



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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: ACM SIGMOD/PODS 2011 (ACM SIGMOD International Conference on Management of Data & ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems)
datum konání: 12.6.-16. 6. 2011
místo: Athény, Řecko
účastník konference: Mgr. Jan Konečný, Ph.D.

Stručný popis konference:

ACM SIGMOD/PODS je hlavní mezinárodní fórum pro databázové vědce, provozovatele, vývojáře i uživatele. Konference je sponzorována největšími firmami v oblasti informatiky (Microsoft, Google, Yahoo, IBM apod.).

Zajímavá čísla

Acceptance rate: SIGMOD (výzkumné články 23%, demo 39%, krátké prezentace 38%)
PODS (22%)

Počet účastníků: 655

Počet prezentovaných příspěvků: SIGMOD (87 výzkumných článků 32 demonstrací, 14 krátkých prezentací), PODS (25 přednášek)

Počet plenárních přednášek: SIGMOD (2), PODS (1)

Počet paralelních sekcí 7

Zajímavé přednášky

A. Plenární přednášky

Videostreamy jsou k dispozici zde:

<https://services.choruscall.eu/links/sigmod1106.html#>

James Hamilton – *Internet-Scale Storage*.

Přednášející analyzuje datové centrum s ekonomických hledisek (cena energie, serverů, úložišť, sítí, chlazení,...) vyvozuje závěry, do čeho je potřeba investovat – nejen při nákupu, při rozšiřování/škálování center, ale také ve výzkumu. Upozorňuje na některá technologická omezení, na které se naráží v cloudových databázích.

Anastasia Ailamaki – *Managing Scientific Data (The 18-month rule)*

Přednášející se zabývá přesnou analýzou experimentálních dat pro vědecké účely (koloběh vody, výzkum vesmírných těles, simulace zemětřesení, ...). Protože složitost simulovaných procesů neustále roste, roste i velikost datasetů pro tyto procesy, přitom jsou udržovány na starých systémech. Měly by se tedy vyvinout systémy a techniky pro správu vědeckých dat. Přednášející probírá práce týmů, které se o to snaží.

Tova Milo – *A Quest for Beauty and Wealth.*

Přednášející nejprve vyjmenovává dobré vlastnosti, které čekáme u systémů správy a business procesů dat (vtipně je ze ženského pohledu srovnává s vlastnostmi mužů). Ukazuje, které z nich mají současné technologie, upozorňuje, že jedny se zaměřují na data a druhé na chování procesu, ale málo z nich se dostatečně zaměřuje na obojí. Poukazuje na možné směry výzkumu v této oblasti.

B. Řádné přednášky

M. Zhang et al. – *Automatic Discovery of Attributes in Relational Databases* – přednáška pojednávala o automatickém sdružování sloupců tabulek v relační databázi do klastrů – atributů. Autoři ve svém algoritmu využívají rozložení podobnosti, čím podobnější rozložení, tím pravděpodobnější, že sloupce náleží stejnému atributu.

G. Konstantinidis, José Luis Ambite – *Scalable Query Rewriting: A Graph-Based Approach* – přednáška se týkala problému hledání odpovědi na dotaz nad databázovým schématem používá pouze v pohledů nad stejným databázovým schématem (v kontextu integrace dat). Autor uvádí algoritmus (založený na výsledcích z teorie grafů), který je výrazně rychlejší, než ten dosavadní.

H. Elmeleegy et. al – *Leveraging Query Logs for Schema Mapping Generation in U-MAP* – Přednáška pojednávala o mapování schémat – problem přenesení dat z jednoho datového zdroje podle daného schématu do zraje s jiným schématem (spadá do kategorie správy heterogeních databází). Autoři navrhuje systém U-Map na generování mapovacích schémat, který využívá informace z query logů daných databází.

B. Alexe et. al – *Designing and Refining Schema Mappings via Data Examples* – Pojednává také o mapování schémat. Autoři prezentují nové paradigma, a systém pro interaktivní návrh mapování pomocí příkladů z dat.

Jun Gao et. al – *Neighborhood-Privacy Protected Shortest Distance Computing in Cloud* – Přednáška se týká studie problému výpočtu nejmenší vzdálenosti v cloudu tak, aby nebyla vydána citlivá informace o struktuře grafu a přitom byla minimalizována práce na straně klienta.

