



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

ROČNÍKOVÁ PRÁCE

Jméno Příjmení

Závěrečná zpráva z ročníkového projektu

Název katedry nebo ústavu

Vedoucí bakalářské práce: Vedoucí práce

Studijní program: studijní program

Studijní obor: studijní obor

Praha 2024

Obsah

Úvod	3
1 Sada nástrojů	4
1.1 MM-cat	4
1.2 MM-infer	5
1.3 MM-evocat	5
1.4 MM-quecat	6
2 Cílové skupiny uživatelů	7
2.1 Definice cílových skupin	7
2.2 Popis cílových skupin	7
2.2.1 Skupina mladých lidí a dětí	8
2.2.2 Skupina začínajících ve věku 40-50 let	8
2.2.3 Skupina expertních uživatelů	8
3 HTA (Hierarchical Task Analysis)	10
3.1 Vytváření diagramů	10
4 Formát PDF/A	12
Závěr	13
Seznam použité literatury	14
Seznam obrázků	15
Seznam tabulek	16
Seznam použitých zkratk	17
A Přílohy	18
A.1 První příloha	18

Úvod

V rámci GAČR projektu Unified Management of Multi-Model Data (č. 20-22276S) byla navržena sada nástrojů pro modelování a správu multi-modelových dat, např. MM-cat a MM-evocat. Nástroje vznikaly postupně a za podpory různých lidí. Díky tomu nemají jednotné rozhraní pro práci. Navíc jsou současná rozhraní často nepřehledná a uživatelsky nepřívětivá. Proto se v rámci ročníkového projektu zabýváme návrhem sjednoceného rozhraní. Důraz je kladen na návrh vhodného uživatelského rozhraní a jeho udržitelnost.

Práce je soustředěna hlavně na nástroj MM-evocat. A to tak, aby bylo možné rozšířit aplikaci o další nástroje. Výsledkem práce je návrh low-fidelity prototypu MM-evocat a jeho otestování na vybraném vzorku reálných uživatelů.

TODO: nahradit mezery za no-break space všude v dokumentu... pomocí ~

1. Sada nástrojů

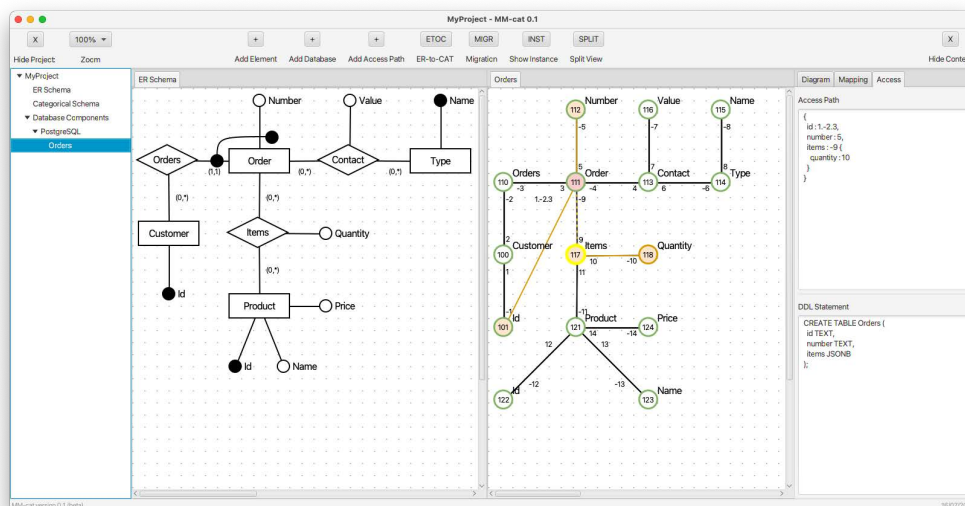
TODO: Rychlý úvod do všech nástrojů. Rozebrat hlavně mm-evocat, protože ten je základem návrhu. Názvy článků přidat do literatury na konci.

1.1 MM-cat

MM-cat framework je navržený tak, aby řešil složitosti spojené s návrhem a správou multi-modelových databází. Jeho hlavním úkolem je modelování multi-modelových schémat a jejich mapování na příslušné Systémy řízení bází dat (SŘBD). Slouží i jako základ pro rozšíření o složitější úkoly.

