

Otevřený přístup, digitální knihovny a citační služby

Eva Bratková

Ústav informačních studií a knihovnictví FF UK v Praze

Eva.Bratkova@ff.cuni.cz

1 Úvod

Narůstající fondy vědeckých informačních materiálů, které jsou dostupné zdarma v prostoru WWW, jsou předmětem zájmu nejen koncových uživatelů, kteří je stále více vyhledávají a využívají ve své práci, ale také předmětem značného zájmu informačních profesionálů a řady informačních služeb, které tyto materiály zpracovávají a dále zpřístupňují ve formě zajímavých výstupů. Ke službám velkého významu v oblasti vědy patří služby citační, které, kromě jiného umožňují sledovat vliv vědeckých prací na další vývoj vědeckého výzkumu.

Následující příspěvek v první části představuje problematiku informačních materiálů, které jsou šířeny v prostředí WWW zdarma. Definován je termín „otevřený přístup“, představeny jsou hlavní typy zdrojů s volně dostupnými informačními materiály a také instituce a iniciativy, které otevřený přístup k informacím podporují. Hlavní část příspěvku je věnována stručnému rozboru a hodnocení významných vlastností hlavních citačních služeb, které zpracovávají volně dostupné materiály uložené ve volně dostupných digitálních knihovnách nebo elektronických archivech a také a citace v nich obsažené. Příspěvek vychází z dostupné literatury, zejména pak ze zprávy britských odborníků z roku 2005 [HARDY, 2005], aktuální informace byly získány prostřednictvím online rozhraní jednotlivých citačních systémů. Zařazeny jsou jak systémy komerční (WOS, Web Citation Index, Scopus a CrossRef), tak především systémy volně dostupné (Citebase Search, CiteSeer, Google Scholar). Závěrečná část příspěvku zahrnuje stručné informace o problematice citačního propojování a strukturace záznamů citací pro potřeby jejich strojového zpracování podle standardu OpenURL a doporučení „Směrnice pro kódování bibliografických citačních informací“ Metadatové iniciativy Dublin Core.

2 Otevřený přístup (Open Access)

2.1 Definice otevřeného přístupu

Komunikace informací ve vědě doznala v průběhu posledních 15 let velkých změn. Výrazným trendem dneška je další rozšiřování otevřeného přístupu k informacím. Termín „**otevřený přístup**“ (Open Access) v sobě zahrnuje význam **trvalého a bezplatného online přístupu k dokumentům, zejména úplným textům, pro všechny uživatele**. Otevřený přístup se vyznačuje dvěma podstatnými rysy: dokumenty jsou **dostupné bezplatně**, i když nemusejí být bezplatně vytvářeny, a vlastník copyrightu dává jednoznačné svolení k jejich neomezenému **čtení, stahování, kopírování, sdílení, ukládání, tištění, vyhledávání a hypertextovému propojování** [HARDY, 2005]. Lze doplnit, že otevřený přístup je v rámci tohoto příspěvku míněn vždy v kontextu bezplatného zpřístupňování informací v prostředí WWW. Termín „otevřený přístup“ lze tedy ztotožnit s termínem „**bezplatný online přístup**“. V různých oborech lze bezplatně poskytovat různé typy dokumentů (preprinty, články z recenzovaných časopisů, technické zprávy, konferenční sborníky aj.).

Prvním významným reprezentantem bezplatného online přístupu k informacím ve vědě byl mezinárodní archiv elektronických preprintů vědeckých recenzovaných článků **arXiv.org** (<http://arXiv.org/>), založený komunitou fyziků již v roce 1991 (nyní je provozován na Cornellově univerzitě). Vznik WWW na počátku 90. let 20. stol. a jeho rychlé rozšiřování po

celém světě vedlo řadu autorů i z dalších oborů k vystavování kopií publikovaných vědeckých článků volně (neorganizovaně) na jejich **personálních webových stránkách**, lokalizovaných zpravidla na univerzitních serverech. To nebyl optimální jev, proto začaly na přelomu tisíciletí podle modelu archivu arXiv.org postupně vznikat institucionální repozitáře elektronických tisků, do kterých autoři mohli pomocí **auto-archivace** svoje dokumenty uložit a popsat. Velký impuls k tomu dala **Iniciativa otevřených archivů** (Open Access Initiative, OAI, <http://www.openarchives.org/>), která se zrodila v roce 1999 právě za účelem integrace lokálních repozitářů prostřednictvím sklizení jejich metadat na základě protokolu **OAI-PMH** (Protocol for Metadata Harvesting) [LAGOZE, 2004]. Koncepce OAI však počítala se všemi typy dokumentů, které se mohou v digitálních knihovnách vyskytovat. Lze konstatovat, že se tak z koncepce OAI do značné míry vytratil původní cíl bezplatného online přístupu jen k recenzovaným časopisům [HARDY, 2005].

Tato myšlenka opět ožila a byla formalizována jako „otevřený přístup“ v rámci **Budapešťské iniciativy otevřeného přístupu** (Budapest Open Access Initiative, BOAI, <http://www.soros.org/openaccess/read.shtml>). Zdůrazněny byly dva způsoby otevřeného přístupu: uložit preprint článku do univerzitního nebo tematického repozitáře a pokusit se potom předložit tento text buď do komerčního nebo otevřeného online časopisu.

2.2 Zdroje informací s otevřeným přístupem

K nejdůležitějším typům informačních zdrojů s otevřeným přístupem patří:

- a) předmětově profilované archivy a institucionální repozitáře
- b) online časopisy s otevřeným přístupem
- c) povrchový web

Již výše zmíněný archiv **arXiv.org** je reprezentantem typu centralizovaných **předmětově profilovaných archivů** s mezinárodním záběrem dokumentů. Zahrnuje preprinty, postprinty vědeckých článků, ale také jinou šedou literaturu z oborů fyziky, matematiky, počítačové vědy, nelineárních věd a kvantitativní biologie. Jde o jeden z nejstarších a také největších archivů, jeho repozitář obsahuje již téměř 400 000 dokumentů. Takových archivů je v současnosti řádově několik desítek, a to v oblasti přírodních, ale také společenských nebo humanitních oborů. Jmenovat lze například archiv postprintů recenzovaných **článků PubMed Central** pro biomedicínské obory, archiv **CogPrints** pro kognitivní vědy, **E-LIS** pro knihovní a informační vědu aj. Jejich provozy zajišťují zpravidla určité univerzitní nebo vědeckovýzkumné organizace z nějaké země. Mohou využívat i případnou finanční podporu dalších korporací včetně mezinárodních.

Druhým typem archivů jsou **institucionální repozitáře** (Institutional repositories, **IR**), které začaly vznikat po roce 2000 na univerzitách po celém světě. Jak dokládá jejich světový registr **ROAR** (Registry of Open Access Repositories, <http://archives.eprints.org/>), k dubnu 2006 bylo registrováno cca 676 IR. K největším IR patří **Archiv Evropského sdružení pro jaderný výzkum** s téměř 850 000 záznamy a cca 360 000 elektronickými dokumenty (CERN Document Server, <http://cdsweb.cern.ch>) a Archiv **ECS** s více než 10 000 dokumenty (ECS Eprints Service, <http://eprints.ecs.soton.ac.uk>), který spravuje Fakulta elektroniky a počítačové vědy Southamptonské univerzity. Hlavním úkolem IR je zajistit otevřený přístup k recenzovaným výsledkům výzkumu dané instituce a jejich autorů, zároveň tím ale také zajistit větší vliv jejich prací a v důsledku toho zvětšit také viditelnost a prestiž dané instituce. S nasazováním nových informačních technologií se objevují nové **externí citační služby**, jež pomáhají vliv a prestiž zvyšovat. Objevují se studie, které dokladují tezi, že dokumenty dostupné v otevřeném přístupu získávají značný vliv (impakt) ve srovnání s dokumenty dostupnými komerční cestou [HARNAD, 2004]. Informační vstupy do archivů/repozitářů

jsou zpravidla zajišťovány pomocí **auto-archivace**. Návštěvnost největších mezinárodních archivů koncovými uživateli (vědci) z celého světa je enormní (desítky tisíc spojení denně).