Guoliang Li et. al – *Faerie: Efficient Filtering Algorithms for Approximative Dictionary-based Entity Extraction* – Autoři navrhuje jednotný framework, který pokrývá mnoho současných metod extrakce entit v dokumentu (např. identifikace předdefinovaných entit – osobních jmen, lokací, ...). Dále navrhuje efektivní metodu filtrování, která zabraňuje redundantním výpočtům.

Hwanjo Yu et. al – *Exact Indexing for Support Vector Machines* – Autoři věnují pozornost funkci SVM (Support Vector Machine) pro nalezení horních k výsledků; Obvyklé algoritmy nejsou na tento problém aplikovatelné (mají nároky na vlastnosti ohodnocovací funkce, nebo jsou založeny na metrikách). Proto je navržen nový způsob ohodnocování.

Venu Satuluri et. al – *Local Graph Sparsification for Scalable Clustering* – Přednáška se týkala sparsifikace grafu – tj. vytvoření řídicího grafu odebráním hran tak, aby bylo umožněno rychlejší shlukování na grafu, aniž by došlo ke ztrátě na kvalitě shlukování. Autoři toto řeší ohodnocováním hran na základě jednoduché heuristiky (založené na podobnosti) a ponecháním těch s nejvyšší hodnotou.

Francesco Gullo et. al – *Advancing Data Clustering via Projective Clustering Ensembles* – Přednáška pojednávala o prvotních formulacích PCE (Projective Clustering Ensemble) problému ve shlukování, zejména o jejích nevýhodách; autoři pak navrhuji alternativní formulaci PCE problému, která netrpí některými z těchto nevýhod.

Zengfeng Huang et. al – *Sampling Based Algorithms for Quantile Computation in Sensor Networks* – Autoři prezentují algoritmus, založený na vzorkování, na výpočet aproximačních kvantilů s lepší složitostí, která roste sublineárně s velikostí sítě (až na patologické případy). Výsledky ukazují na experimentech s umělými i skutečnými daty.

Serge Abiteboul et. al – *Rule-based Language for Web Data Management (PODS)* – Přednáška pojednávala novém jazyku založeném na pravidlech (rule-based) ke specifikaci moderních distribuovaných aplikací, především na Webu.

Ronald Fagin – *Rewrite Rules for Search Database Systems (PODS)* – Autoři vyvíjí framework, který formalizuje pojem prepisovacího programu. Ten je v podstatě kolekce prah-přepisujících pravidel (hedge-rewriting).

Liang Jeff Chen, Yannis Papakonstantinou – *Context-sensitive Ranking for Document Retrieval* – Autoři studují problém kontextové závislého rankování pro document retrieval, kde kontext je definován jako podkolekce dokumentů a je specifikovaná uživatelem.

Nathan Bales et. al – *Score-Consistent Algebraic Optimization of Full-Text Search Queries with GRAFT* – Autoři se věnují dvěma otevřeným problémům zahrnujícím algebraické vykonávání full-textového vyhledávání.

Mingyang Zhang et. al – *Mining a Search Engine's Corpus: Efficient Yet Unbiased Sampling and Aggregate Estimation* – Autoři navrhuji novou techniku k vzorkování dokumentu a k odhadu agregací (count, sum, avg).

Arijit Khan et. al – *Neighborhood Based Fast Graph Search in Large Networks* – Autoři navrhuji aproximační algoritmus na hledání podgrafu ve velkých grafech založený na sousednosti.

Sebastiaan J. van Schaik, Oege de Moor – *Memory Efficient Reachability Data Structure Through Bit Vector Compression* – Autoři navrhnou jednoduchou alternativu struktur a algoritmů pro reprezentaci tranzitivního uzávěru grafu založené na nové formě bit-vektorové komprese.

Wenfei Fan et. al – *Incremental Graph Pattern Matching* – Autoři zkoumají inkrementální algoritmy na pattern matching v grafech definovaný v rámci grafové simulace, ohraničené simulace a isomorfismu podgrafů.

Ziyu Guan et. al – *Assessing and Ranking Structural Correlations in Graphs* – Autoři předkládají nové měření pro ohodnocování strukturálních korelací v heterogenních grafových datasetů s událostmi.

