

Zajišťování plných textů a faktografických údajů

ing. Rostislav Kudláček, CSc.
Informační specialista
rostislav.kudlacek@seznam.cz

INFORUM 2006: 12. konference o profesionálních informačních zdrojích
Praha, 23. - 25.5. 2006

***Abstrakt.** Jedním z primárních úkolů informačního pracovníka je zajišťování plných textů odborných dokumentů, získávání údajů z faktografických databází. Zpřístupnění záznamů o dokumentech, technické možnosti informačního prostředí, způsob publikování mu umožňují plnit toto poslání. Má dostatek podkladů pro nalezení optimálního postupu, vedle placených zdrojů se mu mnohdy nabízí neplacené, efektivně může používat vyhledávací stroje, usnadňovat získávání dokumentů i údajů pomocí agregátorů i dalších informačních služeb. Na zkušenostech se zajišťováním informací ve farmacii i zdravotnictví lze ukázat mnoho variant realizace i je porovnávat jejich efektivnost. Nesmí však spolu s koncovým uživatelem opomíjet to, že informace jsou zbožím, ve kterém vedle nákladů na publikování a zpřístupnění jsou v nich uloženy prostředky na získání i ověření jejich obsahu.*

Může být milou a příjemnou povinností knihovníka, informačního pracovníka IP zprostředkovávat koncovému uživateli plné texty a hodnověrné faktografické údaje, být tak do určité míry prospěšný v jeho odborné a vědecké činnosti. IP má dostatek podkladů k výběru zdroje informací, ze kterého lze získat dokument nebo údaje bezplatně nebo jen za úhradu. K uspokojení uživatele není možná (i když mnohdy pochopitelná) orientace jen na neplacené, protože nemohou být nikdy dostatečné, pokrýt potřeby speciálních, odborných dokumentů. (Pomineme, i když odsoudíme, nelegální přístup na př. neoprávněné využití licenčních práv¹ - poskytovatel licence pozná IP adresu uživatele, lze zkontrolovat původ plného textu a pod²). V této službě se musí IP dobře orientovat v postupu při uspokojování potřeb uživatelů v rozsahu svého pole působení volit:

- a) odebírání celých čísel a ročníků nebo jen obstarání konkrétního dokumentu;
- b) rešerše s bibliografickými výstupy a jen abstraktem nebo plné texty;
- c) tištěné – elektronické formáty;
- d) mezi možnostmi ICT zvláště internetu.

¹V tomto případě lze volit dostatečnou ochranu závazkem na úrovni pracovních povinností.

²Otázka etiky získávání informací, které je zbožím se všemi atributy, přesahuje záměr tohoto příspěvku.

K rozhodování volí pak následující kritéria – rychlost dodání, konečné náklady, frekvence používání, úplnost splnění požadavku, hodnověrnost použitých zdrojů, dostup k seznamu sledovaných, pokrytých titulů, formát dokumentu, embargované čísla (způsobující snížení aktualizace); služba jiných rešeršerů, propojení s jinými službami, které poskytují full-text; porovnání služby poskytovatelů zpracovatelů dokumentů a vlastní vyhledávání – vazba mezi dokumenty jako je derogace, novelizace, změna názvu (příkladem můžem posloužit srovnání přímého přístupu [EUROPA - Portál Evropské unie](#) a zpracování dokumentů [ESO - Web Directory](#) nebo [IDRAC.com](#) obsahující farmaceutické dokument). IP nemá vždy možnost sledovat a pamatovat si při malé frekvenci vstupu do konkrétního zdroje postupy vyhledávání, které se v něm používají, a je výhodné volit služby agregátorů nebo distributorů elektronických informačních zdrojů, využít jejich zkušenosti, větší možnosti metodického přístupu, zjednodušenou administrativu, jednotné vyhledávání ve více zdrojích.