Elektronické archivy mohou být z hlediska programového zabezpečení provozovány buď v komerčních nebo volně dostupných softwarech. K nejrozšířenějším volně dostupným patří americký program **DSpace** (<http://www.dspace.org/>), vyvinutý na Technologickém institutu v Massachusetts (MIT), a britský program **Eprints.org** (<http://software.eprints.org/>), který je rozvíjen na Southamptonské univerzitě. Oba programy jsou již připraveny na sklízení metadat podle protokolu OAI-PMH libovolnými informačními službami včetně. Významnou službou, která indexuje metadata z volně dostupných archivů a zajišťuje efektivní vyhledávání akademických dokumentů, je služba **OAIster** (<http://oaister.umdl.umich.edu/o/oaister/>). Zajímavou službou je i bezplatný vyhledávací systém **Scirus** (<http://www.scirus.com/srsapp/>) komerčního vydavatele Elsevier, který kromě dat ze svých publikací zpracovává také data z volně dostupných institucionálních repozitářů, nikoliv ale ze všech.

Volně dostupné online časopisy, které mohou zachovávat tradiční charakteristiky klasických časopisů, jsou určeny k bezplatnému využívání koncovými uživateli. Jejich významným rysem zpravidla bývá budování archivu či repozitáře k ukládání všech publikovaných článků. Někteří vydavatelé zajišťují volný přístup ke všem článkům (například vydavatel časopisů Public Library of Science, <http://www.plos.org/>). Jejich provoz je finančně zajišťován sponzory, případně i internetovou reklamou. Někteří nakladatelé praktikují zpřístupňování nových článků až po uplynutí určité doby (6 či více měsíců), starší články z archivu bývají dostupné bezplatně. Komerční nakladatelé mohou zveřejňovat bezplatně články z recenzovaných časopisů s tím, že náklady hradí autor nebo instituce za více autorů (například BioMed Central, <http://www.biomedcentral.com>, který publikuje asi 100 online časopisů s otevřeným přístupem). Mnohé online časopisy jsou připraveny na sklizeň metadat na základě protokolu OAI-PMH, takže se dostávají, stejně jako je tomu u elektronických archivů/repozitářů, ke zpracování v nadstavbových bibliografických a citačních službách.

Různé registry volně dostupných časopisů vykazovaly v současné době kolem 2 200 titulů (převážná část z nich je z oborů přírodních věd), což je zatím malé číslo ve srovnání s přibližně 24 000 tituly recenzovaných komerčních časopisů [HARDY, 2005]. Registraci online volně dostupných časopisů zajišťuje například systém **DOAJ** (Directory of Open Access Journals, <http://www.doaj.org/>), který spravuje Systém knihoven Univerzity v Lundu (Švédsko). V rámci registru je k dispozici také vyhledávání na úrovni článků z 633 online časopisů (systém registruje celkem 2 234 online titulů).

Elektronické kopie vědeckých článků z recenzovaných časopisů nebo i libovolné jiné odborné texty řada autorů vystavuje na svých **personálních webových stránkách**. Pro koncové uživatele se stávají vyhledatelnými prostřednictvím univerzálních webových vyhledávacích služeb nebo prostřednictvím specifických citačních služeb, jako jsou **Google Scholar** (viz část 3.2.4) nebo **CiteSeer** (viz část 3.2.5).

2.3 Podpora otevřeného přístupu k informacím a jeho růst

V rozšiřování otevřeného přístupu k dokumentům hrají klíčovou roli **autoři**. Ačkoliv velké množství autorů ví o existenci volně dostupných elektronických archivů a využívá informace v nich uložené, jen málo z nich do nich aktivně přispívá, jak dokladují již realizované průzkumy [HARDY, 2005]. Velké množství autorů často využívá dokumenty, které jsou lokalizovány volně na personálních webových stránkách jejich autorů - jsou jednoduše nalézány při vyhledávání pomocí internetových vyhledávačů (Google aj.). Příznivější situace je v oblasti přírodních věd. Proces auto-archivace dokumentů v archivech či repozitářích si jen pozvolně razí cestu kupředu, nicméně vůle realizovat tento proces je u autorů průzkumy

doložena. Počet autorů přispívajících do archivů se každým rokem zvětšuje. Autoři, kteří využívají auto-archivaci v otevřených archivech, oceňují jednak možnost bezplatného využívání jejich dokumentů koncovými uživateli, poskytující možnost zvyšování citovanosti jejich prací, jednak rychlost publikování. Průzkumy realizované v zahraničí také potvrzují, že velké procento autorů o možnosti auto-archivace svých prací vůbec neví. V tomto směru je pro další období důležité rozšiřování části informační gramotnosti, zejména na univerzitách.

Pro autory je důležitý vliv jejich práce. Pokud autor zjistí zlepšení **vlivu** své práce díky bezplatnému přístupu k ní, bude pravděpodobně více ochotný využívat této cesty. Mnozí odborníci uvádějí, že v některých oborech (například počítačové vědě) byla citovanost preprintů recenzovaných článků, které byly uloženy v otevřených archivech, větší než citovanost článků získaných z předplacených zdrojů. V některých průzkumech byla zjištěna vyšší citovanost otevřeně přístupných dokumentů i ve srovnání s citovaností uvedenou ve světové citační databázi WOS společnosti Thomson [HARDY, 2005].

Zvětšování počtu otevřeně přístupných zdrojů a zároveň zvětšování jejich využívání koncovými uživateli se neobejde bez **podpory řady institucí na lokální, národní i mezinárodní úrovni**. Proto ve vyspělých zemích poskytují různé výzkumné instituce a nadace nebo státní orgány mnoho finančních prostředků na rozvoj aktivit v této oblasti. Značné prostředky byly již vynaloženy ve Velké Británii (například institucí The Wellcome Trust nebo The Research Councils UK), v USA (The National Institutes of Health aj.), Finsku (Ministerstvo výchovy) aj. Velikou podporu otevřenému přístupu k informačním zdrojům nabízejí mezinárodní organizace. Jmenujme zejména Institut otevřené společnosti (Open Society Institute), který v roce 2002 vyhlásil **Budapešťskou iniciativu otevřeného přístupu** (BOAI). Tato iniciativa, mimo jiné, zdůraznila potřebu auto-archivace elektronických článků do elektronických archivů a zavádění volných online časopisů. V rámci této iniciativy byly mobilizovány finanční zdroje na pomoc vzniku nových online časopisů s otevřeným přístupem a na změnu obchodních modelů u již existujících komerčních časopisů ve prospěch auto-archivace elektronických článků. Principy a doporučení v oblasti elektronického publikování s otevřeným přístupem v biomedicínských oborech jsou kodifikovány také v **Prohlášení z Bethesdy o publikování s otevřeným přístupem** (Bethesda Statement on Open Access Publishing, <http://www.earlham.edu/~peters/fos/bethesda.htm>) z roku 2003. Ve stejném roce byla v říjnu přijata také **Berlínská deklarace o otevřeném přístup ke znalostem ve vědě a humanitních oborech** (Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>), která byla podepsána hlavními německými vědeckými institucemi. Berlínská deklarace se stala impulzem k přijetí stejných principů k podpoře otevřeného přístupu v institucích dalších zemí.

Také řada **komerčních vydavatelů** se rozhodla v určité míře podporovat otevřený přístup k informačním zdrojům. Jednou z možných cest je udělit autorům svolení k auto-archivaci vlastních kopií publikovaných dokumentů (preprintů nebo postprintů). Registrací nakladatelů, kteří přijali politiku podpory auto-archivace, se zabývá projekt ROMEO. Jeho databáze je dostupná ze serveru projektu **SHERPA** (<http://www.sherpa.ac.uk/romeo.php>). Doplnková databáze poskytuje také informace o jednotlivých titulech časopisů, v rámci kterých je možné politiku auto-archivace realizovat (tzv. „**zelená politika**“). V rámci adaptace „zelené politiky“ dnes velký počet komerčních časopisů (vydavatelů jako jsou Elsevier, Springer aj.) dává automatické svolení k auto-archivaci - žádost autora se již nemusí předkládat oproti dřívějšímu období [SUBER, 2006]. Někteří nakladatelé přijali hybridní politiku podpory veřejného přístupu k publikovaným dokumentům. V rámci ní požadují od autorů zaplacení poplatku za možnost zpřístupnění dokumentu ve zdroji s otevřeným přístupem. Jde o jeden z experimentálních modelů pomoci autorům zpřístupnit dokument v otevřeném zdroji.

