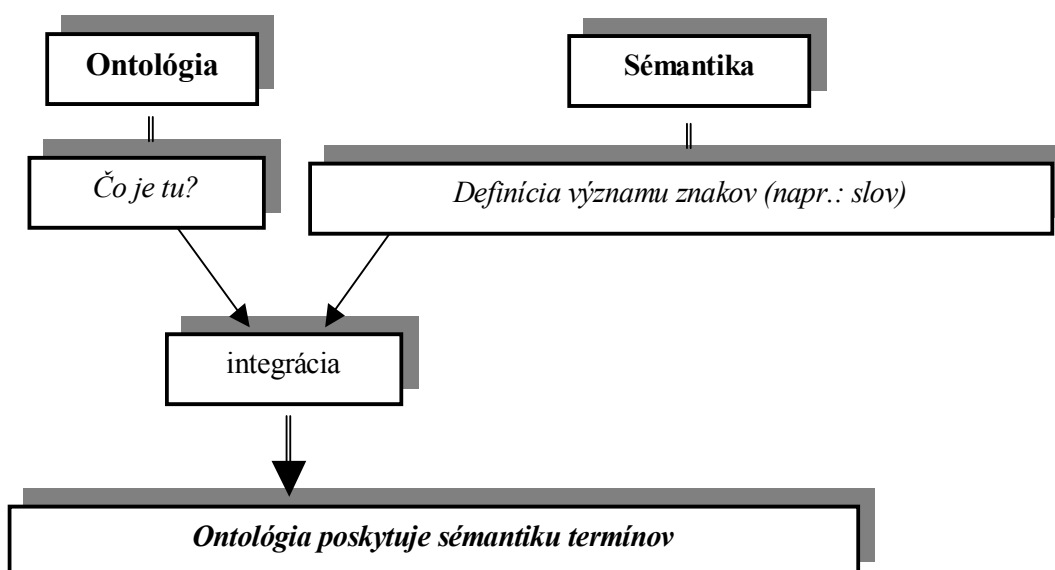
Ako upozorňuje Frank (2005), princípy ontológie hrajú pri budovaní priestorových informačných systémov kľúčovú úlohu, najmä v týchto situáciách:

- pri integrovaní báz priestorových údajov;
- pri budovaní užívateľského rozhrania priestorových informačných systémov;
- pri stanovení hodnoty priestorových údajov.

2.1 ONTOLÓGIA A SÉMANTIKA

Pri používaní priestorových báz údajov je kľúčové poznať význam týchto báz, presnejšie význam objektov nachádzajúcich sa v bázach priestorových údajov. Z pohľadu kvality priestorových údajov je sémantická presnosť, resp. správnosť (pozri kap.2) bázy údajov

jeden z rozhodujúcich parametrov pri stanovení ceny priestorových údajov. Vzájomný vzťah ontológie a sémantiky je znázornený na obrázku č.1.

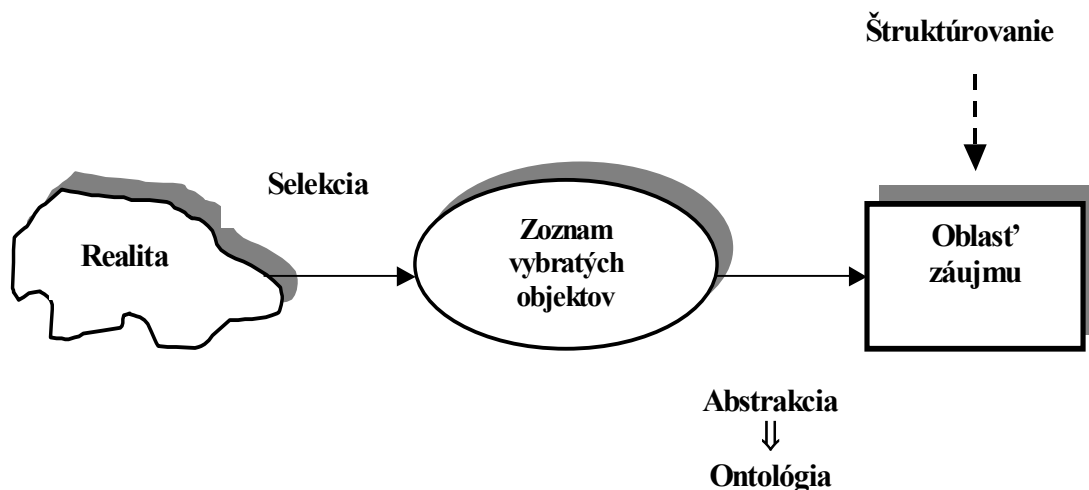


Obr.1 Vzťah ontológie a sémantiky

Z mnohých prác (napr.: Frank, 2005) vyplýva, že posun vo výskume ontológie a sémantiky nastáva ak diskusie na tému týchto dvoch predmetov nie sú oddelené – čo vlastne vyplýva aj z povahy oboch pojmov (obr. 1).