Typickým scénářem použití je vytvoření ER (entity-relationship) schématu uživatelem. Takové schéma je pak automaticky převedeno do jednotné kategorické reprezentace, která umožňuje namapování na kombinaci SŘBD. Díky specifikaci schématu je vytvořen skript s příkazy CREATE, které se aplikují na přiřazené SŘBD. Uživatel může následně dále upravovat ER diagram, nebo provádět SELECT dotazy.

Na ukázce uživatelského rozhraní (Obrázek 1.1) je výsledek typického scénáře rozebraného v předchozím odstavci. Na levé straně je uživatelem vytvořený ER diagram, na straně pravé pak jeho kategorická reprezentace. Na pravé straně ještě stojí za povšimnutí panel s přístupovými cestami (značené oranžově) a CREATE příkaz.



Obrázek 1.1: Ukázka uživatelského rozhraní nástroje MM-cat.

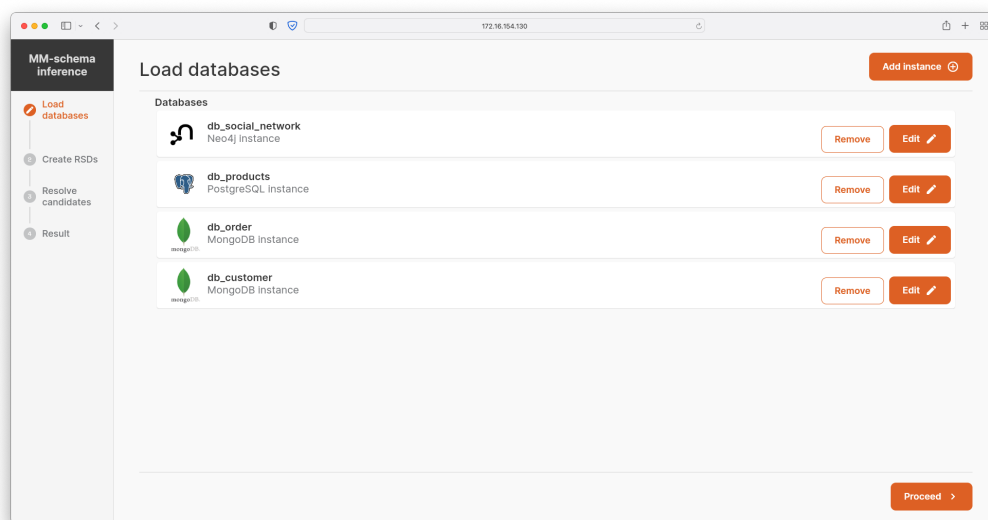
Další možnosti použití, i podrobnější popis nástroje, lze nalézt v článku Koupil a kol. (2021).

1.2 MM-infer

Ne všechna data mají předem definované schéma. Nástroj MM-infer se snaží schéma zpětně zrekonstruovat z již uložených multi-modelových dat. Je schopen odhalit intra- a inter-modelové reference a překrývání modelů. Zároveň nástroj umí efektivně zpracovávat velké množství dat.

MM-infer podporuje tři druhy SŘBD, vybrané tak, aby bylo pokryto co nejvíce funkcí takových systémů. PostgreSQL byl vybrán jako zástupce schema-full, relační SŘBD. Neo4j reprezentuje schema-less a MongoDB obojí.

Rozhraní aplikace nás provede hned několika kroky, jejichž výsledkem je globální schéma pro vybranou množinu SŘBD. Uživatel musí nejdřív takové SŘBD vybrat (Obrázek 1.2). V levém panelu je osa ukazující již splněné a následující kroky, pro lepší orientaci v procesu.



Obrázek 1.2: Načtení databází v nástroji MM-infer.

Podle vybraných databází a dalších parametrů je vytvořen RSD (Record Schema Description), sjednocující reprezentace. Po automatickém načtení je pak předána ruka uživateli, který kontroluje navržené kandidáty. Výsledkem je vizualizace globálního schématu (Obrázek 1.3).