3 Citační indexace a citační služby

3.1 Citační indexace její význam

Citační indexací rozumíme proces vytváření rejstříku citací. **Citační rejstřík** je databází, která spojuje citující dokumenty s dokumenty citovanými prostřednictvím jejich záznamů. Zatímco seznam záznamů použité (citované) literatury v nějakém dokumentu míří vždy jen k **dříve publikovaným dokumentům**, specifikem citačního rejstříku je, že může poskytnout také seznamy **později publikovaných dokumentů**, které daný dokument citovaly (vazba typu „cited by“). Citační rejstříky jsou významnými nástroji citační analýzy, v rámci které jsou sledována citovaná díla za účelem určení vlivu těchto děl nebo jejich autorů. Citace mohou být využity k měření **vlivu** nějakého dokumentu ve specifickém oboru. Čím více je čten a citován nějaký článek, tím více může ovlivňovat badatele v tomto oboru. Na úrovni citačních služeb je měřen tzv. „**impakt faktor**“ (faktor vlivu) vědeckých časopisů.

Se zdokonalováním informačních technologií se citační rejstříky staly sofistikovanějšími. Oproti dřívějšímu období vstoupily do hry **úplné texty dokumentů** v dobře **organizovaných digitálních knihovnách nebo archivech** s předplaceným nebo otevřeným přístupem. Takové prostředí umožňuje poskytování velkého počtu nových zajímavých a užitečných výstupů a služeb. Jsou-li připravena kvalitní hypertextová propojení, může uživatel získávat úplné texty jak citujících tak citovaných dokumentů. Zdroje s otevřeným přístupem (archivey a volné online časopisy), komentované v předchozí kapitole, jsou zahrnovány dnes i do komerčních citačních služeb, které se tak stávají ještě více komplexnějšími. S nástupem nových technologií skončila zároveň dlouhá doba monopolního postavení zatím největšího a nejkvalitnějšího světového citačního rejstříku na světě provozovaného dnes v systému **Web of Science (WOS)**, který je součástí komplexního portálu **Web of Knowledge**. Objevily se a objevují i nadále další nové moderně koncipované citační služby, které se na nekomerční bázi věnují informačním materiálům s otevřeným přístupem.

3.2 Současné citační služby

V dalším přehledu jsou uvedeny současné nejvýznamnější citační služby. Zahrnuty jsou jak služby komerční, které zpracovávají vybrané množiny vědeckých časopisů včetně online volně dostupných materiálů (patří k nim systém WOS, Scopus a CrossRef), tak bezplatně dostupné služby, které zpracovávají především dokumenty dostupné volně v prostoru webu (patří k nim systém Google Scholar, CiteSeer, Citebase Search a také nová služba Web Citation Index, která ale bude dostupná komerčně).

3.2.1 Web of Science

WOS je polyoborový citační systém zahrnující následující základní citační rejstříky: **Science Citation Index (SCI, 1900-současnost)**, **Social Sciences Citation Index (SSCI, 1956-současnost)** a **Arts & Humanities Citation Index (A&HCI, 1975-současnost)**. Do systému WOS patří ještě dvě chemické databáze, které ale zatím vykazují malý počet záznamů (cca 100 000): **Chemical Reactions Index** a **Index Chemicus**. Největší počet záznamů je obsažen v části SCI (až 77 %), část SSCI obsahuje zhruba 14 % a část A&HCI zhruba 9 % záznamů. Databázi citačních rejstříků dnes produkuje a přes webové rozhraní také zpřístupňuje společnost The Thomson Corporation (<http://isiwebofknowledge.com/index.html>). Celek databáze zahrnuje již přibližně **37 000 000** záznamů zdrojových (citujících) článků. Abstrakta se vyskytují u záznamů od roku 1971 (u části SCI) a od roku 1972 (u části SSCI). Významným rysem databáze WOS ve srovnání s novými citačními systémy je fakt, že **citace** byly předmětem zpracování od samého počátku existence tohoto zdroje, což je jeho velkou

výhodou. Přesný počet zaznamenaných citací (záznamů citovaných dokumentů) v bázi WOS není znám, ale někteří odborníci odhadují, že jich je zhruba **500 000 000** [JACSÓ, 2004b]. Excerptní základnu tvoří v současnosti kolem **9 000** vybraných vědeckých časopisů především komerčního charakteru, do tohoto počtu patří již také zhruba **230 online časopisů s otevřeným přístupem**. Celkový počet excerptovaných časopisů za celé období existence tohoto zdroje je zhruba **15 000** [JACSÓ, 2005a, s. 1541]. Producent již také rozšiřuje excerptní základnu o vybrané periodické konferenční sborníky a edice monografií.

Producent ustanovil a přísně dodržuje **standards pro jednotný popis zdrojových článků** a zejména **jednotný formalizovaný popis citovaných dokumentů**, který je zajišťován pomocí velkého vkladu lidské práce. Různé formy citací uplatňované autory z celého světa musejí být v systému WOS zapsány v unifikovaném tvaru (jde o citace různých typů dokumentů, nejen o článků). K zabezpečení správného propojení záznamů citovaných dokumentů se záznamy příslušných zdrojových dokumentů, jsou-li v bázi WOS uloženy, je využito algoritmicky vytvářených specifických klíčů, které tyto dokumenty reprezentují.

Systém WOS umožňuje sofistikované **online vyhledávání** a další využívání záznamů citujících (zdrojových) článků i citovaných dokumentů (citací). Kromě jiného nabízí vyhledávání podle **tématu** (Topic) na základě klíčových slov z názvů článků, autorských abstraktů a autorských klíčových slov. Výsledné zkrácené záznamy zdrojových článků lze uspořádat podle data publikování, citovanosti (záznamy nejcitovanějších prací jsou na první pozici), relevance, prvního autora a názvu zdroje. Výsledné záznamy (maximálně 100 000) lze podrobit **bibliometrické analýze** (Analyze Results). U záznamů zdrojových článků lze prohlížet **citace dokumentů jiných autorů** (Cited References) a lze také využívat jedinečné funkce k zobrazení **záznamů článků, které výchozí dokument již v daném okamžiku citovaly** (Times Cited), tedy zjišťování vlivu dané práce na jiné autory. Záznamy citovaných dokumentů, které jsou v databázi zároveň reprezentovány záznamem jako zdrojové články, jsou s nimi hypertextově propojené. Lze také zjišťovat **předmětově relevantní články** (Find Related Records), nastavit službu typu „**alerts**“ aj. Systém WOS nabízí hypertextové odkazy na **úplné texty článků** (View Full Text), jejichž využívání je však vázáno na další licenci přístupu do digitálních knihoven vydavatelů nebo přístupu k dodavatelským službám.

WOS je odborníky vyhodnocován jako **nejlepší citační služba** současnosti, která propojuje miliony vědeckých prací publikovaných od 60. let 20. století, resp. od roku 1945 nebo 1900, pokud jde o přírodní a aplikované obory. WOS je tedy, co do retrospektivy, rozsáhlejší databází než databáze služby Scopus, jak ve své recenzi poznamenal P. Jacsó [JACSÓ, 2004b]. Jako nejstarší citační služba je využívána velkým počtem institucí z celého světa. Náměty či podněty odborníků na další vylepšování rešeršního systému realizuje producent postupně v každé další nové verzi – viz inovace v roce 2005-2006 [ESPECHE, 2006.]. Nový směr v systému WOS reprezentuje připravovaný nový rejstřík **Web Citation Index**, jehož předmětem jsou volně dostupné materiály a jejich citace (viz část 3.2.7).

