

NOVÁ DATABÁZE "CSA ILLUSTRATA: TECHNOLOGY" - KOMU JE URČENA?

Boris Škandera

Informetal, zájmové sdružení

V červnu 2008 firma ProQuest spustila databázi „CSA Illustrata: Technology“. Tato databáze využívá technologii tzv. „hlubokého indexování“ (deep indexing). Jaký je princip této technologie?

Zatímco doposud se při zpracování anotací a indexů vycházelo především z textové části informačních pramenů, „hluboké indexování“ se zaměřuje na indexaci „netextových informací“, tj. grafů, obrázků, tabulek, fotografií, apod. Jinými slovy, zpracovává dosud opomíjené informační bohatství uložené v grafických a tabulkových částech odborných článků či publikací.

Tento přístup přináší několik výhod, jako např.:

- „zviditelňuje“ dosud opomíjené cenné informace obsažené v pramenech - již na úrovni informačního záznamu (sekundární informace); uživatel tedy přesně vidí, jaké grafické informace daný text obsahuje již při prohlížení databáze a nikoliv až při získání a čtení vlastního primárního pramene
- podstatným způsobem zvyšuje počet indexačních termínů, čili rozšiřuje a zpřesňuje informační záznam a tím umožňuje i přesnější vyhledatelnost daného pramene
- do určité míry odbourává jazykové bariéry, neboť grafické informace jsou v naprosté většině případů mezinárodně srozumitelné

Jak již napovídá název zmíněné databáze, jejím obsahem jsou informace z oblasti technických věd.

Vzhledem k tomu, že naše zájmové sdružení Informetal zajišťuje informační služby zaměřené především na hutní průmysl, tato novinka nás velice zajímala a proto jsme si vyžádali bezplatný zkušební přístup k této databázi na 30 dnů.

Při naší práci využíváme zejména dvě dokumentografické databáze – naši vlastní databázi „METAL“ budovanou od r. 1972, a databázi firmy ProQuest „METADEX“, získávanou v rámci projektu MŠMT 1N04143 „Integrovaný informační a knihovnický systém pro rozvoj hutního průmyslu“.

Snažili jsme se zjistit využitelnost této nové databáze – zejména z pohledu našich uživatelů, jakož i provést určité srovnání všech tří zmíněných databází. Toto srovnání je samozřejmě poznamenáno tím, že jsme se zaměřili pouze na naši hutnickou tematiku, čili naše poznatky nelze absolutizovat či zevšeobecňovat pro jiné tematické oblasti.

Nejprve jsme se snažili porovnat uvedené databáze na základě seznamů periodik, které zpracovávají.

Porovnání počtů periodik zpracovaných v databázích „Metadex“ a „CSA Illustrata: Technology“ ukázalo, že překrytí obou seznamů představuje 1944 titulů (z toho ale pouze 111 titulů zaměřených převážně na hutnictví) – viz Tabulka č. 1.

Tabulka č. 1

Porovnání seznamů periodik zpracovávaných pro databáze „Metadex“ a „CSA Illustrata: Technology“ 06/2008

	CSA Illustrata: Technology	METADEX	Periodika zpracovávaná oběma databázemi	Periodika zpracovávaná pouze pro CSA Illustrata: Technology	Periodika zpracovávaná pouze pro METADEX
	Počet periodik	Počet periodik			
CELKEM	3 398	3 459	1 944	1 454	1 515

Copyright © 2008 ProQuest LLC. All rights reserved.

Průnik všech tří databází pak představuje pouhých 48 titulů. Poněkud překvapivým zjištěním pro nás bylo, že 61 titulů periodik je zpracováváno pouze naší tuzemské databázi „METAL“.

Dále jsme se zaměřili na praktickou využitelnost databáze „CSA Illustrata: Technology“ pro naše uživatele z oblasti hutního průmyslu, výzkumu a vývoje. Použili jsme pro tento účel rešerše zpracované v 1. pololetí 2008 na základě skutečných požadavků uživatelů. Všechny těchto 128 rešerší jsme se pokusili zpracovat v databázích „METAL“, „METADEX“, „CSA Illustrata: Technology“ a pro zajímavost i „CSA Illustrata: Natural Sciences“.