3 ABSTRAKCIA OBJEKTOV REÁLNEHO SVETA – TVORBA OBLASTI ZÁUJMU PRE PRIESTOROVÚ BÁZU ÚDAJOV

Oblasť záujmu (OZ, z ang. Universe of Discourse) je podľa normy STN EN ISO 19101:2005 – Geografická informácia – Referenčný model definovaná ako pohľad na reálny alebo hypotetický svet, ktorý obsahuje všetko, čo skúmajúceho zaujíma. OZ je abstraktným modelom databázy obsahujúci výlučne objekty záujmu. Tak ako je veľakrát počúvanie veľmi subjektívne, keď každý počuje čo chce, aj pohľad na realitu vyššie uvedeným spôsobom je ovplyvnený pomyselnými klapkami na oči, ktoré prepúšťajú 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** vybratých objektov ich popisom, definovaním atribútov a vzťahov medzi objektmi,
- **Štruktúrovanie** objektov a vzťahov medzi nimi.

3.1 ŠPECIFIKÁCIA OZ

Špecifikácia OZ je aplikačne závislá a rozumie sa pri nej tvorba konceptuálnej schémy a popis obsahu bázy priestorových údajov. Predchádzajúcu definíciu je možné priblížiť pomocou jazykovednej analógie: tvorba konceptuálnej schémy je v tom prípade identická s tvorbou gramatiky jazyka a popis obsahu bázy je porovnateľný so slovníkom pre konkrétny jazyk. Proces špecifikácie môžeme rozdeliť na dva základné kroky:

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

V procese špecifikácie OZ je možné uvažovať napr. aj nad špecifikáciou vstupu, ktorá definuje proces zberu objektov reálneho sveta do bázy priestorových údajov. Súčasťou špecifikácie OZ je aj množina parametrov kvality, ktorú producent očakáva po splnení podmienok pre vytvorenie predmetnej bázy priestorových údajov.

3.2 TVORBA PRIESTOROVEJ BÁZY ÚDAJOV NA ZÁKLADE ŠPECIFIKÁCIE OBLASTI ZAUJMU

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

- definícia objektov, ich atribútov a vzťahov medzi nimi;
- geometrická definícia referenčného systému, používaných jednotiek, rozlíšiteľnosti;
- definícia logických vzťahov v báze priestorových údajov – topológie, atribútových obmedzení a pod.
- definícia kvality údajov a sémantickej kvality bázy priestorových údajov;
- definícia času – používanej časovej sústavy;
- aktualizácia bázy priestorových údajov – frekvencia, metódy a pod.
- popis pôvodu bázy údajov;

- možnosti aplikácie – prístup, prípady použitia a pod.

Štandardy ISO pre priestorové informácie špecifikujú ontológiu na úrovni definície objektov, ich atribútov a vzťahov medzi nimi. Takisto umožňuje dodatočnú definíciu operácií objektov. Všetky definície by mali byť popísané 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 popis ontológie použitím jazyka UML.

Pri tvorbe ontológie interoperabilnej bázy priestorových údajov, je nutné dbať na splnenie nasledovných podmienok (Aalders, 2002):

- popísané sú iba tie objekty a ich atribúty, ktoré sa nachádzajú v OZ;
- fyzikálne charakteristiky objektov je možné popísať pomocou ich atribútov;
- ontológia je aplikačne nezávislá;
- ontológia je flexibilná;
- klasifikácia objektov je platná pre všetky druhy použitia bázy priestorových údajov;
- vzťahy medzi objektmi sa zachovávajú.

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

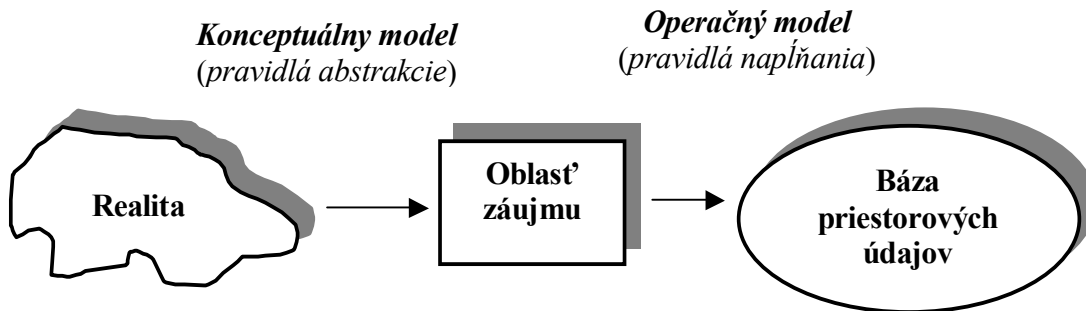
4 KVALITA PRIESTOROVEJ BÁZY ÚDAJOV

Koncepcia hodnotenia kvality vytvára priestor pre producentov aj používateľov bázy údajov pre jej ďalšie použitie. Producent hodnotí kvalitu z pohľadu dodržania vopred stanovených kvalitatívnych parametrov v špecifikácii priestorovej bázy údajov. Používateľ sa na základe popisu kvality údajov môže rozhodnúť pre použitie bázy na svoj konkrétny účel.

