

Název projektu: Mezinárodní centrum pro informaci a neurčitost

Registrační číslo: CZ.1.07/2.3.00/20.0060

Zpráva z účasti na konferenci

název konference:	DASFAA 2012: The 17th International Conference on Database Systems for Advanced Applications
datum konání:	15. 4. 2012 - 18. 4. 2012
místo:	Pusan, Korejská republika
účastník konference:	Mgr. Petr Krajča, Ph.D.

Stručný popis konference:
DASFAA patří mezi přední konference zaměřené na databázové systémy konající se pravidelně každý rok v asijsko-pacifickém regionu. Jedná se o platformu sloužící k setkávání vědců zabývajícími se databázemi, vývojářů a uživatelů z prostředí akademické sféry a průmyslu. Tématicky se konference zaměřuje na databáze, zpracování rozsáhlých dat, dolování informací z dat, vyhledávání, doporučovaní a Web.

Zajímavá čísla

Acceptance rate:	27%
Počet účastníků:	přes 320
Počet prezentovaných příspěvků:	52 (44 článků, 8 krátkých prezentací), 8 demonstrací, 8 na průmysl orientovaných prezentací
Počet plenárních přednášek:	2
Počet paralelních sekcí	4 (z toho jedna sekce tutoriálů)

Zajímavé přednášky

A. Plenární přednášky

Divesh Srivastava (AT&T Research, USA): Enabling Real Time Data Analysis
Na přednášce byla představena databázová infrastruktura používaná telekomunikační společností AT&T pro sběr dat a jejich následnou analýzu. Tato infrastruktura se skládá z několika modulů. (1) Daytona: DBMS schopný v reálném čase pracovat až s triliony záznamů a několika sty terabyty dat při zachování dostupnosti 24x7. (2) DataDepot: systém určený pro konverzi příchozích dat do datového skladu založeném na systému Daytona. (3) GS Tools: výkonný systém pro monitoring síťového provozů. (4) Bistro: systém pro úpravu různých zdrojů dat do jednotné podoby.

Sang Kyun Cha (Seoul National University & SAP Labs Korea, South Korea): A New Paradigm of Thinking and Architecture for Real-Time Information Processing at Fingertips
Hlavním tématem přednášky byly koncepční změny v architektuře informačního systému SAP. Vývoj hardware v posledních letech umožnil odstranit některé historicky zažité bariéry tykající se zejména nutnosti rozdělit informační systém do několika oddělených vrstev. Nově navržená architektura systému SAP HANA umožňuje sjednotit prostředí pro OLTP a OLAP a poskytovat uživatelům informace v reálném čase. Mimoto posun k in-memory databázím a možnost použití běžného hardwaru umožňuje optimalizovat výkon systému a snížit náklady na pořízení a další provoz.

B. Řádné přednášky

Wenjie Zhang, Jing Xu, Xin Liang, Ying Zhang, Xuemin Lin: Top-k Similarity Join over Multi-valued Objects
Přednáška se zabývala prací s objekty skládajících se z více hodnot stejného typu (např. sportovní tým), zejména pak problematikou vyhledávání podobných objektů. Přednáška zahrnovala zavedení metriky, algoritmus kNN, který rozšiřoval základní vyhledávání na R-stromech o prořezávání.

Patrick Roocks, Markus Endres, Stefan Mandl, Werner Kiebling: Composition and Efficient Evaluation of Context-Aware Preference Queries
Na přednášce byl prezentován systém dotazování založený na modelování preferencí. Vedle striktních podmínek je možné zadat preference (slabé podmínky). Navržený model umožňuje generovat další podmínky na základě kontextu – historie, aktuální situace, atd. Navržený systém byl prezentován na aplikaci, která doporučovala outdoorové aktivity na základě aktuálního počasí, informací o uživateli, jeho předchozí aktivitě, apod.

Linlin Ding, Junchang Xin, Guoren Wang, Shan Huang: ComMapReduce: An Improvement of MapReduce with Lightweight Communication Mechanisms
Příspěvek se zabýval rozšířením MapReduce frameworku o koordinační uzel, který umožňuje na základě výstupu jednotlivých mapperů redukovat vstup do reduceru a tím redukovat množství přenášených a zpracovávaných dat.

Christoph Lofi, Ulrich Guntzer, and Wolf-Tilo Balke: Malleability-Aware Skyline Computation on Linked Open Data
Autoři tohoto příspěvku navrhly mechanizmus, který umožňuje zeslabit podmínky při dotazování nad Linking Open Dat (typicky RDF, DBPedia) tak, aby systém vracel i data, která sice striktně neodpovídají dotazu, ale stále jsou relevantní z pohledu uživatele.