3.2.2 Scopus

Scopus je rozsáhlou online dostupnou polyoborovou bibliografickou a abstraktovou databází s přidanou hodnotou. Je obohacena o **citační rejstřík a citační vyhledávání**. Jde o novou komerčně založenou službu, která byla do provozu uvedena v listopadu 2004 vydavatelskou skupinou Reed Elsevier. Scopus je novým konkurentem citační služby WOS, a byl proto podroben již mnoha hodnocením v odborném tisku. Scopus vznikl na jiném základě než systém WOS. Byl vytvořen pomocí **extrakce záznamů** z tradičních abstraktových databází tohoto vydavatele (například GEOBASE, BIOBASE, EMBASE) i vydavatelů dalších **s obohacením o záznamy citací z článků z časopisů**. Citace ale jsou, jak uvádějí analytici,

doplněny zatím jen u záznamů dokumentů publikovaných po roce 1995 [JACSÓ, 2005a, s. 1539]. Z tohoto hlediska tedy Scopus zatím databázi WOS určitě nepřehání. Scopus ale v současné době pokrývá již zhruba **15 000** komerčních časopisů a konferenčních sborníků z produkce přibližně **4 000** vydavatelů z oblasti vědy, techniky a medicíny a jeho retrospektiva sahá k roku 1966. Z oblasti inženýrských oborů je zahrnuto také zhruba **20 000 knih**. Scopus nezahrnuje vůbec zdroje z humanitních disciplín, zdrojů z oblasti sociálních věd je minimální množství. Vykazovaný objem záznamů zdrojových článků s abstrakty je **27 500 000** záznamů (k dubnu 2006). Tyto parametry signalizují, že Scopus by mohl být větším zdrojem než WOS. Recenzenti však doložili, že počet abstraktů je ve skutečnosti nižší, jsou zahrnuty jen u 67 % záznamů [JACSÓ, 2005a, s. 1539]. Předmětem zpracování databáze Scopus je dále zhruba **400 online časopisů** s otevřeným přístupem. To je pozitivní fakt, jde o větší počet než u systému WOS. Scopus také zpracovává vědecké dokumenty dostupné volně na WWW, které jsou získávány pomocí jeho vyhledávače **Scirus** [HARDY, 2005]. Jejich záznamy jsou zahrnovány do výsledných rešerší. Počet zaznamenaných citací v databázi Scopus je podle oficiální dokumentace zhruba **230 000 000**. Jsou získávány algoritmicky extrakcí z úplných textů.

SCOPUS Register | Login

Search Sources My Alerts My List My Profile Help Scopus Labs

Quick Search [] Go Search Tips Brought to you by Univerzita Karlova Souborný katalog UK

Citation Overview Citations received since 1996

Author: Brody, T export print

Sort documents Date Range
 year descending 2004 to >2006 Go

54 Cited Documents save to list	Total	Citations					total	
		<2004	2004	2005	2006	>2006		subtotal
1 <input type="checkbox"/> 2005 Regulation of temporal identities d...					1		1	
2 <input type="checkbox"/> 2005 EVOPRINTER, a multigenomic comparat...					1		1	
3 <input type="checkbox"/> 2005 Nerfin-1 is required for early axon...						0	0	
4 <input checked="" type="checkbox"/> 2004 The access/impact problem and the g...				2	1		3	
5 <input checked="" type="checkbox"/> 2004 Comparing the impact of Open Access...			2	12	4		18	
6 <input checked="" type="checkbox"/> 2002 Open citation linking. The way forw...	2		2				4	
7 <input type="checkbox"/> 2002 Cellular diversity in the developin...	4	2	10				16	
8 <input checked="" type="checkbox"/> 2002 Evidence of hypertext in the schola...	1	3					4	
9 <input type="checkbox"/> 2002 Sale of your product can prevent pa...						0	0	
10 <input type="checkbox"/> 2002 Identification of novel Drosophila ...	1	7	4	1			13	
11 <input type="checkbox"/> 2001 A search for Drosophila neural prec...	2	1					3	
12 <input type="checkbox"/> 2000 Nerfin-1 and -2, novel Drosophila Z...	4	1	2				7	
13 <input type="checkbox"/> 2000 Programmed transformations in neuro...	16	9	9	2			36	
14 <input type="checkbox"/> 2000 Drosophila melanogaster G protein-c...	31	6	11	4			52	
15 <input type="checkbox"/> 2000 Comparative genomics of the eukaryo...	528	80	65	11			684	
		1138	151	141	35	0	327	1465

Obr. č. 1: přehled počtu citací na díla více autorů s identickým jménem (T. Brody) za poslední 3 roky (cílem bylo najít práce počítačového odborníka s tímto jménem - viz označené záznamy) [2006-04-28]

System Scopus je zpřístupňován na základě předplatného, které se liší podle charakteru a velikosti instituce. Rešeršní systém nabízí standardní způsoby online vyhledávání záznamů zdrojových článků (**základní, autorské a pokročilé**). Výsledné záznamy mohou být uspořádány podle data, relevance, autora, názvu zdrojového časopisu a počtu citací (Cited by). V rámci záznamů jsou zobrazovány nejdůležitější údaje, jako je datum vydání, název

článku, autoři, název zdrojového časopisu a citovanost. Scopus dále uživateli umožňuje prohlížet záznamy citovaných dokumentů, zobrazovat citace jednotlivých dokumentů, nastavit užitečnou specifickou citační službu typu „alerts“, v jejímž rámci bude uživatel získávat také záznamy nových článků, které jež citují určitý dokument. U **citací** (References) lze prohledávat název citované práce a všechna jména citovaných autorů. V rámci autorské rešerše systém nabízí specifickou funkci vyhledání a zobrazení přehledu **citovanosti prací zvoleného autora** (Citation tracker). Scopus má stejný problém při vyhledávání jmen autorů jako systém WOS - užívání iniciál křestních jmen vede k nalezení záznamů prací různých autorů (s identickým jménem - viz obr. č. 1). Podobně jako v systému WOS je i v systému Scopus možné vyhledávat **záznamy příbuzných dokumentů** (Related Documents) kalkulované na základě sdílených citací.

Systém je připraven na propojování informací na základě standardu **OpenURL**. Hypertextové odkazy se u záznamů objevují, pokud má daná instituce předplatné k úplným textům časopisů. Systém také zahrnuje hyperlinky na bezplatně dostupné informační zdroje, a to jak pokud jde o zdrojové články, tak citované dokumenty.

Systém Scopus je zajímavým a perspektivním informačním zdrojem s připojenou citační službou. V mnoha aspektech je srovnatelný se systémem WOS. Jde o mladý systém, jehož databáze dynamicky narůstá. Také rešeršní systém je předmětem neustálého zdokonalování.

3.2.3 CrossRef

Komerčně dostupný systém **CrossRef** (<http://www.crossref.org/>) byl do provozu uveden v roce 2000 neziskovou a nezávislou Asociací pro mezinárodní propojování vydavatelů (Publishers International Linking Association, PILA). Jde o síť kolaborativního charakteru, jejímž hlavním posláním je **propojování online vědeckých publikací** na základě **propojování bibliografických záznamů** včetně **záznamů citací**. Založením této sítě její tvůrci reagovali zejména na **problém neplatných hyperlinků při odkazování na dokumenty lokalizované v síti Internet** [BRAND, 2004, s. 227]. Systém CrossRef byl původně vytvořen s cílem **zlepšit přístup** k materiálům z oblasti přírodních a technických věd a medicíny, ale postupně byl rozšířen na všechny obory. Systém samotný nezahrnuje databáze úplných elektronických textů a neposkytuje propojovací služby přímo koncovým uživatelům. Umožňuje účastníkům se vydavatelům sdílet metadata dokumentů a poskytuje jejich online službám podporu při **propojování záznamů**. Výsledkem je vysoce účinný propojovací systém, jehož prostřednictvím vědec-uživatel může kliknout na libovolnou citaci v online časopiseckém článku a během okamžiku tento citovaný dokument v elektronické formě získá.

