



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 stáži

|                        |                                                 |
|------------------------|-------------------------------------------------|
| Datum konání stáže:    | 25.2.2013 – 24.4.2013                           |
| Navštívené pracoviště: | University of Wollongong, Wollongong, Austrálie |
| Zahraniční garant:     | prof. Peter Eklund                              |
| Účastník stáže:        | Mgr. Jan Outrata, Ph.D.                         |

#### Stručný popis navštíveného pracoviště

Centrum pro digitální ekosystémy (Centre for Digital Ecosystems) je výzkumnou skupinou na School of Information Systems & Technology, Faculty of Engineering and Information Sciences, University of Wollongong, která se zabývá výzkumem systémů „podporujících interakci a zapojení ve spolupracujícím prostředí“. V rámci aktuálního systému CollectionWeb pro katalogizaci dat a navigaci v digitálních multimediálních knihovnách a kolekcích jde o kombinaci metod zpracování přirozeného jazyka (natural language processing) a dolování dat (data mining), zejména formální konceptuální analýzy (FCA). Ukázkovým projektem nového konceptu katalogizace a navigace je zde známé virtuální muzeum kulturní kolekce tichomoří, Virtual Museum of the Pacific. Pracoviště je známé a významné právě pro aplikaci FCA v oblasti informačních systémů, je prvním s úspěšnými projekty katalogizace, vyhledávání a navigace v nich (další známý projekt je SearchSleuth, rozšíření internetového vyhledávače o navigaci založenou na FCA). Centrum tvoří 1 profesor a 4 doktorandi, na škole je kolem 30 pracovníků (z toho 1 profesor a 5 docentů) a 50 doktorandů. Výzkum členů centra a školy je financován převážně na univerzitní úrovni státem (New South Wales), částečně také průmyslovým sektorem.

Během stáže jsem spolupracoval a diskutoval vědecké problémy s prof. Peterem Eklundem a jeho doktorandem Timem Wrayem. Prof. Eklund vede centrum a dlouhodobě se zabývá vedle rozvoje samotné FCA (reprezentace a vizualizace konceptuálních svazů, podobnost formálních konceptů, ontologie) zejména aplikacemi a využitím FCA v oblasti informačních systémů a webu (information retrieval s využitím FCA, navigace v konceptuálních svazech, social tagging). V této oblasti dosahuje se svými doktorandy a spolupracovníky z různých pracovišť v Austrálii (University of Queensland, R. Cole) i jinde ve světě, např. TU Darmstadt (R. Wille), SAP Research Dresden (F. Dau), KU Leuven (J. Poelmans), LORIA

Nancy Francie, a nově i UP v Olomouci (J. Outrata), významných výsledků. Pro svůj výzkum získává i regionální zdroje.

### **Průběh stáže**

S prof. Eklundem a jeho doktorandem Timem Wrayem jsem, spolu s Martinem Trnečkou z katedry informatiky PřF UP, který byl na stejném pracovišti v prvním měsíci stáže na stáži také, diskutoval oblasti jejich aktuálního výzkumu na téma tvorba a navigace v digitální multimediální kolekci muzejních dat s využitím FCA (v rámci systému CollectionWeb). Po zjištění, že pro výpočet množiny formálních konceptů (konceptuálního svazu) je použit algoritmus PCbO, jehož jsem spoluautorem, bylo prakticky ihned dohodnuto vylepšení algoritmu pro jejich potřeby, do podoby algoritmu aktualizujícího množinu formálních konceptů při inkrementální úpravě vstupních dat (přidání, změna atributů a odebrání objektů). Důležitým aspektem nového algoritmu bylo to, že pro aktualizaci nevyžaduje množinu konceptů v datech před úpravou, počítá pouze rozdíl mezi novou a předchozí množinou. Nový algoritmus jsem vyvinul (jako úpravu našich algoritmů PCbO, FCbO a PFCbO založených na algoritmu CbO) a implementoval a po jeho prezentaci a demonstraci prof. Eklundovi a Timovi bylo dohodnuto vytvoření společné publikace na téma výkonnostní vyhodnocení algoritmu na rozsáhlejších reálných (muzejních) datech, která bude zaslána na konferenci CLA 2013. Moje část publikace (popis algoritmu) byla spolu s implementací algoritmu vytvořena během stáže. Prodiskutována byla během stáže integrace algoritmu do jejich systému CollectionWeb a jejich část publikace. Dále dohodnuta byla úprava algoritmu do podoby distribuovaného algoritmu, ve které bude moci být lépe využit v jejich výzkumu. Během stáže jsem také vyvinul (významné) rozšíření našich algoritmů založených na algoritmu CbO o výpočet relace uspořádání v konceptuálním svazu, tj. do podoby algoritmu pro výpočet konceptuálního svazu, a rovněž také sloučení s předchozí úpravou do podoby algoritmu pro aktualizaci konceptuálního svazu (bez požadavku konceptuálního svazu dat před úpravou).