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

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

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



Obr. 3 Pravidlá abstrakcie reálneho sveta do bázy priestorových údajov

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

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

4.1 SEMANTICKÁ KVALITA PRIESTOROVÝCH 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 priestorových informácií rozumieme pod pojmom sémantika (-cký) 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.

4.2 SÉMANTICKÁ PRESNOŠŤ

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

4.2.1 Sémantická presnosť z pohľadu producenta bázy priestorových údajov

Zámer producenta pri dokumentácii parametrov kvality priestorovej bázy údajov z hľadiska sémantickej presnosti má podľa (Salgé, 1995) dve roviny:

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

Prvú rovinu je možné si predstaviť ako dokumentáciu špecifikácie priestorovej bázy údajov (pozri kap. 2). Sémantická interpretácia bázy priestorových 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 priestorových údajov.

4.2.2 Sémantická presnosť z pohľadu používateľa bázy priestorových údajov

Používateľ priestorovej bázy sa pri posudzovaní sémantickej presnosti riadi podobným princípom ako producent – v prvom rade je v jeho záujme pochopiť obsah ponúkanej bázy priestorových údajov. Zvyčajne sa táto snaha uskutočňuje porovnaním modelu používateľskej aplikácie a modelu producenta pri tvorbe priestorovej bázy údajov. Inými slovami, porovnávajú sa koncepty jednotlivých objektov, napr. čo rozumie producent pod pojmom „most“ 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.

4.3 HODNOTENIE SÉMANTICKEJ PRESNOSTI

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

1. **úplnosť** – porovnaním bázy priestorových údajov s realitou pre daný moment. Tento parameter je možné vyjadriť pomocou prebytku alebo vynechania objektov. Z pohľadu sémantickej presnosti sa hodnotí existencia objektov určitej triedy, ich atribútov resp. vzťahov medzi nimi.
2. **konzistentnosť** – hodnotenie platnosti sémantických obmedzení, ktoré definuje špecifikácia (napr. trieda objektov C vznikne zlúčením tried A a B) alebo vyplývajú z reality (napr.: rieky tečú smerom od prameňa po ústie). Parameter konzistentnosti je vyjadrený pomerom počtu neplatných testovaných prvkov k počtu všetkých testovaných prvkov v báze priestorových údajov.
3. **atribútová presnosť** – vyjadruje správnosť priradenia atribútov k objektom priestorovej bázy údajov.
4. **aktuálnosť** – je mierou zmeny objektov bázy priestorových údajov za daný čas. Je vyjadrená pomerom vyjadrujúcim zmenu objektov bázy voči úrovni aktualizácie bázy priestorových údajov.

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

5 ZÁVER

Diskusia na tému ontológií a s nimi súvisiacou sémantickou kvalitou, ako naznačuje príspevok je potrebná. Z príspevku vyplýva aj fakt, že po dlhom čase budovania priestorových báz údajov systémom „pokús – improvizácia – omyl“ prichádzame k diskusii o princípoch definovania priestorových báz údajov.

Príspevok naznačuje proces tvorby bázy priestorových údajov od definície objektov (kap.2), cez abstrakciu reálneho sveta do priestorovej bázy údajov – tvorby oblasti záujmu (kap.3) až po zhodnotenie celého procesu (kap.4).

Interoperabilita báz priestorových 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. Koncepty v ontológii a pravidlá modelovania slúžia na vytvorenie referenčného modelu OZ. Pravidlá modelovania popisujú vzťahy medzi ontológiou a jej možnými aplikáciami. Sémantiku bázy priestorových údajov potom definujú vzťahy medzi konceptmi ontológie a konceptmi aplikačnej ontológie.

Štandardy pre priestorové údaje (napr. séria ISO 19100) sa tvoria pre všetky kroky v procese budovania báz priestorových údajov – definíciu objektov (čiže ontológiu bázy údajov) nevynímajúc. Štandardizovaným procesom tvorby bázy priestorových údajov je možné zabezpečiť homogenitu logických vzťahov v procese jej tvorby a dokumentácie

výsledku. K používaniu štandardov je, ako ukazujú skúsenosti, najviac chýbajúcim prvkom odvaha producentov.

Príspevok vznikol ako súčasť riešenia výskumnej úlohy VEGA č. 1/1035/04 „Štandardizácia a interoperabilita geografických informácií“.

POUŽITÁ LITERATÚRA :

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

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

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