



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: IEEE ICDM 2011 (The 11th IEEE International Conference on Data Mining)
datum konání: 11. 12. - 14. 12. 2012
místo: Vancouver, BC, Canada
účastník konference: doc. RNDr. Vilém Vychodil, Ph.D.

Stručný popis konference:

IEEE ICDM je spolu s ACM SIGKDD jednou ze dvou největších konferencí, která se zabývá analýzou rozsáhlých dat, získáváním znalostí z dat, a data mining. Jedná se o vysoce selektivní konferenci, na které se setkává světová špička v oblasti dolování znalostí z dat, zástupci z předních světových výzkumných univerzit a výzkumných laboratoří velkých softwarových firem. Týmu vědců z katedry informatiky se podařilo v této konkurenci uspět. Letošní ročník sponzorovali NSF (National Science Foundation), Google, IBM Research a NASA.

Zajímavá čísla

Acceptance rate:	18%
Počet účastníků:	více jak 600
Počet prezentovaných příspěvků:	148
Počet plenárních přednášek:	3
Počet paralelních sekcí	4 (z toho 1 sekce tutoriálů)

Zajímavé přednášky

A. Plenární přednášky

Dr. Cynthia Dwork, Distinguished Scientist, Microsoft Research
The Promise of Differential Privacy

Přednáška pojednávala o principech zachování (skrytí) soukromých informací při možnosti analyzovat informace v datech a zabývala se především algoritmickými aspekty tohoto problému. Navržené metody umožňují do určité míry analyzovat citlivá data bez toho aniž by bylo potřeba zavádět speciální bezpečnostní, legislativní, nebo jiné opatření na jejich ochranu.

http://icdm2011.cs.ualberta.ca/downloads/Dwork_Keynote.pptx

Dr. C. Lee Giles

Data Mining and Information Extraction for CiteSeerX and Friends

Přednáška popisovala open-source architekturu SeerSuite, která představuje modulární a rozšiřitelný systém postavený na úspěšných projektech operačních systémů jako jsou Lucene-Solr a zabývala se specifickými problémy vytváření vyhledávacích a analytických enginů pro vědecká data a pro použití v akademické sféře.

Dr. Renée J. Miller

On Schema Discovery

Přednáška se věnovala tématice dolování informací ze strukturovaných dat. Strukturovanými daty byla myšlena data, která jsou vybavena dodatečnou informací (zvanou schéma) popisující logickou strukturu a sémantiku dat. Přednáška se soustředila na nový význam schémat zejména s přibývajícím rolí dat získávaných z webu. Na přednášce se mimo jiné objevila tematika funkčních závislostí v datech.
http://icdm2011.cs.ualberta.ca/downloads/Miller_Keynote.pdf

B. Vybrané řádné přednášky

Eirini Spyropoulou, Tijl De Bie: Interesting Multi-Relational Pattern.

Příspěvek ukazuje přístup k dolování informací z relačních databází, které nejsou omezené pouze na binární objekt-atributová data.

Xufei Wang, Jiliang Tang, and Huan Liu: Document Clustering via Matrix Representation

Příspěvek ukazuje maticovou reprezentaci dokumentů a získávání informací o dokumentech pomocí nezáporné faktorizace matic.

Pauli Miettinen: Boolean Tensor Factorizations

Příspěvek ukazuje přístup k faktorizaci tenzorů, svým tématem je příbuzný k výsledkům, které byly získány na katedře informatiky. Příspěvek obsahuje teoretický vzhled do problému, algoritmy a jejich srovnání.

Miao Zhang, Chris Ding: Tensor Fold-in Algorithms for Social Tagging Prediction

Příspěvek ukazuje aplikaci tenzorových dekompozic pro zpracování dat získávaných ze sociálních sítí skládající se z informací o uživatelích, položkách (items) a jejich visáčkách (tags).

Context-Aware Multi-Instance Learning based on Hierarchical Sparse Representation: Interesting Multi-Relational Pattern

Příspěvek ukazuje přístup k dolování informací z relačních databází, které nejsou omezené pouze na binární objekt-atributová data.

Vlastní prezentace

P. Krajča, J. Outrata, V. Vychodil: Using Frequent Closed Itemsets for Data Dimensionality Reduction.

Přednáška trvala 20 minut. V diskuzi padl dotaz na vztah k tiling databases.

Shrnutí konference (perspektivní témata apod.)

Na konferenci byly prezentovány příspěvky z tradičních oblastí data miningu, například dolování asociačních pravidel, redukce dimenzionality dat, získávání informací z grafů, analýza sekvencí a aplikace v bioinformatice, klasifikace

a shlukování. Dále byly předneseny příspěvky i v oblastech, které jsou příbuzné, jako je například získávání informací z obrazu a podobně. V souběhu s hlavní konferencí se konalo 19 workshopů, obvykle v rozsahu jednoho dne nebo půl dne; dále byly pořádány čtyři tutoriály. Součástí konference byla i soutěž ve vytvoření modelu pro predikci počtu editací stránek na Wikipedii (soutěže se zúčastnilo 193 příspěvovatelů).

Navázání kontaktů

Na konferenci byl navázán kontakt s P. Miettinenem (Max Planck Institute for Informatics, Saarbrücken, Německo), který se věnuje problematice booleovských dekompozic matic a tenzorů. Kontaktovaný vědec předběžně souhlasil s prezentací výsledků na vědeckém semináři pořádaném MCIN.

Přílohy

Program konference

Seznam publikovaných příspěvků:

<http://www.informatik.uni-trier.de/~ley/db/conf/icdm/icdm2011.html>