Významným nástrojem služby CrossRef je **trvalý identifikátor DOI** (Digital Object Identifier). Jde o alfanumerický kód přidělovaný digitálním objektům (článkům, kapitolám nebo i celým dokumentům). V systému CrossRef je spojován se základními metadaty a s URL identifikovaného objektu. DOI, který je americkým standardem (ANSI/NISO Z39.84), je určen k jednoznačné globální identifikaci elektronického obsahu a k jeho trvalé lokalizaci. Celosvětově je systém DOI řízen **Mezinárodní nadací DOI** (International DOI Foundation, IDF, <http://www.doi.org/>). CrossRef je oficiální agenturou (agentem) pro přidělování identifikátorů DOI. Každému digitálnímu dokumentu, který má být v systému CrossRef registrován, vytváří příslušný vydavatel kód DOI (identifikace vydavatele je uvedena v jeho prefixu). Kódy DOI jsou propojovány s **metadaty (včetně abstraktu, citací a URL)**, která tvoří vydavatelé. Ti je v definovaném formátu XML ukládají ve specifické databázi metadat **MDDB** systému CrossRef. Párová informace DOI+URL je pak následně automatickým procesem transferována do centrálního adresáře DOI s tím, že se vytváří trvalá vazba mezi DOI a dokumenty lokalizovanými v repozitářích vydavatelských digitálních

knihoven. V rámci vyhledávacích procesů je využívána **směrovací služba Mezinárodní nadace DOI** (<http://dx.doi.org/>). (Příklad zápisu identifikátoru DOI v záznamu citovaného dokumentu – viz [HARNAD, 2004; BRAND, 2004]). Systém CrossRef integroval a využívá také identifikátorů **OpenURL**. Uživatel, který v rámci vyhledávání sleduje hyperlink DOI, vyvolá automatický proces vyhledání metadat z databáze systému CrossRef a následné vytvoření OpenURL, které je pak zasláno do lokální směrovací služby. Úplný text hledaného článku pak může být získán z různých zdrojů v rámci knihovnických služeb.

V rámci odděleného procesu vydavatelé v současnosti předkládají také **citace** obsažené v každém článku. Jde **zatím o dobrovolný proces**. Vydavatelé seznamy citací vkládají buď společně s metadaty nebo zvlášť do předřazené komponenty databáze MDDDB (Reference Resolver) za účelem zjištění identifikátorů DOI citovaných dokumentů. Vydavatel tak může vytvořit trvalou vazbu mezi citacemi a záznamy dokumentů, které již jsou v systému CrossRef zaregistrované. Zainteresovaní vydavatelé proto nyní mohou v rámci vyhledávacího procesu zjišťovat vazby typu „**Cited by**“, tj. vazbu na dokumenty, které citovaly jimi publikovaný digitální obsah. Jde o novou službu systému CrossRef, která slouží účastníkům systému, veřejně přístupný tento citační systém není. Tohoto systému ale bude využito v novém projektu zvaném **CrossRef Search**, který je připravován se systémem Google Scholar (zatím se ho účastní 45 vydavatelů registrovaných v systému CrossRef).

V rámci sítě služby CrossRef bylo k 26.4. 2006 zaregistrováno již **20 000 000 dokumentů** (s přiděleným identifikátorem DOI). Databáze narůstá velkým tempem, polovina dokumentů z uvedeného počtu byla zaregistrována v posledních dvou letech. Většina z nich byla přidělena článkům z přibližně **14 600 elektronických časopisů**. Zhruba **1 600 000 identifikátorů** bylo přiděleno již také knihám a konferenčním sborníkům (na úrovni titulu nebo jeho části). V posledním období bylo zahájeno přidělování identifikátorů také titulům šedé literatury (zprávám a disertacím). Pokračuje také **retrospektivní přidělování identifikátorů DOI** starším dokumentům. Systém zahrnuje téměř **1 650 vydavatelů** a dalších producentů a distributorů společností, které publikují materiály v digitální formě. Za účast platí poplatky, jejich výše je stanovena také na základě velikosti jejich produkce. K dubnu 2006 se systémem CrossRef spolupracuje již také zhruba **1037 knihoven**.

Systém CrossRef zřejmě míří k vybudování velké databáze záznamů vědecké literatury (ne nepodobné základním databázím známých velkých citačních služeb). Velký nárůst registrovaných citačních vazeb zároveň potvrzuje, že v oblasti citačních služeb vzniká další konkurent.

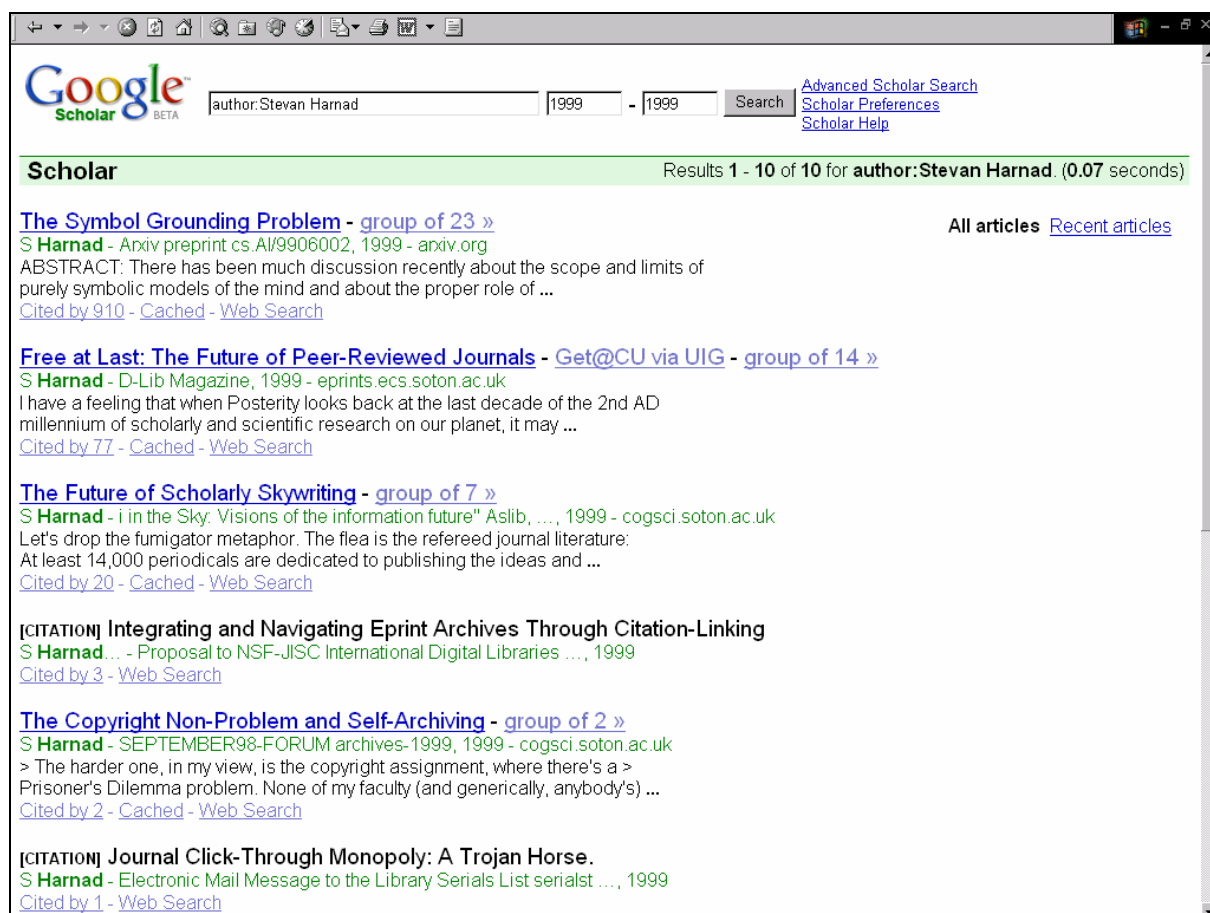
3.2.4 Google Scholar

Společnost **Google**, která provozuje stejnojmennou světově známou vyhledávací službu Internetu (<http://www.google.com>), spustila v listopadu 2004 do provozu nový systém pro zpřístupňování vědecké literatury zvaný **Google Scholar** (<http://scholar.google.com/>). Systém je bezplatný a zahrnuje také **citační služby**. Jeho uvedení bylo doprovázeno nadšením i nedůvěrou a četnými otázkami, které se týkaly rozsahu a pokrytí jeho databáze. I po více než roce a půl však řada těchto otevřených otázek zůstává nezodpovězena [NEUHAUS, 2006, s 127]. Volně přístupná databáze systému Google Scholar je navíc pro světové uživatele k dispozici stále pouze v beta verzi. Nicméně popularita a návštěvnost systému je velká.

Systém Google Scholar je výsledkem úsilí skupiny vydavatelů společností, jež zajistily **obsah** (především metadata, tj. dodali určité části ze svých databází bibliografických záznamů), a společnosti Google, která se postarala o **programové a technické zabezpečení** systému. Společnost Google ovšem poskytuje zatím, na rozdíl od největších komerčních

citačních služeb, jen **velmi málo informací** o svém novém produktu. Není zcela jasné obsahové zaměření databáze, její tvorba, způsoby doplňování, softwarové zabezpečení aj.