Další prezentace (panely, demonstrace, postery)

V rámci SIGMOD/PODS proběhlo 8 sekcí demonstrací a 5 tutoriálů, těch jsem se neúčastnil, protože probíhaly paralelně s řádnými přednáškami. Vše, co bylo prezentováno formou Posteru (vyjma studentských prezentací), bylo též prezentováno orálně jako řádná přednáška.

Vlastní prezentace

Na SIGMOD/PODS 2011 jsem neměl vlastní prezentaci.

Shrnutí konference (perspektivní témata apod.)

Velmi úspěšná konference, byly prezentovány průlomové myšlenky, nové algoritmy a metody v databázových technologiích. Oproti tradičním tématům se stále perspektivnějším stává správa dat v cloudových databázích.

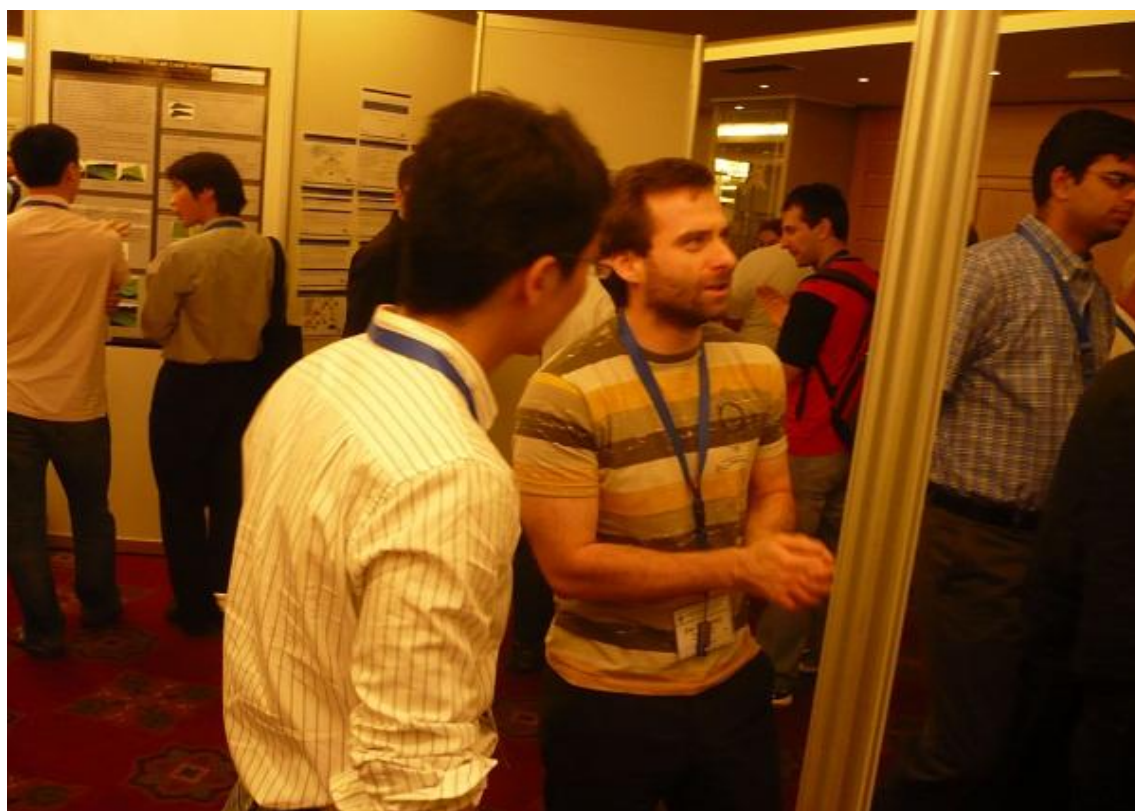
Účast na konferenci mírně komplikovaly nepokoje v Athénách.

Navázání kontaktů

Oliver Kennedy (*École Polytechnique Fédérale de Lausanne*) – používání fuzzy logiky v simulaci a analýze obchodních scénářů s nejistými daty.

Wei Zhang (*Oracle corporation*) – vědecká práce v ORACLE.

Fotografická dokumentace



1. Přílohy

Program konference