



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Mezinárodní centrum pro informaci a neurčitost

Zkušenosti z konferencí a stáží v rámci projektu

Vilém Vychodil

22. března 2012

ICDM 2011

Co, kdy, kde, kdo

název konference:

IEEE ICDM 2011 (The 11th IEEE International Conference on Data Mining)

datum konání:

11.–14. 12. 2012

místo:

Vancouver, BC, Canada

účastník konference:

doc. RNDr. Vilém Vychodil, Ph.D. (já)

Charakteristika

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)

Plenární přednášky

Dr. Cynthia Dwork, Distinguished Scientist, Microsoft Research

The Promise of Differential Privacy

Principy zachování (skrytí) soukromých informací při možnosti analyzovat informace v datech; algoritmické aspekty problému.

Dr. C. Lee Giles

Data Mining and Information Extraction for CiteSeerX and Friends

Open-source architektura SeerSuite, specifickými problémy vytváření vyhledávacích a analytických enginů pro vědecká data.

Dr. Renée J. Miller

On Schema Discovery

Dolování informací ze strukturovaných dat (data vybavená dodatečnou informací popisující „logickou strukturu a sémantiku dat“)

Vybrané řádné přednášky

Eirini Spyropoulou, Tijl De Bie:

Interesting Multi-Relational Pattern

Xufei Wang, Jiliang Tang, and Huan Liu:

Document Clustering via Matrix Representation

Pauli Miettinen:

Boolean Tensor Factorizations

Miao Zhang, Chris Ding:

Tensor Fold-in Algorithms for Social Tagging Prediction

P. Krajča, J. Outrata, V. Vychodil:

Using Frequent Closed Itemsets for Data Dimensionality Reduction

THE BOOLEAN CP TENSOR DECOMPOSITION

$$x_{ijk} \approx \bigvee_{r=1}^R a_{1r} b_{jr} c_{kr}$$

mpc

Constraint Discovery

14

- Functional Dependencies (includes Keys)
 - FDEP [Flach, Savnik AICommunications99], TANE [Huhtala et al. Computing99], FastFDs [Wyss et al. DaWaK01]
- Inclusion Dependencies (includes Foreign Keys)
 - General Inclusion Deps [Bouckmann+07, Die Marchi+09]
 - Foreign Keys [Zhang et al. PVLDB10]
- Mining may give large number of dependencies
 - not always intuitive
 - useful for mining the data

Itemset Mining for Multi-Relational Databases (MRDs)

- frequent itemset mining on the join of all tables²

²[K. Ng et al., 2002; A. Koopman et al., 2005; ... et al., 2010]