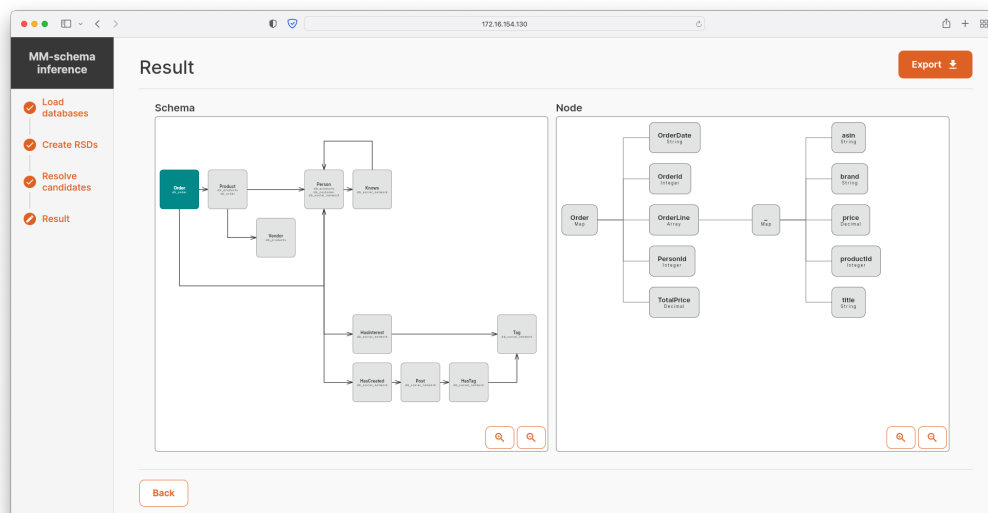
Aplikace příjemně provádí uživatele všemi potřebnými kroky. Nemá zbytečně příliš různých grafických prvků, které by uživatele zahlcovaly. Stejně jako mm-cat využívá dvou oken pro diagramy vedle sebe.

Další informace o návrhu a možnostech použití aplikace lze nalézt v Koupil a kol. (2022b).

1.3 MM-evocat

MM-evocat je nástroj pro modelování a správu evoluce v multi-modelových datech. Podchycuje tak případy, kdy se struktura dat v čase mění tak, aby například vyhovovala novým uživatelským požadavkům.

Více v Koupil a kol. (2022a).



Obrázek 1.3: Vizualizace globálního schématu v nástroji MM-infer.

1.4 MM-quecat

Více v Koupil a kol. (2023).

2. Cílové skupiny uživatelů

Pro dobré nastavení projektu si definujeme cílové skupiny uživatelů, kteří budou naši aplikaci používat. Pokud už k předchozím projektům, které se snažíme sjednotit, cílové skupiny existují, je dobré se s nimi alespoň seznámit.

Jakmile nastavíme potenciální cílové skupiny uživatelů, zvolíme si konkrétní persony. Pod personou si můžeme představit detailní popis fiktivní osoby, která reprezentuje cílovou skupinu. Persony nám pomohou lépe rozebrat ekonomický status a vlastnosti lidí ve skupinách. Nezapomínáme ani na jejich motivace a zájmy. Nesnažíme se vystihnout všechny ze skupiny, ale vybíráme si hlavně stereotypní vlastnosti a chování. Stejně jako kladné rysy a zájmy chceme vystihnout i co dotyčný nemá rád, případně vůči čemu je úplně odmítavý.

Pro práci si definujeme jenom persony, které reprezentují uživatele na které cílíme. Lze ale i definovat vylučující persony. Tedy takové, na které cílit nechceme.

Vytváření cílových uživatelů je důležitým nástrojem, který nás bude provádět v dalších krocích projektu. Díky konkrétním představám nás budou lépe napadat konkrétní řešení a realizace. Snažíme se uzpůsobit návrh personám a vyvolat v nich kladné emoce. Díky tomu získáme konkrétně zaměřený projekt, který nebude tolik odtržený od reálných uživatelů.

2.1 Definice cílových skupin

Protože k předchozím projektům, které se snažíme sjednotit, persony a skupiny neexistují, definovali jsme skupiny nové.

Nejdřív jsme potenciální uživatele rozdělili na začínající a expertní. Jedni očekávají jednoduché koncepty, protože neznají ty složitější. Seznamují se s doménou poprvé a proto bývají mnohem rychleji frustrovaní, či odrazení. Nemají zkušenost ani nadhled, kterým by se učili aplikaci používat rychleji. Tak by se dala v krátkosti vystihnout skupina začínajících. Druzí už mají nějakou zkušenost v oboru, jsou zvyklí na dotazovací jazyky. Nerozumí přímo teorii co stojí za aplikací, ale spíše se opírají o již zmíněnou zkušenost v oboru.

