

## Zpráva ze služební cesty

Ve dnech 31.5. až 3.6. jsem se zúčastnil pravidelné výroční konference ELAG (European Library Automation Group), kterou v tomto roce pořádala knihovna fyzikálního institutu CERN (Ženeva, Švýcarsko - viz <http://elag2005.web.cern.ch/elag2005>). O den dříve, 30.5. jsem se v knihovně CERNU účastnil pracovní zchůzky přípravné a organizační skupiny ELAG, na které byl upřesňován program pro letošní rok a vytvářen nový program pro konferenci ELAG, která se bude příští rok konat v Rumunské Bukurešti.

Konference byla stejně jako v předchozích letech organizovaná tak, že účastníci byli rozděleni do malých skupin a každá skupina se zabývala svým tématem. Dvakrát až třikrát denně se všichni účastníci scházeli na společných sekcích, kde byly prezentovány informační systémy, postupy a technologie, relevantní k probíraným tématům a kde také zástupci jednotlivých skupin prezentovali výstupy z jednání a diskuzí v těchto skupinách (workshopech).

Účastnil jsem se workshopu „Metasearch systems versus Google Scholar“, ve kterém se probíraly a srovnávaly výhody a nevýhody dvou různých přístupů k vyhledávání informací na Internetu. První přístup, který se obvykle označuje jako „metasearching“ spočívá v tom, že dotaz na informaci je rozeslán na více zdrojů současně a jeho zpracování probíhá na těchto zdrojích simultánně v reálném čase. Uživatel poté obdrží odpověď v homogenizované podobě, jako by přišla z jednoho zdroje. Tato metoda se označuje jako vyhledávání „just in time“. Druhý přístup je používán nástoji jako Google, Yahoo, atp. Spočívá v tom, že tyto nástroje provádějí pravidelný periodický sběr dat z mnoha různých zdrojů po celém světě (tzv. harvesting). Harvesting probíhá obvykle jednou denně a odkazy na informace takto získané jsou pak ukládány a indexovány v jedné databázi, kde jsou pomocí výše uvedených nástrojů přístupné, jako by pocházely z jediného zdroje. Tento přístup se označuje jako vyhledávání „just in case“. Rerezentantem prvního přístupu je například systém Metalib od Izraelské firmy ExLibris. Česká instance tohoto systému je přístupná na adrese <http://metalib.cuni.cz>. Neznámějším zástupcem druhé metody je už zmíněný Google (<http://www.google.com>).

Diskuze ve workshopu „Metasearch systems versus Google Scholar“ probíhala podle mého názoru poněkud jednostranným způsobem ve prospěch tzv. „metasearchingu“. Pravděpodobně to bylo dáno tím, že workshop moderovala zástupkyně firmy ExLibris, paní Tamar Sadeh. Tato firma vyvíjí a distribuuje jeden z neznámějších „metasearch“ systémů Metalib (viz výše). Jako hlavní výhoda metasearchingu byl v diskuzi stavěn fakt, že u tohoto způsobu vyhledávání má uživatel plnou kontrolu nad výběrem databázových

zdrojů. Tento argument byl kvitován hlavně knihovníky, kteří se workshopu účastnili. Od techničtější zaměřených účastníků zazníval jako hlavní protiargument fakt, že naprostá většina uživatelů potřebuje při vyhledávání informace rychlou odezvu a nezajímá je příliš, z jakých zdrojů informace přišla. Osobně se přikláním ke druhému argumentu. Nemyslím si ale, že je možné v současném stadiu vývoje jednoznačně říct, který postup je lepší. Oba mají své výhody a nevýhody a každý slouží něčemu trochu jinému, ale ukazuje se, že naprostá většina současných uživatelů preferuje druhý postup, tedy co nejjednodušší dotaz za kterým následuje velmi rychlá odpověď. Pokud je odpověď příliš rozsáhlá, dotaz se v dalším kroku upřesní, až dostaneme přijatelně malé výsledky. Podle mé dosavadní zkušenosti myslím, že vývoj vyhledávacích systémů, určených pro širokou veřejnost se bude dále ubírat spíše cestou, po které jde v současné době vyhledávač Google, ale vedle toho budou pravděpodobně i nadále existovat metasearch systémy, které budou pravděpodobně sloužit užší skupině lidí, kteří ke své práci potřebují i údaje o „holdingu“ nalezených informací a dokumentů a potřebují (nebo si myslí že potřebují) mít kontrolu nad výběrem prohledávaných zdrojů.

Hlavní přínos letošního ELAGu pro mě ležel trochu jinde než ve vlastním programu konference. Pro letošní rok získalo ICUP grant z FRVŠ (2938/2005/a/c). Mým příspěvkem projektu je vybudování a provoz systému, na kterém se budou postupně vystavovat digitalizované dokumenty, které jsou příliš cenné, aby se půjčovaly běžným způsobem v knihovně. Průzkumem stávajících systémů jsme s kolegy koncem loňského roku dospěli k řešení, které je už 10 let vyvíjeno v CERN Computer Centre. Programátoři z CERNu vyvinuli systém CDSWare (CERN Document Server – viz <http://cdsware.cern.ch>), který původně sloužil jako preprintový server pro vystavování zatím nepublikovaných vědeckých prací fyziků z CERNu. Za deset let vývoje se systém CDSWare změnil ve velmi robustní a výkonný nástroj pro budování digitální knihovny a zdá se být i velmi vhodným nástrojem pro katalogizaci a vystavování tzv. „šedé literatury“ (diplomové a dizertační práce atp.). Systém CDSWare dali jeho tvůrci k dispozici světu zdarma včetně jeho zdrojových kódů. Systém je dnes velmi propracovaný, ale má jednu nevýhodu, společnou pro hodně podobných „open source“ a „free“ projektů. Celý systém sestává z rozsáhlé komplikované sady mnoha python-scriptů, PHP-scriptů a programů napsaných v jazyce C a celý je velmi úzce navázán na www server Apache a relační databázový stroj MySQL. Proces jeho instalace je velmi komplikovaný a těžkopádný. Pokoušel jsem se nainstalovat a nakonfigurovat systém přibližně od konce února do konce května a velmi pracně jsem dospěl ke stavu, kdy cca 80% systému bylo plně funkčních. V CERNu jsem měl možnost strávit téměř dva dny v CERN Computer Centr s hlavním vývojářem systému Tiborem Šimkem a během cca 3 hodin jsme s jeho pomocí systém kompletně

nainstalovali a nakonfigurovali na mém notebooku. Následnými diskuzemi mě pan Tibor Šimko plně saturoval podrobnými informacemi o struktuře, fungování a správě systému. V těchto dnech jsem s Tiborem Šimkem nadále ve spojení a naše diskuze i práce na rozjetí CDSWare na výkonném serveru intenzivně pokračuje.

Když mluvím o přínosech mého pobytu v CERNu, nemohu pominout ani poněkud emotivní náboj. 4 dny jsem se pohyboval v místech, kde pod vedením pana Tima Bernerse-Lee v roce 1989 vzniklo www...

V Olomouci dne 7.6.2005

Aleš Horák