Popis zkušeností se zajišťováním plných textů i faktografických údajů je vhodné seřadit podle nejčastějšího a mnohdy osvědčeného postupu. Z bibliografických databází, ze skryté bibliografie i z jiného zdroje je předložen informačnímu pracovníkovi název dokumentu, resp. bibliografický popis s abstraktem. To náročnému uživateli event. jeho náročné činnosti nedostačuje. Požaduje plný text. Je přirozené, že IP při potřebě zajištění plného textu nebo konkrétního údaje (chemického, farmaceutického, technického) o některé entitě se snaží pátrat mezi tituly **bezplatných časopisů**. Je třeba si uvědomit, že mezi nimi jako pravidlo nejsou časopisy s vysokým impakt faktorem a hluboce odborné, prestižní; při podrobnějším posouzení těchto titulů zjistíme, že mají nejčastěji okrajový charakter. K jejich poznání mohou posloužit seznamy časopisů: DOAJ [Directory of open access journals](#), který deklaruje, že zahrnuje jen vědecké a odborné časopisy ze širokého spektra odborností i jazyků, celkem na 2200, z nichž až v 600 lze vyhledávat i podle názvu článku (k usnadnění vyhledávání v tomto zdroji jsou k dispozici dílčí seznamy – širší [Directory of open access journals-health](#) a něm dílčí pro obor farmacie [Directory of open access journals - pharmacy](#)); [Databases@Ovid Alphabetical Listing](#) (obsahující odkazy jak na plné texty, tak na faktografické databáze); pro zdravotnické obory [Current Contents - Life Sciences](#) (jsou zde uvedeny i důležité web adresy z oboru); z oboru medicíny [BioMed Central | The Open Access Publisher](#) i s odkazem vyhledávání na PubMed; [Directory of Medical Journals - OmniMedicalSearch.com](#); [FindArticles - News, Magazines, Reference - On Health, Fitness, Business, Home, Arts, Computers](#) umožňuje vyhledávání jak plných textů volně přístupných, tak placených. [Elektronická knihovna časopisů \(EZB\)](#) (hlavní organizátor [Universitätsbibliothek Regensburg](#)) zajišťuje odkazy na celkem více než 11 300 volně

dostupných, uvedených v samostatném adresáři [Directory of Free Access Journals](#); [Scopus Info - Scopus in Detail - What does it cover?](#). Zásady zajištění volného přístupu k dokumentům jsou shrnuty [Open Archives Initiative](#).

Obsahově rozsáhlejší jsou **seznamy plnotextových časopisů**, dostupných za úplatu v rámci předplatného nebo úhradou za dodaný dokument. Tyto služby zajišťuje několik center včetně agregátorů, zde uvedme [Ovid](#), [ProQuest Information and Learning-UMI](#), [ProQuest Information and Learning](#), [HighWire Press](#), [Wiley InterScience](#) a další; spojení s nimi zajišťuje v ČR např. [AiP - Albertina icome Praha](#), [Medistyl, s.r.o.](#) Spolupráce mnoha knihoven v systému [Elektronická knihovna časopisů \(EZB\)](#) zajišťuje odkazy na celkem více než 15 000 časopisů s placeným přístupem. Na základě předplatného má NK ČR k dispozici téměř 11.000 titulů prostřednictvím [dostupných zdrojů](#). Elektronické zdroje zahrnuté do EZB jsou m.j. z produkce eIFL Direct | ProQuest | PCI Full Text | Literature Online | Science Direct | Springer Link | Electronic Journal Service. Za jednu z výhod používání služeb agregátorů lze považovat to, že odstraňují duplicitu výstupu. Především je na místě konstatovat, že v jejich souborech jsou tituly, které se nenachází mezi volně přístupnými. Z dalších rozsáhlých seznamů uvedme: [Questia - The Online Library of Books and Journals](#), který se představuje jako největší světová knihovna periodik (na 1 000 000) včetně jednotlivých článků, celých textů knih, včetně encyklopedií (přes 65 000) – obsahově je zaměřena na humanitní a sociální obory; [Scopus](#) - the world's largest abstract and indexing database - vedle bibliografických údajů a abstrakta umožňuje bezprostřední link na plné texty i dalších titulů.