Celkový **počet zdrojových záznamů** v databázi Google Scholar není znám (jde zřejmě řádově o miliony záznamů). Znáno není ani celkový počet zastoupených vydavatelů, preprintových či reprintových serverů nebo samotných databází. Mnoho odborníků se pokouší různými způsoby zjistit tuto informaci. V jedné z **posledních studií z roku 2006** [NEUHAUS, 2006], která prezentovala výsledky měření rozsahu a obsahu databáze (porovnávána byla s obsahem 47 jiných vědeckých databází), její autoři zjistili, že Google Scholar má ve své beta verzi některé silné stránky, ale také řadu stránek slabých. Pokrytí online časopisů s otevřeným přístupem, zdarma zpřístupňovaných vědeckých databází a jednotlivých vydavatelských databází je poměrně veliké. Zastoupení databází z humanitních disciplín a umění je však minimální, zastoupení databází ze společenských věd, výchovy a obchodu je zhruba 50% a je spíše nahodilé. Největší množství literatury je zastoupeno z oblasti přírodních a aplikovaných věd a medicíny (převaha volně dostupných zdrojů z daných oborů je zjevná). Slabé je rovněž pokrytí ne-anglicky psané literatury a chybí starší literatura. Při rešerších tedy uživatelé mnohdy netuší, že nalézají jen určitou část z celkového množství existující literatury. V databázi se vyskytuje také duplicita řady záznamů.



Obr. č. 2: rešerše prací vysoce citovaného autora Stevana Harnada a také citací jeho prací v systému Google Scholar (beta verzi); záznamy zahrnují hyperlinky na záznamy citujících autorů [2006-05-19]

Google Scholar nabízí především **vyhledávání záznamů zdrojových dokumentů** s možností návazného získání **úplných textů** z řady volně dostupných online zdrojů. Mnohé volně dostupné dokumenty však stále nejsou registrovány, stejně tak chybí řada vazeb na existující

úplné texty. Systém proto intenzivně spolupracuje s mnoha knihovnami a dalšími dodavatelskými službami na **propojování záznamů s fondy úplných textů**, které tyto subjekty vlastní (viz například hyperlink u druhého záznamu na obr. č. 2 k propojení na rozhraní Jednotné informační brány Univerzity Karlovy v Praze - „Get@CU via UIG“, tj. Uniform Information Gateway). Využito je technologií SFX a OpenURL.

Systém standardně nabízí rozhraní pro **jednoduché a pokročilé vyhledávání**. Vyhledávat lze podle libovolného klíčového slova, autora zdrojového článku, názvu zdrojové publikace, data vydání a dokonce podle široké předmětové kategorie (nabízeno je 7 oblastí poznání).

V rámci vyhledávacího procesu lze předem nastavit další kritéria: jazyk komunikace se systémem Google (aktuálně 13 nabízených jazyků), jazyk dokumentu, nastavení propojení na fondy určité knihovny, počet výsledných záznamů zobrazovaných na stránce a nastavení linku k exportu záznamu do bibliografických manažerských systémů (aktuálně 5 nabízených systémů). Řazení nalezených záznamů je dáno specifickým algoritmem systému Google („page rank“), který je postaven především na webové citovanosti dokumentu, a do výpočtu je navíc zahrnuta také jeho klasická citovanost. Jiná pořadání záznamů, kromě možnosti zobrazení nejnovějších dokumentů (Recent articles), zatím nejsou nabízena.

Jak ve své analýze systému Google Scholar z listopadu 2005 uvedl P. Jacsó, systém Google Scholar ve své beta verzi vykazuje **při vyhledávání řadu problémů**. Některé rešeršní funkce jsou chybně koncipované a špatně implementované a výstupy záznamů obsahují mylné informace [JACSÓ, 2005a, s. 1538]. Problémy byly vykážány při využívání základní booleovské operace OR. Značné nejasnosti se týkají také vyhledávání záznamů podle roku vydání či vytvoření dokumentu (systém například při postupném zvětšování časového období předkládá postupně menší počty výsledných záznamů). Testovací rešerše všech dokumentů za období let 868 až 2005, provedená P. Jacsó v roce 2005, vedla k číslu zhruba 1 360 000 záznamů. Zdá se, že šlo o mylné číslo celkového počtu záznamů (ve srovnání se systémem WOS a Scopus by šlo o velmi malý počet). Zavádějící výsledky byly zjištěny rovněž u citovanosti dokumentů [JACSÓ, 2005a, s. 1541].

Google Scholar nabízí zdarma službu sledování **citovanosti zdrojového dokumentu**, která je reprezentována hyperlinkem „Cited by“ (viz obr. č. 2) vedoucím k záznamům citujících dokumentů. Je potřeba ale brát do úvahy fakt, že jde o citovanost měřenou pouze na základě dokumentů, které jsou indexovány systémem Google Scholar. Specifické záznamy začínající návěštím „[Citation]“ reprezentují záznamy citované literatury (viz poslední záznam na obr. č. 2), které byly extrahovány ze seznamů použité literatury zdrojových dokumentů (nejsou prozatím předmětem indexace). Záznamy těchto citací ale nejsou klikatelné na primární dokumenty přímo, je nutné použít hyperlinku „Web Search“ a titul dohledat na Internetu. V databázi se vyskytují také duplicity citací. Celkový počet citací není veřejnosti znám.

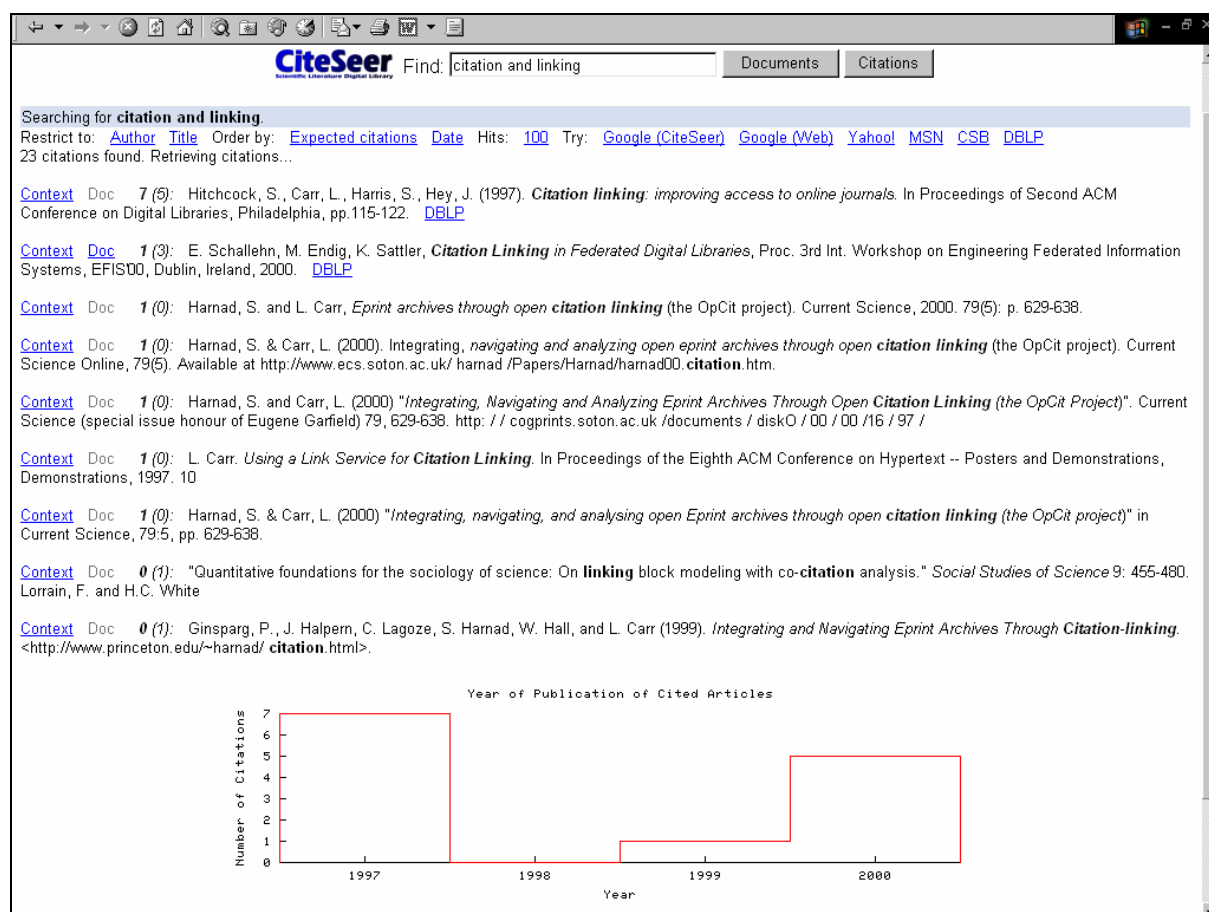
Systém zatím také nenabízí žádné možnosti bibliometrických analýz databáze Google Scholar. Nelze zjistit její velikost a složení, tak jak tomu je u systému WOS a Scopus. K dispozici není ani seznam vydavatelů, ani seznam zdrojových časopisů nebo sborníků. Není udána retrospektiva databáze ani pokrytí zastoupených oborů.

