

## Konceptuální Modelování Neurochirurgických dat

Cílem předkládané diplomové práce bylo studium možností modelování neurochirurgických dat na konceptuální úrovni pomocí klasického ER modelu a jeho alternativ založených na použití grafové technologie strukturalizace medicínských znalostí. Cílem také bylo srovnání těchto přístupů a jejich zhodnocení. Dále měl být implementován systém pro zadávání dat, který respektuje obě zmíněné varianty.

V úvodu autor popisuje specifické vlastnosti lékařských informačních systémů (IS) a strukturu zdravotních záznamů pacientů. Z uvedených poznatků autor vyvozuje, že lékařský IS by měl být navržen jako *polymorfní*. Takový polymorfní IS je postavený na centrálním datovém úložišti. Každá skupina uživatelů (např. různé typy lékařů - specialistů) definuje vlastní schéma ukládaných dat. IS pak automaticky zajišťuje přístup k datům v centrálním datovém úložišti a dodává je uživatelům ve schématem definované podobě. Autorovým cílem je zavést konceptuální model pro takový IS. Požaduje hierarchickou strukturu tohoto modelu a jeho reprezentovatelnost pomocí jazyků pro popis schémat XML dokumentů. Autor tak zobecňuje problém konceptuálního modelování neurochirurgických dat na problém konceptuálního modelování hierarchických (XML) dat. V Kapitole 2 autor zdůvodňuje nevhodnost klasického E-R modelu pro konceptuální modelování XML dat a podává podrobný výčet vlastností, které by měl konceptuální model pro XML data mít. V Kapitole 3 autor zavádí známý pojem regulární stromová gramatika jako vhodný formalismus pro popis schémat XML dokumentů.

Kapitola 4 je stěžejní kapitolou diplomové práce. Autor zde zavádí vlastní konceptuální model pro XML data nazvaný A(C)-model. Model je striktně rozdělen na dvě části. Nejprve je vybudován A-model jako abstrakce regulárních stromových gramatik pro modelování struktury XML dokumentů. Autor pak popisuje převod A-modelu na regulární stromové gramatiky. Dále je na základě logiky 1. řádu vybudován C-model pro popis integritních omezení instancí popsaných A-modelem. V Kapitole 5 autor uvádí formální příklady schémat v A(C)-modelu a především popisuje vlastnosti A(C)-modelu. Dokazuje, že ke každé regulární stromové gramatice splňující drobné předpoklady vždy existuje odpovídající A-model, který generuje stejný jazyk. V Kapitole 6 se autor věnuje popisu známých jazyků pro zápis XML schémat a porovnává je. Zmiňuje, že pomocí jazyka Relax NG lze popisovat obecně regulární stromové gramatiky, zatímco pomocí jazyka XML Schema pouze jednotypové regulární stromové gramatiky. Popisuje i další výhody jazyka Relax NG nad jazykem XML Schema. Konečně v Kapitole 7 autor popisuje svojí prototypovou implementaci nástroje pro modelování v A(C)-modelu a v Kapitole 8 popisuje současný stav v oblasti konceptuálního modelování XML dat.

Práce má kvalitní úpravu a je napsána na vysoké formální úrovni. Zaváděný A(C)-model je zpracován velmi kvalitně. Autor uvádí řadu tvrzení, které také dokazuje, což v jiných pracích zavádějících konceptuální modely pro XML často chybí. V některých částech práce se však autor věnuje terminologii a formálnímu popisu používaných pojmů až příliš podrobně (např. od základů buduje predikátovou logiku 1. řádu). Na druhou stranu některé pojmy uvádí bez dalšího vysvětlení (např. pojmy teorie grafů, abelovská grupa, apod.). Autor by se neměl držet pouze formálních definic. Měl by také více uvádět neformální vysvětlení a motivaci k zaváděným pojmům. Vhodné by také bylo uvést více praktických příkladů. Ty se v práci téměř nevyskytují a přitom jejich doplnění není v případě návrhu konceptuálního modelu obtížné. Pokud se příklady v práci vyskytují, pak se netýkají lékařských dat, což by bylo žádoucí.