Katalogy knih - odborných a vědeckých titulů - usnadňují orientaci při vyhledávání tohoto typu zdrojů. Pro informatiky i z řad svých studentů je uveden takový seznam na webu ÚISK [Digitální knihovna : seznamy digitálních knihoven](#), která je též zvlášť členěna na př. na části medicína farmacie i další obory; [Wiley InterScience](#) vedle plných textů časopisů zpřístupňuje i on-line texty knih a vzorové protokoly laboratorních manuálů.

K hledání plných textů dokumentů a mnohdy i článků slouží **katalogy webovských stránek**: [Librarians' Internet Index: http://www.lii.org/advanced](#) představuje jeden z takových katalogů, který je členěn dosti podrobně podle oborů (avšak nejen odborných a vědeckých); [Welcome to INFOMINE: Scholarly Internet Resource Collections](#); [BUBL Information Service](#); k oborovým, speciálně ve zdravotnictví, je možné řadit seznamy [Welcome to Medscape](#), [Health Medicine refdesk.com](#) – [Science Technology refdesk.com](#) (jsou součástí FACTS ENCYCLOPEDIA [Encyclopedia Resources- refdesk.com](#)). Soubory

linkovaných adres u tohoto typu zdrojů vedou na seriózní, hodnověrné zdroje, protože vznikají na akademické základně.

Vyhledávací stroje, jejich použití patří při současném rozmachu internetu k velmi frekventovaným i při zajišťování plných textů. Není možné je ani zavrhnout, odmítat, ale ani brát jako jediné prostředky k tomuto účelu. V první řadě je třeba prověřovat hodnověrnost poskytovaných informací touto cestou např. podle „pozadí“ vyhledávače – kdo jej vystavuje, ke které službě je přiřazen. Žádný vyhledávač však neovlivňuje možnost získání plného textu bezplatně. Z praxe však uvedme pozitivní poznatek – jestliže je známo úplné znění názvu práce, pak může být úspěšné pátrání, zdali jí autor ještě zveřejnil na jiném místě (např. v publikaci odborné společnosti, jejímž je členem, nebo ve zprávě své organizace apod.) a pak je pravděpodobné, že se prostřednictvím vyhledávače najde. Vědecké a odborné, volně přístupné, zaručují vedle hodnověrnosti i naději na vyhledávání s vyšší relevancí, i když nejsou určeny jen k prezentaci plných textů dokumentů. Uvedme jen jejich názvy - OmniMedicalSearch.com a jeho část [Overview - OmniMedicalSearch.com](#); [Scirus - for scientific information](#); [IngentaConnect Home](#); [Fuzzycrawler](#); [Vivísimo // Vivísimo Clustering - automatic categorization and meta-search software](#). Jejich přínos pro vyhledávání odborných a vědeckých informací spočívá i v tom, že jsou „zaštítěny“ spojením s akademickou a výzkumnou sférou. Mnohé z nich též uvádí katalogy odborných webovských adres. Přehled vyhledávačů tohoto typu je uveden na [Specialty Search Engines](#), např. medicínské [Medical Search Engines](#) nebo vědecké [Science Search Engines](#).

Nejčastěji uváděný vyhledávač obecného typu [Google](#) neustává ve své aktivitě přiblížit se uživateli. Způsob vyhledávání – bez hodnocení hodnověrnosti a libovolného umístění hledaného slova v textu - se snaží nahradit vymezením okruhu prohledávaných stránek. Příkladem je soustředění se na medicínskou problematiku: How Google is changing medicine: A medical portal is the logical next step. BMJ, 2005, vol 331, s. 331, s.1448; plný text dostupný na [How Google is changing medicine -- Giustini 331 \(7531\): 1487 -- BMJ](#); nebo weblogs [googlemedicine.pdf \(application/pdf objekt\)](#). Výhradně plným textům z titulů knih a dalších publikací se věnuje [Google Scholar](#) (s popisem [About Google Scholar](#)). Vyhledávací stroj v ruštině je např. [Яндекс](#), který obsahuje i rozsáhlý katalog web stránek (ruské buky lze psát pomocí [Yndex: The Russian keyboard emulator](#)). Obecné vyhledávací stroje jsou různě kombinované, agregované, označované metasearch apod: [Info.com - Displays results from 14 search engines](#) (se speciální částí [Info.com - Displays results from 14 search engines Health](#)); [Dogpile Web Search Home Page](#). Přehled vyhledávacích strojů i další informace o nich lze nalézt na [Search Engine Watch: Tips About Internet Search Engines &](#)

[Search Engine Submission](#) ; na těchto stránkách je možné nalézt též seznamy vyhledávacích strojů podle oborů [Specialty Search Engines](#).

