

POZVÁNKA na 26. setkání PRAŽSKÉHO INFORMATICKÉHO SEMINÁŘE

# TOMÁŠ SKOPAL:

## Podobnostní vyhledávání v nestrukturovaných datech



*po přednášce bude následovat diskuse*

### ANOTACE PŘEDNÁŠKY

V dnešní době velkých dat („Big Data“) se stále častěji setkáváme s daty, která pocházejí ze senzorů digitalizujících „signály přírody“ a jejichž technická datová struktura slouží pouze k manipulaci a reprodukci. Nejčastěji mezi tato nestrukturovaná data řadíme multimédia (obraz, zvuk), ale senzorová data mohou být daleko rozmanitější. Pro vyhledávání v nestrukturovaných datech se používají abstraktní podobnostní modely, kde jsou datové entity reprezentovány doménově specifickými deskriptory (např. vysokodimenzionálními vektory, časovými řadami nebo řetězci). Podobnost entit se pak měří jako vzdálenost jejich deskriptorů a celý problém je tak geometrizován jako vyhledávání nejbližších deskriptorů k deskriptoru dotazového objektu. Geometrie podobnostních prostorů je velmi důležitá pro indexování, tj. pro databázovou techniku urychlující vyhledávání, ale také pro samotné modelování podobnosti a deskriptorů. V přednášce ukážeme, že obvyklé euklidovské vnímání prostoru není zdaleka jedinou možností reprezentace, velmi používaný je obecnější metrický model. Lze dokonce vyrobit unikátní vzdálenostní prostory, jejichž topologické vlastnosti jsou přímo odvozeny z dat. Budeme také diskutovat otázky spojené s modelováním podobnosti, zejména problematiku sémantických deskriptorů vs. chytrých podobností.

### O PRAŽSKÉM INFORMATICKÉM SEMINÁŘI

Seminář se schází vždy 4. čtvrtek v měsíci v 16 hod. (s výjimkou letních měsíců a prosince), a to buď v budově FEL ČVUT na Karlově náměstí, nebo v budově MFF UK na Malostranském náměstí. Jeho program je tvořen hodinovou přednáškou, po níž následuje časově neomezená diskuse. Základem přednášky by mělo být něco (v mezinárodním měřítku) mimořádného nebo aspoň pozoruhodného, na co přednášející přišel a co vysvětlí způsobem srozumitelným a zajímavým i pro širší informatickou obec. Přednášky jsou standardně v angličtině.



**Tomáš Skopal** se věnuje výzkumu v oblasti podobnostního vyhledávání a tématům okolo multimediálních databází a information retrieval. Je docentem a vedoucím katedry softwarového inženýrství MFF UK. V roce 2006 založil úspěšnou výzkumnou skupinu SIRet (SImlarity RETrieval), kterou až do současnosti vede. Magisterský diplom získal na Univerzitě Palackého v Olomouci, doktorát na VŠB - Technické univerzitě v Ostravě. Od té doby působí v Praze na Univerzitě Karlově a pracoval také jako hostující profesor a výzkumník na Univerzitě v Kostnici (Německo) a na DCC, University of Chile v Santiagu.

**Kontakt:** [info@praguecomputerscience.cz](mailto:info@praguecomputerscience.cz)

**Informace:** [www.praguecomputerscience.cz](http://www.praguecomputerscience.cz)