

SPOLUFINANCOVÁNO ZE STRUKTURÁLNÍCH FONDŮ EU (EVROPSKÉHO FONDU PRO REGIONÁLNÍ ROZVOJ) PROSTŘEDNICTVÍM IOP

Zpráva ze služební cesty
Projekt „Vytvoření Národní digitální knihovny“

CZ 1.06/1.1.00/07.06386

Jméno a příjmení účastníka cesty	PhDr. Ladislav Cubr
Pracoviště - dle organizační struktury	Úsek digitálních fondů a technologií
Pracoviště - zařazení	Pracovník Oddělení pro standardy
Důvod cesty	Konference TPDL 2011
Místo - město	Berlín
Místo - země	Německo
Datum (od - do)	25.-29.9.2011
Podrobný časový harmonogram	25.9. cesta Praha - Berlín 26.-28.9. konference 29.9. cesta Berlín Praha
Spolucestující z NK	Mgr. Jiří Kremser
Finanční zajištění	IOP „Vytvoření Národní digitální knihovny“
Vztah k projektu	
Cíle cesty	Návštěva konference, konzultace s kolegy
Plnění cílů cesty	Všechny cíle splněny
Další podrobnější informace	Viz text níže
Podpora publicity projektu	Propagace projektu při komunikaci se zahraničními kolegy

Související materiály	
Materiál	Místo uložení

SPOLUFINANCOVÁNO ZE STRUKTURÁLNÍCH FONDŮ EU (EVROPSKÉHO FONDU PRO REGIONÁLNÍ ROZVOJ) PROSTŘEDNICTVÍM IOP

Datum předložení zprávy	10.10.2011
Podpis předkladatele zprávy	

Datum	Podpis
-------	--------

Podpis nadřízeného

10.10.2011

Vloženo na intranet

Přijato v mezinárodním oddělení

Cestovní zpráva

Výše uvedení pracovníci Národní knihovny ČR se zúčastnili 15. evropské konference věnované výzkumu a vývoji v oblasti digitálních knihoven, která se konala ve dnech 25.-29. září 2010 v Berlíně v Německu na Humboldtově univerzitě (Humboldt-Universität). Konference byla letos přejmenována na „International Conference on Theory and Practice of Digital Libraries“ (zkratka TPDL), předchozí název byl „The European Conference on Research and Advanced Technology for Digital Libraries“ (zkratka ECDL). Důvodem změny byla kolize názvu s evropským certifikačním programem počítačové gramotnosti, který je znám též po zkratku „ECDL“ (European Computer Driving Licence) a který je mezinárodně známější.

Pracovníci se zúčastnili hlavního konferenčního programu. Letošní program byl zaměřen zvláště na oblast průniku mezi digitálními knihovnami a vědeckým a vzdělávacím sektorem.

Níže uvádíme přehled nejdůležitějších příspěvků z hlediska zaměření výše uvedených pracovníků a příbuzných oblastí ohodnocených jako relevantních.

Relevantní příspěvky

V úvodní řeči byla prezentována potřeba větší spolupráce mezi digitálními knihovnami a vědeckým a vzdělávacím sektorem. Důležitým faktorem pro rozvoj digitálních knihoven bude dlouhodobá zpřístupnitelnost a interoperabilita vědeckých dat.

Thomas Hofmann, výzkumník pracující pro Google prezentoval vizi prohlížení webu napříč jazykovými bariérami a představil posuny ve vývoji automatizovaného překladu. Oznámil též, že slibný projekt Google Squared byl v září tohoto roku pozastaven.

Ray Larson v příspěvku „Connecting Archival Collections: The Social Networks and Archival Context Project“ prezentoval velmi zajímavý projekt SNAC, řešící problémy vzniklé ze slučování autoritních údajů z několika kolekcí. Pro desambiguaci a spojování pojmenovaných entit (v tomto případě názvů autorů) využívá Larson kombinaci klasických metod, tedy Naive Bayes, editační

SPOLUFINANCOVÁNO ZE STRUKTURÁLNÍCH FONDŮ EU (EVROPSKÉHO FONDU PRO REGIONÁLNÍ ROZVOJ) PROSTŘEDNICTVÍM
IOP

vzdálenost a multinomialni model. Získaná data poté vizualizuje formou sociální sítě ve své prototypové aplikaci. Jeho příspěvek tak pomáhá řešit problémy duplicity v záznamech atd.

Hamed Alhoori v příspěvku „Understanding the Dynamic Scholarly Research Needs and Behavior as Applied to Social Reference Management“ varoval před tím, že akademické knihovny v současnosti nejsou hybatelem ve zvyšování povědomí o dostupných výzkumných technologiích.

Pro NK patrně nejvýznamnější z posterů pro naše potřeby byl příspěvek „Interconnecting DSpace and LOCKSS“ od Husseina Sulemana informující o způsobu zapojení systému postaveného na DSPACE do distribuovaného archivačního systému LOCKSS, který je zaměřen na dlouhodobou ochrany digitálních publikací provozovanou konsorciem knihoven a vydavatelů. Ve stručnosti. Využívá protokolu OAI-PMH pro komunikaci mezi jednotlivými DSPACE repozitáři a součástí metadat vystavených přes tento protokol je také odkaz na stažení digitálního objektu, což umožňuje replikaci pro systém LOCKSS. Po diskuzi s Husseinem jsme bohužel nedokázali posoudit, zda podobné řešení lze aplikovat na Fedoru, nicméně žádné nepřekonatelné problémy na první pohled nevystaly.

Dalším důležitým příspěvkem pro naše využití byl „An Organizational Model for Digital Library Evaluation“ Michaela Khooa and Craiga MacDonalda, představující nástroj pro hodnocení organizačních procesů v digitálních knihovnách. Tento nástroj by bylo možná vhodné využít pro hodnocení digitálních knihoven typu Kramerius nebo Manuscriptorium.