Bylo těžké najít jednotné zástupce skupiny začínajících. I proto jsme se rozhodli skupinu rozdělit na podskupinu mladých lidí a dětí, kteří mají více sklony k hravosti a experimentování při objevování nových konceptů a na podskupinu lidí ve věku 40 až 50 let, kteří se chtějí rekvalifikovat pro práci s daty, ale zatím nemají vhodné zkušenosti.

Nakonec jsme si ještě pohrávali s myšlenkou, že by aplikaci využíval stroj. Jednotlivé úkony by byly prováděny pomocí skriptů. Ačkoli se taková skupina nedá zařadit do skupin potenciálních uživatelů, je potřeba i s takovým využitím počítat. Nicméně ji dál nebudeme rozebírat.

2.2 Popis cílových skupin

S charakterizací skupin nám pomůže představa konkrétních person. Do skupiny mladých začínajících můžeme zařadit bakalářského studenta informatického oboru. Pod starším začínajícím uživatelem si představíme padesátiletou pracov-

nici České pošty, která se chce rekvalifikovat pro práci s daty. Člověk co vystudoval MFF UK a dále pracuje v oboru s daty, třeba učitel Datového inženýrství, bude reprezentovat zástupce expertní skupiny uživatelů.

Za pomoci person rozebereme jednotlivé skupiny potenciálních uživatelů. Každou skupinu krátce charakterizujeme, následně se snažíme vystihnout její chování. Nakonec se zamyslíme nad úkoly, které budou uživatelé v aplikaci dělat.

2.2.1 Skupina mladých lidí a dětí

Je skupinou začínajících laiků, kteří se s aplikací seznamují úplně poprvé. Nemají příliš zkušeností v oboru, neznají dotazovací jazyk. Budou pracovat hlavně s grafickým prostředím, které by jim mělo práci usnadnit a motivovat je.

Chovají se spontánně, rozhodují se impulzivně a experimentují. Z toho co dělají vyzařuje hravost. Tímto přístupem velmi rychle prozkoumají různá zákoutí aplikace a dobře otestují její funkčnost. Učí se velmi rychle, ale stejně tak je dokáže rychle odradit i maličkost. Lépe si představují věci vizuálně a ocení odezvu nebo odměnu za úkol, který provedou. Odměna pro ně může být rychlá zpětná vazba, třeba odezva na úkony v grafickém prostředí. Pro pohyb v prostředí využívají hlavně počítačovou myš. Ocení barevnost, různorodost a zajímavé podání.

V aplikaci budou využívat možnosti dotazování nad daty, jednoduché modelování schematických kategorií, bez přidávání mapování a dalších složitějších operací. Nebude se jich týkat převod konceptuálního znázornění dat do multimodelové databáze a funkce nástroje MM-infer.

2.2.2 Skupina začínajících ve věku 40-50 let

Je druhá skupina začínajících uživatelů. Je to skupina, co se typicky chce rekvalifikovat do práce s daty, ale mají zatím jen nevhodné zkušenosti. Jsou seznámeni se základní prací na počítači, umí používat webový prohlížeč, tabulkový a textový procesor. Aplikace by pro ně měla být ideální na přeučení. Nemusí se učit komplexní technologie a dál budou dělat úkony přes grafické rozhraní.

Narozdíl od mladých začínajících mají metodický přístup chování na webu, přistupují k věci konzervativně. Upřednostňují umírněné podání grafiky. Jsou pro ně důležité popsané konkrétní kroky a argumenty. Nepracují spontánně, dělají zadanou práci. Je potřeba servírovat jim množství práce postupně v dávkách, aby nepřišlo informační zahlcení a odrazení od aplikace. Stejně jako mladí začínající jsou klikací typ, k práci využívají počítačovou myš.

Úkony prováděné v aplikaci jsou podobné skupině mladých začínajících lidí.

2.2.3 Skupina expertních uživatelů

Poslední skupina, která už zná koncepty ve světě databází. Zvyklá na textové dotazovací jazyky, databázové modely. Jsou schopní psát si vlastní skripty a jinak automatizovat a zefektivnit práci.