Systém Google míří k vytvoření velkého zdroje vědecké literatury. Očekává se však podstatné doplnění informačních zdrojů (včetně materiálů z elektronických volně dostupných archivů), zlepšení vyhledávacího procesu systému (včetně vytvoření specifických rejstříků pro řadu údajů zahrnutých bibliografických záznamů). Rozšířeny by měly být možnosti řazení záznamů. Konsolidovány by měly být také citace na základě rozrůstajícího se registru DOI (o projektu zvaném CrossRef Search viz informaci v části 3.2.3). Měly by být zavedeny nástroje pro zpracování záznamů citujících a citovaných dokumentů [HARDY, 2005].

3.2.5 CiteSeer

Zajímavý směr v oblasti zpracování a využívání citací v prostředí digitálních knihoven je zastoupen systémem **CiteSeer.IST** (<http://citeseer.ist.psu.edu/>). Systém dvakrát změnil jméno (původně se nazýval CiteSeer, později ResearchIndex). Byl vytvořen v roce 1997 zaměstnanci Výzkumného ústavu společnosti NEC (Princeton, N.J., USA) – S. Lawrencem, L. Gilesem a K. Bollackerem. V současnosti je systém již provozován Školou informačních věd a technologií (Pensylvánská státní univerzita) a jeho správou a dalším rozvojem je pověřen L. Giles (na jaře 2005 byly získány grantové prostředky ve výši 1,2 milionů dolarů na obohacení a zlepšení funkčnosti celého systému [JACSÓ, 2005b]).

Specifikem systému je **citační rejstřík** budovaný na bázi digitální knihovny, která zahrnuje volně dostupné dokumenty v elektronické formě z oboru **počítačové vědy** a nově také z oboru **informační vědy** v celosvětovém záběru. Získávanými zdrojovými dokumenty jsou volně dostupné preprinty statí z konferencí nebo časopiseckých článků, výzkumné zprávy, disertace aj. Limitem jsou formáty získávaných typů dokumentů (PostScript a PDF). V oboru počítačové vědy jde ale o podstatné formáty, které pomáhají udržovat vysokou kvalitu databáze. Navíc se na jejich základě provádí automatická katalogizace daných dokumentů.



Obr. č. 3: přehled záznamů citací včetně grafické prezentace na téma „citace a propojování“ v systému CiteSeer – v uspořádání podle citovanosti daných dokumentů [2006-04-28]

Veškeré procesy v systému probíhající jsou zajišťovány programovými prostředky. Unikátní technologie zajišťuje **autonomní indexaci citací (ACI, Autonomous Citation Indexing)**. Na základě ní je vytvářen specifický index citací (vedle základního indexu úplných textů zdrojových dokumentů). Program ACI umožňuje zjišťovat a extrahovat z dokumentů citace,

identifikovat citace reprezentující stejný dokument, byť se vyskytují v různých formách, a také identifikovat zajímavé **kontexty citací** ve zdrojových dokumentech. Dokumentace systému proklamuje zjištěnou **chybovost** jenom cca 5 %. (Podrobnosti o zpracování informací v systému CiteSeer viz publikovaný článek z roku 2001 [BRATKOVÁ, 2001]). Databáze zahrnuje v současné době již téměř **750 000** úplných textů dokumentů a řádově **miliony citací** (počet není na WWW stránkách systému v současnosti uveden). Uživatelům systém poskytuje zajímavé výstupy a služby, řada z nich je zcela nového charakteru.

Systém CiteSeer.IST nabízí v současnosti několik málo způsobů vyhledávání informací. V rámci přímé formulace dotazů může být uplatněno libovolné klíčové slovo, dotaz lze uplatnit jak vůči indexu zdrojových dokumentů, tak indexu citací. Zajímavé výstupy nabízí vyhledávání citací (viz obr. č. 3). Výsledkem je rešerše **sjednocených záznamů citovaných dokumentů** (doplněná grafickou prezentací vývoje citovanosti na dané téma), ze kterých se lze dostat pomocí specifické funkce (Context) ke zkráceným záznamům zdrojových dokumentů (zahrnují části textů lokalizovaných okolo citačního identifikátoru a původní záznam citace). Ze zkráceného záznamu se lze pomocí hyperlinku dostat k úplnému záznamu zdrojového dokumentu a z něho k úplnému textu (v různých formátech). Seznamy záznamů citací lze třídit na základě mnoha údajů, k nejzajímavějším patří uspořádání podle **očekávané citovanosti** (Expected citations - viz obr. č. 3). Systém nabízí také několik zajímavých statistik: seznam nejvíce využívaných zdrojových dokumentů, nejvíce citovaných zdrojových nebo nejvíce citovaných dokumentů (vlastní citační rejstřík), seznam nejvíce citovaných autorů aj. Zajímavostí je také automaticky vytvářený předmětový rejstřík (zatím jen pro počítačovou vědu). K návazným funkcím systému CiteSeer.IST patří služba **průběžného informování** uživatelů o nových přírůstcích ve zdrojových dokumentech nebo citacích.

Systém CiteSeer představuje zajímavý experiment v současném rozvoji citačních služeb na Internetu. Reprezentuje modelovou situaci výrazného trendu propojování informací a integrace informačních systémů. Přispívá k celkovému zlepšení organizace, zpracování, vyhledávání, rozšiřování a zpřístupňování vědecké literatury v rámci sítě Internet. Kvalita softwaru byla oceněna také tím, že společnost The Thomson Corporation si ho vybrala jako základ pro svůj nový projekt Web Citation Index (viz část 3.2.7).

3.2.6 Citebase Search

Z mnoha hledisek pozoruhodná citační služba **Citebase Search** (<http://www.citebase.org/>) je produktem původního americko-britského projektu **OpCit** (The Open Citation Project, <http://opcit.eprints.org/>). Jak ve svém hodnocení uvedl P. Jacsó, Citebase Search je jeho „královským klenotem“ [JACSÓ, 2004a, s. 57]. Projekt OpCit (1999-2002) se zabýval otázkami vnitřního a vnějšího propojování bibliografických záznamů a citací uložených ve volně dostupných archivech elektronických tisků. Jeho neméně důležitým cílem bylo také zajišťování komplexních bibliometrických či infrometrických analýz, které pomáhají lépe porozumět tomu, jak je vědecká literatura využívána v tvůrčím procesu.

Systém **Citebase Search** byl vyvinut na Southamptonské univerzitě (Velká Británie). Jeho existence byla odborné veřejnosti poprvé oznámena v prosinci 2001, uživatelé jej však mohli začít využívat až po jeho integraci s archivem arXiv.org v srpnu 2002. Citebase Search je autonomním scientometrickým nástrojem ke zjišťování potenciálu informačních materiálů uložených v elektronických archivech s otevřeným přístupem. Systém je stále v **experimentální fázi** svého vývoje, jak je uvedeno na jeho hlavní webové stránce, nelze ho zatím využívat například pro potřeby reálných akademických evaluací, protože pokrytí citací a jejich analýza není (prozatím) úplná. Jde o jeden z prvních systémů, který začal pracovat na bázi principů OAI (jako poskytovatel služby).