Po pravdě řečeno, výsledky nás poněkud zklamaly, neboť v databázi „CSA Illustrata: Technology“ bylo nalézáno v průměru o jeden řád méně informací, než v databázích „METADEX“ a „METAL“ – viz Tabulka č. 2.

Tabulka č. 2

Porovnání výsledků rešerší*			
Počet hitů v databázích			
METADEX	METAL	Illustrata: Technology	Illustrata: Natural Science
26887	15727	2267	440
100 %	58,5 %	8,4%	1,6%

* Porovnání provedeno na základě 128 stejných rešerší, zpracovaných ze všech těchto databází

Je to zřejmě způsobeno tím, že tato nová databáze není zatím (pro náš obor) dostatečně naplněna záznamy v potřebné retrospektivě, jinými slovy, je příliš „mladá“ pro uspokojivé využití. Byla by nepochybně dobrým doplňkem ke stávajícím databázím, ovšem pouze pro tento účel je její cena neúměrně vysoká.

Příjemným zjištěním pro nás byl naopak fakt, že databáze „METAL“ v počtech hitů nezaostává za „METADEXEM“ tolik, kolik by odpovídalo poměrné velikosti obou databází – viz Tabulka č. 3.

Tabulka č. 3

Porovnání objemů databází		
Databáze	celkový počet záznamů	objem v %
Metadex	cca 4 300 000	100,0
Metal	cca 200 000	4,7
CSA Illustrata: Technology	cca 1 500 000	34,9

Závěry

Jakkoliv je databáze „CSA Illustrata: Technology“ vsutku inovativním produktem, přinášejícím uživatelům řadu výhod, její informační pokrytí konkrétně pro potřeby uživatelů z oblasti hutnictví je zatím nedostatečné. Dá se očekávat, že tato situace se v průběhu příštích let zlepší a tato databáze se pak stane významným zdrojem informací.

Vzhledem k širokému tématickému záběru uvedené databáze v rámci technických oborů se jeví její vhodnost spíše pro konsorcium universit a vysokých

škol, jelikož se tak skýtají větší možnosti na využití této databáze v prostředí se širším víceoborovým záběrem.

Zájemcům o databáze využívající tzv. hluboké indexování typu Illustrata, které v současnosti pokrývají tematické oblasti „Technika“ a „Přírodní vědy“ (tematika „Lékařství“ se připravuje), doporučujeme vyžádat si bezplatný zkušební 30-denní přístup a prověřit si tak na praktických příkladech využitelnost těchto databází pro daný obor či okruh uživatelů.

Práce byla řešena v rámci projektu 1N04143 „Integrovaný informační a knihovnický systém pro rozvoj hutního průmyslu“ za finanční podpory MŠMT.

Literatura:

1. CSA Illustrata™ [online]. ProQuest, c2008 [cit. 2008-09-10]. Dostupné na World Wide Web: <<http://www.proquest.com/promos/product/csainstrata.shtml>>.
2. GOLDEN, Alan C. CSA Illustrata: Enhancing Search and Discovery Through Deep Indexing & New Products from ProQuest CSA. In INFORUM 2007: 13. ročník konference o profesionálních informačních zdrojích, Praha 22.-24. května 2007 [online]. Praha : Albertina icome Praha, 2007 [cit. 2008-09-10]. Dostupné na World Wide Web: <<http://www.inforum.cz/archiv/inforum2007/en/proceedings/59/>>. ISSN 1801-2213.
3. LAURIDSEN, Helle. Exploit Your E-collections to Full. In INFORUM 2008: 14. ročník konference o profesionálních informačních zdrojích, Praha 28.-30. května 2008 [online]. Praha : Albertina icome Praha, 2008. [cit. 2008-09-10]. K dispozici jsou plný text i prezentace. Dostupné na World Wide Web: <<http://www.inforum.cz/en/proceedings/98/>>. ISSN 1801-2213.
4. Interní informace poskytnuté firmou ProQuest
5. ŠKANDERA, Boris. Inovativní technická databáze „CSA Illustrata: Technology“ – první dojmy. Ikaros [online]. 2008, roč. 12, č. 10 [cit. 2009-01-17]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/5022>>. URN-NBN:cz-ik5022. ISSN 1212-5075.

Ing. Boris Škandera

Informetal, zájmové sdružení