Srovnání některých vyhledávacích strojů pro nalezení dokumentů s určitým obsahem je v příloze.

U **faktografických databází**, na rozdíl od plnotextových, je důvod pro rozdílný způsob poskytování zřejmější – bezplatný může být dán povinností orgánu zveřejnit údaje, nebo vyplývají ze zřetelného komerčního záměru.

Pro výchozí orientaci je vhodné použít katalogy faktografických databází: (All Topics >>[CompletePlanet - Discover over 70,000+ databases and specially search engines](#) a jeho části [All Topics](#) >> Health [CompletePlanet - Discover over 70,000+ databases and specially search engines](#) (1059 titulů) [All Topics](#) >> [Health](#) >> Drugs & Pharmaceutical: [CompletePlanet - Discover over 70,000+ databases and specially search engines](#) (68 titulů). Mezi volně přístupné databáze z medicíny, farmacie, event. chemie patří [Knovel Sample Book](#), který poskytuje identifikační a fyzikálně-chemické údaje chemických látek i tabulky z inženýrství a fyziky (včetně převodu jednotek); [FIZ Chemie Guides](#) – má též části MedPharmGuide (9009021 web pages) a ChemGuide (8806810 web pages); údaje v plném textu o nebezpečných chemických látkách z 11 zdrojů jsou přístupny bezplatně na [IPCS INCHEM](#).

V případech, kdy faktografická databáze, údaje v nich uvedené, jsou výsledkem poměrně pracné excerpcce a zpracování faktů z placených zdrojů, nelze předpokládat volný přístup do těchto databází. Jako příklad uveďme zjištění historie léčiva (od jeho výzkumu přes klinické zkoušky, registraci, uvedení na jednotlivé trhy a postmarketingové sledování). Tyto údaje jsou obsaženy jen v následujících titulech s placeným přístupem - IMSworld R&D Focus, Pharmaproject a Adis R&D Insight. A do databází tohoto typu nemá přístup žádný bezplatně provozovaný vyhledávač. Vznik faktografických databází volně přístupných je zřejmě dán komerčními důvody (vedoucí zpravidla k „demo“ ukázce omezeného počtu entit), legislativní povinnosti (poskytující pouze aktuální stav s přesně vymezenou strukturou monografie, která je pro odborníky a specifické požadavky nedostatečná), ne jen prostým altruizmem.

Placené dokumenty budou existovat nadále - je v nich uložena práce, přidaná hodnota. Bezplatné poskytování informací nebo za úhradu bude zřejmě záležet vždy a jen na producentovi. Informace jsou zbožím – není zvláštní důvod, proč by se v úhradě nebo bezplatném přístupu příliš odlišovali od jiného zboží. Placený nebo volný přístup k informacím velmi málo, prakticky vůbec nezávisí a ani nemůže záviset na informačních