Pro potřeby dlouhodobé ochrany byly nejvýznamnější dva příspěvky. „Coherence-Oriented Crawling and Navigation Using Patterns for Web Archives“ Myriama Bena Saada, analyzující řešení pro zajištění koherence sklizeného obsahu. Daniel Gomes v příspěvku "A Survey on Web Archiving Initiatives" představil statistický výzkum existujících webarchivů, sbírající např. informace o ukládaných formátech.

Zajímavým příspěvkem byl „How to Become a Group Leader? or Modeling Author Types Based on Graph Mining“, který prezentoval George Tsatsaronis, a který pojednává o modelování spoluautorství formou power grafu, tedy konceptu z teorie grafu, vhodného pro sociální sítě, nebo obecněji pro takové grafy, kde je velmi velký počet hran. Pojmenoval zajímavé struktury a časté vzory v power grafech a zaměřil se zejména na evoluci autora v čase a na rady směřované autorům o tom jak se stát vysoce citovaným autorem.

Uterní blok věnovaný problematice Linked Open Data (LOD) zahájil Naimdjon Takhirov, článkem „Linking FRBR Entities to LOD through Semantic Matching“, ve kterém se zaměřoval, zjednodušeně řečeno, na spojování entit napříč různými systémy, v oblasti sémantického webu známého též jako Ontology Alignment/Matching. Jako referenční ontologie používal autor ontologie Dbpedia a Freebase. Prezentoval taky svoji formu metriky, kterou pak určoval podobnost entit. Tato metrika byla založena na známých metrikách Jaro Winkler, Monge Elkan (edit distance) a Svaled Levenshtein a byla jakousi kombinací těchto tří.

Další prezentací z oblasti LOD byl příspěvek „Interactive Vocabulary Alignment“ spjatý s projektem Europeana, který prednesl Jacco van Ossenbruggen. Příspěvek na mne působil spíše dojmem, že se snaží posluchače v dané oblasti vzdělat, než že by přinášel něco nového. Týkal se také problematiky Ontology Alignment, ale v porovnání s předchozím příspěvkem působil

SPOLUFINANCOVÁNO ZE STRUKTURÁLNÍCH FONDŮ EU (EVROPSKÉHO FONDU PRO REGIONÁLNÍ ROZVOJ) PROSTŘEDNICTVÍM IOP

nedodělaně. Jejich přístup využíval WordNet a uměle vytvořenou hru, která měla nadchnout uživatele, aby formou zábavy dělal užitečné věci, tedy spojoval podobné koncepty. Bohužel vizuální stránka této hry spíše odrazovala a největší přínos článku tak vidím v představení iniciativy pro spojování ontologií tzv. Ontology Alignment Evaluation Initiative (OAEI) .

Další příspěvek, také z oblasti sémantického webu a LOD byl ten s názvem „Transformation of a keyword indexed collection into a semantic repository: applicability to the urban domain“, kde Španěl Javier Lacasta prezentoval svou bohužel ne úplně srozumitelnou angličtinou výsledky své práce o ontologii s názvem DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering). Tato ontologie se snaží být tzv. Upper Ontology, tedy ontologií ne úzce zaměřenou na určitou doménu, ale na velmi obecné a abstraktní koncepty. Téma bylo velmi zajímavé, což ostatně potvrzuje i fakt, že autor si odnesl cenu za nejlepší článek konference, bohužel ale také velmi nesrozumitelné kvůli už zmíněné velmi špatné výslovnosti prezentátora a také velmi vysoké obtížnosti a abstraktnosti obsahu článku. Ve stručnosti ale hlavním motivem pro vznik článku, byl deficit komunikace mezi akademickou a průmyslovou sférou a potřeba zachycení těchto požadavků formou konceptuálního modelování.

Posledním užitečným článkem, který můžeme využít také, byl „DAR: Institutional Repository Integration in Action“, tedy případová studie z Egyptské národní knihovny, kterou prezentoval Youssef Mikhail. Zajímavý je zejména fakt, že chtějí do budoucna napojit svůj systém na LTP systém iRODS. Prezentace dostupná z goo.gl/Xca0L. U neformální večeře jsem s Youssefem pak konzultoval řešení, které máme v MZK. Zajímavým faktem také bylo, že narazili na úplně stejný problém jako Incad s Fedorou, tedy na nedostačující výkon sémantického uložiště (triple store) Mulgara, které má ale jinak spoustu výhod. Incad se s tímto faktem bohužel vyrovnal tak, že rezignoval na sémantické uložiště a tedy na jakoukoliv budoucí interoperabilitu se světem LOD, zatímco DAR využili sémantické uložiště 4Store a mohou tak dotazovat své digitální objekty jazykem SPARQL a vystavit veřejná data komunitě sémantického webu. Jejich řešení také docela dobře škáluje, což je do budoucna možné riziko našich (MZK/NK) Fedor.

Jako zajímavost uvedme, že cenu za nejlepší poster získal český student studující v Británii na Univerzitě za prezentaci systému CORE (COncnecting REpositories) určený pro zpřístupňování a navigaci mezi vědeckými dokumenty uloženými v digitálních repozitářích založených na Open Access.

V panelu věnovaném otázce volného online zpřístupňování dokumentů nacházejících se v oblasti veřejného sektoru byly představeny aktuální problémy související s ochranou duševních práv. Byl představen nástroj „Public Domain Calculation“ (<http://outofcopyright.eu/>), pomáhající knihovnám odhadnout, zda se na jimi uchovávaný dokument pravděpodobně vztahující či nevztahující podmínky pro dokumenty ve veřejném sektoru.