Nechovají se spontánně a k aplikaci přistupují hlavně metodicky. Dokážou se lépe orientovat v aplikaci, díky zkušenosti z oboru. Jsou spíše tolerantní, nemají přehnané nároky na aplikaci. Vědí, že je složité přecházet na jiný systém, narozdíl od začínajících uživatelů, které odradí i maličkost. Více si potrpí na funkce, které

jím urychlí práci. Myšleno možnost používat skripty a klávesové zkratky. Ocení strážlivé podání a strohá data. Pracuje jak s grafickým rozhraním, tak i textovým.

Expertní uživatel využije více funkcí aplikace. Není omezen jen na tvorbu schematické kategorie. Může přidávat mapování a joby, převádět konceptuální znázornění dat do multimodelové databáze. Stejně tak bude využívat funkce nástroje MM-infer.

3. HTA (Hierarchical Task Analysis)

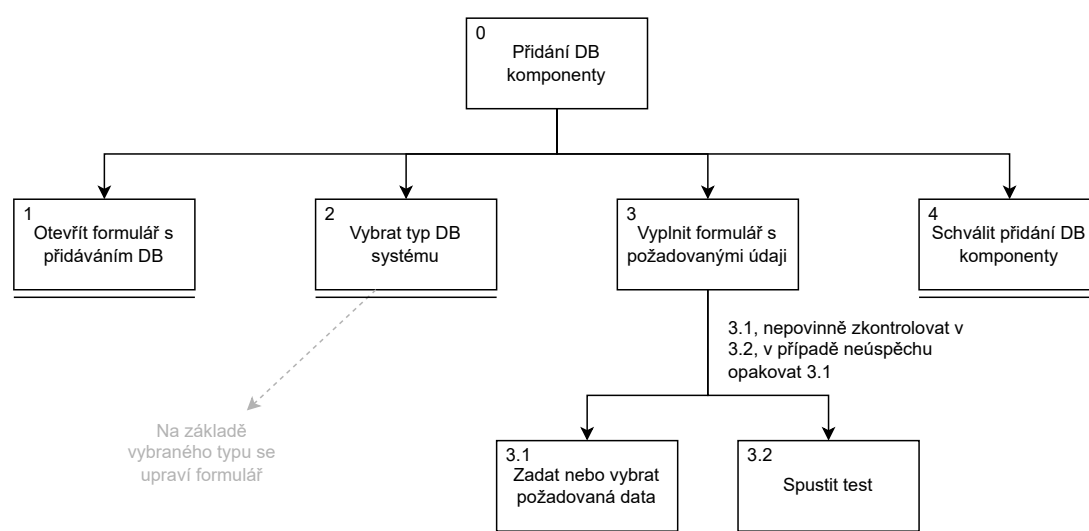
Na definování skupin uživatelů navážeme analýzou úkolů (Task Analysis) prováděných uživateli v aplikaci. Hlavním cílem analýzy je dokumentovat a strukturovat kroky nezbytné k dokončení úkolů. Díky definování různých úkolů, rozdělíme komplexní úlohy na menší, lépe představitelné části. Vše zaznamenejme do hierarchické struktury, odtud se vzalo slovo Hierarchical. Soustředíme se jen na pozitivní průchod.

Hierarchickou analýzu úkolů můžeme rozdělit na několik částí. Jako první určíme úkoly, které chtějí uživatelé v aplikaci dokončit. Každý hlavní úkol rozdělíme na podúkoly a akce, které definují vztah mezi podúkoly. Navíc každému hlavnímu úkolu přiřadíme výchozí situaci. Následně z jednotlivých komponent vytvoříme hierarchický diagram, v našem případě se stromovou strukturou. Jako poslední analyzujeme posloupnost akcí a snažíme se strukturu pochopit a vylepšit.

3.1 Vytváření diagramů

V charakterizaci cílových skupin už jsme narazili na úkoly, které by uživatelé mohli dělat. Vybereme si z nich tři, které budeme analyzovat. Začneme úkolem, kdy uživatel přidá databázovou komponentu. Další hlavní úkol je vytvoření jednoduchého schématu a poslední jeho dekompozice. Vždy začínáme v situaci systému otevřeného na hlavní stránce.

Podívejme se na první jednoduchý diagram (Obrázek 3.1). Uživatel přidá nový databázový systém k již existujícím. Začíná ve spuštěném systému.