Yanwei Xu, Jihong Guan, and Yoshiharu Ishikawa: Scalable Top-k Keyword Search in Relational Databases
Na této přednášce byl prezentován algoritmus, který umožňuje prohledávat relační databázi na základě zadaných klíčových slov. Využívá se propojení tabulek přes cizí klíče. Tento přístup však generuje velké množství potenciálních odpovědí a je proto nutné provést redukci na *k* nejlepších výsledků.

Trieu Minh Nhut Le, and Jinli Cao: Top-k Best Probability Queries on Probabilistic Data
Na této přednášce byl představen databázový model, který vycházel z předpokladu, že řádky spadají do tabulky s určitou pravděpodobností.

Van Bao Tran Le, Sebastian Link, and Mozhgan Memari: Discovery of keys from SQL tables
Příspěvek se zabýval algoritmy určených k nalezení klíčů v relačních tabulkách za použití Armstrongových tabulek.

Md. Rezaul Karim, Md. Mamunur Rashid, Byeong-Soo Jeong, and Ho-Jin Choi: Privacy Preserving Mining Maximal Frequent Patterns in Transactional Databases

Tento příspěvek se zabýval efektivní anonymizací dat určených k hledání nejčastěji vyskytujících se vzorů v transakčních databázích. Navržený algoritmus využívá kódování transakcí za pomoci prvočísel.

Ugur Demiryurek and Cyrus Shahabi: Indexing Network Voronoi Diagrams
Na přednášce byl představen algoritmus určený pro vyhledávání k-nejbližších objektů v eukleidovských prostorech za použití Voroniových diagramů.

Qianlu Lin, Ying Zhang, Wenjie Zhang, and Aiping Li: General Spatial Skyline Operator
Příspěvek se zabýval vyhledáváním k-nejbližších objektů na základě dané preferenční funkce.

Md. Mamunur Rashid, Md. Rezaul Karim, Byeong-Soo Jeong, and Ho-Jin Choi: Efficient Mining Regularly Frequent Patterns in Transactional Databases
Prezentován byl model a související algoritmy sloužící k identifikaci často se opakujících se vzorů v databázi transakcí.

Carson Kai-Sang Leung and Syed K. Tanbeer: Fast Tree-Based Mining of Frequent Itemsets from Uncertain Data
V tomto příspěvku byl představen nový efektivní algoritmus pro hledání často se vyskytujících množin položek v pravděpodobnostních databázích.

Ghim-Eng Yap, Xiao-Li Li, and Philip S. Yu: Effective Next-Items Recommendation via Personalized Sequential Pattern Mining
Prezentovaný příspěvek se zabýval problematikou, jak vylepšit personalizované doporučovací systémy s využitím systému pro hledání často se vyskytujících sekvencí.

Yingjie Shi, Xiaofeng Meng, and Bingbing Liu: Halt or Continue: Estimating Progress of Queries in the Cloud
Přednáška byla zaměřena na model, který umožňuje předem určit dobu trvání dotazu zpracovávaného za pomocí MapReduce frameworku. Předložený algoritmus se umí vypořádat i s heterogenním prostředím.

Lei Chen and Xiang Lian: Query Processing over Uncertain and Probabilistic Databases
Tutoriál byl zaměřený na praktické aspekty práce s daty, ve kterých je potřeba pracovat s nejistotou nebo pravděpodobnostní. Představeny byly jednotlivé modely a způsoby dotazování nad nimi.

Další prezentace (panely, demonstrace, postery)
Workshopů jsem se neúčastnil, protože moje registrace je nezahrnovala (měly další vlastní poplatky). Postery na této konferenci nebyly prezentovány.

Vlastní prezentace
Neměl jsem vlastní prezentaci.

Shrnutí konference (perspektivní témata apod.)
Na konferenci byly prezentovány zásadní příspěvky tykající se všech aspektů databázových systémů – jmenovitě problematikou uložení dat, jejich indexace, efektivního vyhledávání, dolováním dat, prací se strukturovanými a prostorovými daty. Stranou nezůstala ani praktická témata zabývající se pokročilými aplikacemi zaměřujícími se zejména na sémantickou analýzu dat a sociální sítě. Z příspěvků prezentovaných na konferenci lze usuzovat, že mezi perspektivní směry dalšího výzkumu patří zejména práce se strukturovanými a vícerozměrnými daty, práce s rozsáhlými daty a analýza dat v sociálních sítích. Na konferenci byly prezentovány zásadní příspěvky, které se těmito problémy důkladně zabývaly a budou mít významný přínos pro další výzkum na Katedře informatiky.

Navázání kontaktů
Jan Vosecky *(University of Honk Kong)* – diskuze o různých aspektech využití dataminingu při analýze dat v sociálních sítích.

Fotografická dokumentace


Přílohy
žádné