Zavedený A(C)-model lze chápat jako abstraktní formální aparát pro popis regulárních stromových gramatik. Lze však diskutovat o tom, zda se jedná o konceptuální model pro XML data. Některé

jeho vlastnosti jako striktní hierarchická struktura s velmi omezenými možnostmi modelování jiných typů vztahů než otec-syn a přílišná blízkost A-modelu a regulárních stromových gramatik jej posouvají spíše na logickou úroveň než konceptuální. Autor příliš brzy zavrhl možnost využití oblíbeného E-R modelu či navazujících modelů (různá rozšíření E-R či příbuzný UML). Takové modely jsou pro konceptuální úroveň vhodnější neboť striktní hierarchická struktura může být často pro návrháře příliš omezující. Navíc, požadovanou hierarchickou strukturu je možné z nehierarchického schématu odvozovat až dodatečně jako hierarchické pohledy na data. To i více odpovídá počátečnímu cíli práce, kterým byl návrh konceptuálního modelu pro polymorfni IS. Je však možné, že speciálně pro lékařské IS určené k vedení záznamů o pacientech takový přístup stačí. To ale v práci není dostatečně vysvětleno.

Z textu vyplývá autorův názor, že pro modelování na konceptuální úrovni je grafický zápis pouze okrajovou záležitostí. Formalizace a vyjadřovací síla je samozřejmě důležitá, nicméně síla např. ER modelu a důvod jeho oblíbenosti je právě v jednoduchosti a intuitivnosti jeho používání, kdy i člověk bez matematického vzdělání je schopen ER schémata navrhovat, nebo alespoň číst. Příčinou je právě grafická reprezentace, která je proto stejně důležitá jako popis po formální stránce, např. nově zaváděné konceptuální modely pro XML jsou autory často popisovány pouze na neformální, uživatelsky přívětivé, tj. grafické úrovni a formální popis není vůbec (což je na druhou stranu také chyba) nebo až dodatečně v podrobnějších publikacích. Naštěstí autor na str. 13 uvádí požadavek na grafickou reprezentaci jako požadovanou vlastnost konceptuálního modelu.

Autor správně uvádí, že problematika návrhu integritních omezení není dostatečně řešena a zajímavým způsobem zde aplikuje predikátovou logiku 1. řádu. Otázkou zůstává efektivita takového řešení (validace instancí) v případě velkých objemů dat, které se v případě lékařských IS jistě vyskytují. Nicméně je nutné tuto otázku v budoucnu řešit. Bylo by také zajímavé studovat možnosti popisu takových integritních omezení pomocí existujících jazyků určených k popisu omezení v XML dokumentech jako např. Schematron, příp. i pomocí pokročilejších dotazovacích či transformačních jazyků XQuery a XSLT.

Podle zadání bychom od autora očekávali podrobnější srovnání možností klasického ER modelu a zvolené hierarchické reprezentace. Také lze očekávat těsnější vazbu na lékařská data. Dodaná prototypová implementace umožňuje navrhování schémat v A(C)-modelu a jejich validaci. Implementace má však pouze omezené možnosti zadávání dat. Navíc má systém respektovat jak ER model tak hierarchickou reprezentaci, což implementace nerespektuje. Autor v práci vysvětluje, že otázka vstupu dat je výhradně otázkou validace. Není úplně zřejmé, co tím autor myslí a proto by bylo vhodné, aby to při obhajobě podrobněji vysvětlil.

I přes výše uvedené připomínky **doporučuji** předkládanou diplomovou práci **uznat**, především díky velmi kvalitnímu popisu A(C)-modelu, ze kterého je patrné, že autor dokáže samostatně pracovat na složitějším problému a kvalitně ho zpracovat na vysoké formální úrovni. Do budoucna však zůstává řada otevřených otázek týkajících se jak dalších úprav a rozšíření modelu, tak jeho vhodnosti pro konceptuální modelování.

V Praze 4.9.2006

Mgr. Martin Necaský  
KSI MFF UK