pracovnících nebo knihovnách. Jejich role je poměrně výrazně vymezena v procesu akvizice, zpracování a ohlašování, je nezastupitelná (z hlediska zkušenosti, frekvence hledání, orientace ve zdrojích ve srovnání s koncovým uživatelem), ale nespočívá v placení (pokud mu nebudou tyto prostředky předány tak, aby nebyla omezena jeho základní činnost). Pozastavme se nad názorem – „knihovny institucí budou pravděpodobně muset asistovat autorům v procesu publikování a platit poplatky za přijetí rukopisu k publikování“. Není zřejmý důvod, proč by se tak mělo stát. Proč by měla knihovna platit za opublikování práce odborného, vědeckého pracovníka její organizace – čili vynakládat prostředky za neznámého konečného uživatele; za koho by měla platit knihovna veřejná, která není spojena přímo s producenty dokumentů. Jde více o politickou a ekonomickou záležitost – zda zboží a služby se platí nebo zda jsou poskytovány solidárně bezplatně. O tom, že tomu tak je, svědčí to, že Open Access je předmětem jednání Světové obchodní organizace WTO³. Též je vhodné připojit, že nelze vyvozovat z nároku na volný přístup k informacím požadavek na jejich bezplatné poskytování a není možné rozšiřovat povinnost poskytování informací podle zákona č. 106/1999Sb. povinnými subjekty (které mají podle tohoto zákona v rozsahu své působnosti s poskytováním jiných informací (z jimi pořizovaných placených, licencovaných zdrojů a pod.).

Dělba procesů a zajišťování informací informačním pracovníkem se nebude příliš měnit – orientace ve zdrojích, vypracování možností dostupnosti, sledování novinek, metodika získávání dokumentů a údajů a zároveň se nebude měnit podstata vědecké, odborné, projekční práce – work finish publish. A není důvod, aby se měnila práce knihovníka - informačního pracovníka zásadním způsobem, proč by se měl především ekonomicky zapojovat do procesu tvorby dokumentu. Má dostatek odpovědné práce s jeho zpracováním, zveřejněním, uskladněním i poskytováním.

Příloha:

Autor si dovoluje jako ilustraci možností některých vyhledávacích strojů i jejich srovnání a event. hodnocení nostalgicky vycházet ze svého původního vzdělání a spojit s činnostmi na informačním poli konkrétní odbornosti. Zadání vychází z formulace booleanu „Stable

³ Viz na př. Economic Perspectives on a Multilateral Agreement on Open Access to Basic Science and Technology by *John H. Barton, Keith E. Maskus* [SCRIPT-ed: Online Journal](#) - což je jeden z podkladových materiálů pro tuto organizaci.

isotope“ x pharmaceutical x (authentication or counterfeiting)⁴(v uvozovkách označena fráze). Hledalo se, zda-li se získá několik apriori známých publikací a počet hitů relevantních požadavku.

Searchmachine	bere výraz	celkem hitů	relevantní	známé	poznámka
OmniMedicalSearch.com	ano	41	1	0	a)
Vivísimo	ano	69	2	ano	b)
Google	ano	32; 441	4	ano	c)
Google Scholar	ano	3; 32	ano	ano	d)
Info.com	ano	14	5	ano	e)
Dogpile	ano	1; 69	ano	ne; ano	f)

a) bere i jen jednotlivé výrazy, nectí celý boolean, nerespektuje frázi;

b) neumí frázi, 69 hitů označeno jako „top“, seřazeny do clustrs, celkem se uvádí výsledek 615, bere i jednotlivé výrazy;

c) nerespektuje znaky v boolean výrazu (x jako znak, ale jako písmeno textu) při jeho celém zadání; druhý výsledek je získán s využitím advanced search;

d) první výsledek boolean, nerespektuje frázi, nehledá samotné písmeno x; druhý výsledek s využitím advanced search;

e) bere boolean;

f) bere boolean, druhý výsledek v advanced.

Uvedené zadání rešerše, poskytlo ve vybraných vyhledávacích poměrně rozdílné výsledky. Při větším počtu hitů na výstupu je někdy složité a pracné vybrat vhodné, relevantní plné texty. Výraznější pozitivnější výsledky se získaly v posledních třech hledáních. Provedené srovnání platí pro ryze konkrétní zadání. I když výsledky lze považovat spíše za orientační, ukazuje se, že je vhodné udělat si podobné srovnání, nepracovat jen s jedním vyhledávacím strojem, vypracovat si svůj metodický postup.

⁴ Metoda, založená na stanovení koncentrace stabilních izotopů se používá v kriminalistice též ke zjišťování falz vína, potravin, výbušnin nebostanovení identity šarší výrobků - stanovuje se odlišné místo původu výrobku ve srovnání s chráněným originálem; má též své místo v boji s terorismem.