Dalším diskutovaným tématem s prof. Eklundem byla (úspěšně obhájená) diplomová práce studenta Lukáše Havrlanta z katedry informatiky PřF UP pod mým vedením na téma rozšíření (internetového) vyhledávače o navigaci založenou na FCA, tzn. práce přímo vycházející z a navazující na projekt SearchSleuth prof. Eklunda. Diskutováno bylo vzájemné využití a spojení projektů do jednoho nového i vytvoření další společné publikace na toto téma.

Na doporučení prof. Eklunda jsem také, spolu s Martinem Trnečkou, diskutoval s doktorandem prof. Eklunda Peterem Goodallem v Sydney možnosti využití našeho výzkumu Booleovské faktorové analýzy (BFA) s využitím FCA v jeho výzkumu na téma štitkování (tagging) muzejních dat multimediálních kolekcí. Zde se přímo nabízí vytvoření štitků jako faktorů.

S Martinem Trnečkou jsem po celou dobu jeho stáže konzultoval jeho výzkum na téma BFA a algoritmy pro BFA.

### **Přednášky na navštíveném pracovišti**

Nebyla přednesena žádná přednáška.

### **Publikace rozpracované během stáže**

*P. Eklund, J. Outrata, T. Wray: Evaluation of an Algorithm for Updating Formal Concepts based on Incremental Changes to Museum Collection Metadata. 12 pp., submitted to CLA 2013.*

Článek, zaslaný na mezinárodní konferenci CLA 2013 konanou v říjnu 2013 v La Rochelle, Francie, prezentuje výkonnostní vyhodnocení nového algoritmu pro aktualizaci množiny formálních konceptů při inkrementální úpravě vstupních dat (přidání, změna atributů a odebrání objektů) na rozsáhlejších reálných (muzejních) datech. Algoritmus pro aktualizaci nevyžaduje množinu konceptů v datech před úpravou, počítá pouze rozdíl mezi novou a předchozí množinou.

*J. Outrata: A lattice-free concept lattice update algorithm based on  $\square$ CbO. 14 pp., submitted to CLA 2013.*

Článek, zaslaný na mezinárodní konferenci CLA 2013 konanou v říjnu 2013 v La Rochelle, Francie, představuje nový algoritmus pro aktualizaci konceptuálního svazu při inkrementální úpravě vstupních dat (přidání, změna atributů a odebrání objektů). Algoritmus pro aktualizaci nevyžaduje konceptuální svaz dat před úpravou, počítá pouze rozdíl mezi novým a předchozím svazem.

### **Navázání kontaktů**

Stáž velmi významně přispěla k získání nových kontaktů a navázání vědecké spolupráce se skupinou prof. Eklunda. Byl diskutován a významně rozšířen stávající výzkum a navržena nová témata pro jeho další pokračování, zahájena spolupráce na jejich výzkumu a diskutována témata pro další vzájemnou spolupráci.

### **Shrnutí stáže**

Stáž ve více než plné míře splnila svůj účel. Byla navázána nová vědecká spolupráce s jedním z významných pracovišť pro aplikaci formální konceptuální analýzy (FCA) v oblasti informačních systémů. Během stáže byly rozpracovány dvě publikace zaslané na mezinárodní konferenci.

### **Fotografická dokumentace**



Fotografie po prezentaci a demonstraci nového algoritmu a dohodě na vytvoření společné publikace. Na snímku Tim Wray (vlevo), prof. Eklund (uprostřed) a dr. Outrata (vpravo).

*For Orl*