Obrázek 3.1: Diagram pro přidání databázové komponenty.

Kořen oznamuje jaký úkol bude uživatel provádět. V první hladině máme kroky 1 až 4, které nám v základních rysech ukazují, jak bude uživatel postupovat při plnění úkolu. Krok 3 dále rozdělíme na 3.1 Zadání dat a 3.2 Otestování. Na

hraně pak definujeme postup uživatele při plnění kroků 3.1 a 3.2. Když není hrana popsána, předpokládáme průběh kroků podle čísel v hladině vzestupně. Konkrétní pořadí ale není podmíněné. Číslování 3 a 3.x vyjadřuje vztah mezi rodičem a podúkoly. Podtržené stavy v HTA se dál nedělí.

Takto vytvořený diagram nám neříká nic o interakci uživatele s konkrétním systémem. Díky tomu ale můžeme kroky rychle měnit a upravovat. Lépe zhodnotíme, jestli uživatel nedělá příliš mnoho kroků pro dokončení úkolu, nebo naopak. Často se může stát, že jeden krok je ještě potřeba rozdělit.

4. Formát PDF/A

Opatření rektora č. 13/2017 určuje, že elektronická podoba závěrečných prací musí být odevzdávána ve formátu PDF/A úrovně 1a nebo 2u. To jsou profily formátu PDF určující, jaké vlastnosti PDF je povoleno používat, aby byly dokumenty vhodné k dlouhodobé archivaci a dalšímu automatickému zpracování. Dále se budeme zabývat úrovní 2u, kterou sázíme \LaTeX .

Mezi nejdůležitější požadavky PDF/A-2u patří:

- Všechny fonty musí být zabudovány uvnitř dokumentu. Nejsou přípustné odkazy na externí fonty (ani na „systémové“, jako je Helvetica nebo Times).
- Fonty musí obsahovat tabulku ToUnicode, která definuje převod z kódování znaků použitého uvnitř fontu to Unicode. Díky tomu je možné z dokumentu spolehlivě extrahovat text.
- Dokument musí obsahovat metadata ve formátu XMP a je-li barevný, pak také formální specifikaci barevného prostoru.

Tato šablona používá balíček `pdfx`, který umí \LaTeX nastavit tak, aby požadavky PDF/A splňoval. Metadata v XMP se generují automaticky podle informací v souboru `prace.xmpdata` (na vygenerovaný soubor se můžete podívat v `pdfa.xmpi`).

Validitu PDF/A můžete zkontrolovat pomocí nástroje VeraPDF, který je k dispozici na <http://verapdf.org/>.

Pokud soubor nebude validní, mezi obvyklé příčiny patří používání méně obvyklých fontů (které se vkládají pouze v bitmapové podobě a/nebo bez unicodových tabulek) a vkládání obrázků v PDF, které samy o sobě standard PDF/A nesplňují.

Další postřehy o práci s PDF/A najdete na <http://mj.ucw.cz/vyuka/bc/pdfaq.html>.

Závěr

Seznam použité literatury

- KOUPIL, P., SVOBODA, M. a HOLUBOVÁ, I. (2021). MM-cat: A tool for modeling and transformation of multi-model data using category theory. In *2021 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*, pages 635–639.
- KOUPIL, P., BÁRTÍK, J. a HOLUBOVÁ, I. (2022a). MM-evocat: A tool for modelling and evolution management of multi-model data. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, pages 4892–4896. ACM.
- KOUPIL, P., HRICKO, S. a HOLUBOVÁ, I. (2022b). MM-infer: A tool for Inference of Multi-Model Schemas. In *Proceedings of the 25th International Conference on Extending Database Technology, EDBT 2022, Edinburgh, UK, March 29 - April 1, 2022*, pages 2:566–2:569.
- KOUPIL, P., CRHA, D. a HOLUBOVÁ, I. (2023). MM-quecat: A tool for unified querying of multi-model data. In *International Conference on Extending Database Technology*.

Seznam obrázků

1.1	Ukázka uživatelského rozhraní nástroje MM-cat.	4
1.2	Načtení databází v nástroji MM-infer.	5
1.3	Vizualizace globálního schématu v nástroji MM-infer.	6
3.1	Diagram pro přidání databázové komponenty.	10

Seznam tabulek

Seznam použitých zkratek

A. Přílohy

A.1 První příloha