The screenshot shows the Citebase Search web application. At the top, there is a search bar with the text 'citebase Search' and links for 'Help', '(Impact)', 'Health-Warning', 'Metadata', 'Citation', and 'OAI Identifier'. Below the search bar is a form with the following fields: 'Author(s) (explain?)' containing 'Brody Tim', 'Title/Abstract Keywords' containing 'Web', 'Publication title' (empty), and 'Creation Date' with a range from '1995' to '2006'. There are also dropdown menus for 'Rank matches by:' (set to 'Descending') and 'Citations (Paper)'. A 'Search' button and a 'Clear' button are present. Below the form, a status bar indicates 'Showing 1 - 8 of 8 found [1-100 in XML, Refer, BibTeX, RSS]' and 'Query took 0.023 seconds'. The search results are listed below, each with a title, author(s), date, and OAI ID. The first result is 'Earlier Web Usage Statistics as Predictors of Later Citation Impact' by Brody, Tim; Harnad, Stevan; Carr, Les (2005-01-01) with OAI ID 10647. The second result is 'A Scalable Architecture for Harvest-Based Digital Libraries - The ODU/Southampton Experiments' by Liu, Xiaoming; Brody, Tim; Harnad, Stevan et al (2002-05-28) with OAI ID 0205071. The third result is 'Digitometric Services for Open Archives Environments' (null) by Brody, Tim; Kampa, Simon; Harnad, Stevan et al (2003-01-01) with OAI ID 7503. The fourth result is 'Open Citation Linking: The Way Forward' (null) by Hitchcock, Steve; Brody, Tim; Gutteridge, Christopher et al (2002-10-01) with OAI ID 7717. The fifth result is 'Evaluating Citebase, an open access Web-based citation-ranked search and impact discovery service' (null) by Hitchcock, Steve; Woukeu, Arouna; Brody, Tim et al (2003-05-01) with OAI ID 8204.

Obr. č. 4: zkrácené záznamy vyhledaných dokumentů registrované v databázi Citebase Search, uspořádané podle jejich citovanosti [2006-04-14]

Do databáze Citebase Search jsou pomocí protokolu OAI-PMH sklížena **metadata zdrojových dokumentů** z několika elektronických archivů. Metadata jsou získávána ve formátu Dublin Core a běžně k nim patří název dokumentu, abstrakt nebo komentář, jména autorů (v nestruturovaném tvaru, příjmení jsou ale zpravidla odlišitelná pro potřeby dalšího zpracování), lokace zdroje (např. URL), data různých typů (datum revize aj.). V některých archivech jsou součástí sklizených metadat také **citace** ve formě volného textu (jsou „strukturované“ pouze autory podle různých citačních stylů). Název a abstrakt zdrojového dokumentu vcházejí do invertovaného souboru pro potřeby volného vyhledávání v úplném textu. Systém Citebase Search však také na základě **vlastního rozhraní** sklízí úplné texty zdrojových dokumentů (nejsou ale v systému natrvalo ukládány), ze kterých **extrahuje záznamy citovaných dokumentů**, jež jsou v rámci dalšího zpracování s těmito dokumenty propojovány, jsou-li volně na WWW lokalizovány [BRODY, 2003, s. 4]. Extrahovány jsou běžně přesné názvy časopisů (často zkrácené), rok publikování, ročník, číslo a stránkování. Extrakce citací probíhá rozdílně podle toho, v jakém formátu jsou k dispozici úplné texty. Snadno lze například získat strukturované citace z archivu BioMed Central (dokumenty jsou ve formátu XML). Složitější způsob získání citací je nasazen například u většiny textů z archivu arXiv.org (texty jsou ve zdrojovém formátu LaTeX, popř. PDF, nutná je tedy jejich konverze do jednoduchého textu před vlastní extrakcí citací). K rozpoznávání záznamů citací a jejich dílčích údajů je využíváno specifických algoritmů. Sklizená metadata a extrahované záznamy citací slouží k tvorbě citační databáze, která je základem následných webových služeb včetně možného exportu dat pomocí protokolu OAI-PMH a tvorby datových analýz [BRODY, 2003, s. 5].

K dubnu 2006 systém Citebase Search ohlásil téměř **465 000 záznamů zdrojových dokumentů** (s propojením na jejich úplný text). Největší počet pochází z partnerského archivu arXiv.org (až 370 079 záznamů), který byl původně jediným archivem projektu. Další záznamy pocházejí například z archivu PubMed Central (29 173), BioMed Central (25 134), repozitáře Southamptonské univerzity (18 849) aj. Sklizen je také známý archiv E-LIS pro knihovní a informační vědu (téměř 3 960 záznamů) nebo archiv pro kognitivní vědy CogPrints (2 730 záznamů). Předpokládá se, že v budoucnu systém Citebase Search pokryje všechny volně dostupné elektronické archivy. Počet **registrovaných citací** byl v dubnu 2006 téměř **13 380 000**, více než **3 000 000** z nich je propojeno na úplný text. V rámci citací je identifikováno (příjmení, křestní jméno) zhruba 350 000 autorů. Ve stejné době systém zaregistroval již **5 640 129** stažení úplných textů.

Earlier Web Usage Statistics as Predictors of Later Citation Impact
 Authors: [Brody, Tim](#); [Harnad, Stevan](#); [Carr, Les](#)
 See also ([explain?](#)): [oai:arXiv.org/cs/0503020](http://oai.arXiv.org/cs/0503020), [oai:eprints.ecs.soton.ac.uk:10713](http://oai.eprints.ecs.soton.ac.uk:10713), [oai:eprints.ecs.soton.ac.uk:10647](http://oai.eprints.ecs.soton.ac.uk:10647).

THIS IS NOT THE LATEST VERSION: PLEASE SEE THE LATEST VERSION

Full-text available from: <http://eprints.ecs.soton.ac.uk/10647/02/timcorr.htm>
<http://eprints.ecs.soton.ac.uk/10647/>

Export record as: [BibTeX](#)

- [This Article's Citation/Hit History](#)
- [All Articles Cited by this Article \(Reference List\)](#)
- [Top 5 Articles Citing this Article](#)
- [All Articles Citing this Article](#)
- [Top 5 Most Co-cited Articles with this Article](#)
- [All Articles Co-Cited with this Article](#)

Based on record (harvested at) [oai:eprints.ecs.soton.ac.uk:10647](http://oai.eprints.ecs.soton.ac.uk:10647) (2005-08-18)
 date 2005-01-01
 format html <http://eprints.ecs.soton.ac.uk/10647/02/timcorr.htm>
 type Article

This Article's Citation/Hits History ([explain?](#))
 > Top #Reference List #Cited By #Co-Cited All Citations All Co-Citations

Use the [Correlation Generator](#) to explore the correlation between download impact ("hits") and citation impact.

Summary

Citebase is currently only an experimental demonstration. Users are cautioned not to use it for academic evaluation yet. Citation coverage and analysis is [incomplete](#) and hit coverage and analysis is both [incomplete](#) and [noisy](#).

Caution!	Citations	Full-text Downloads
To this Article	4	342
To authors (mean)	2.73987	77.5479

These figures represent citations to and downloads of the work, and not necessarily the record shown. For more information see [information on duplicate records](#).

History

— Citebase-identified Citations (4)
 — Full-text Downloads (342)

Obr. č. 5: část obrazovky s úplným záznamem dokumentu z databáze Citebase Search včetně údajů a grafu o vývoji citovanosti vyhledaného titulu [2006-04-14]

Rozhraní pro vyhledávání informací systému Citebase Search nabízí standardně **vyhledávání v indexu metadat**, a to podle jména autora, klíčových slov z názvu/abstraktu zdrojového článku, názvu časopisu a data vytvoření nebo publikování dokumentu. Zvláště lze vyhledávat podle citací a podle OAI identifikátoru (viz například oai:arXiv.org/cs/0503020 v záznamu na obr. č. 5). Výsledky rešerší vedou vždy k seznamu zkrácených záznamů zdrojových dokumentů. Specifickou nabídkou systému je **uspořádání či řazení nalezených záznamů** (Rank matches by - viz obr. č. 4): podle citačního vlivu článku (Citations (Paper)) - viz také obr. č. 4, citačního vlivu autorů článku (Citation (Author)), počtu webového stažení článku

(Hits (Paper)), počtu webového stažení dokumentů autorů článku (Hits (Author)), data vytvoření článku (Date (Creation)) a data poslední aktualizace (Date (Update)).

Ze zkráceného záznamu zdrojového dokumentu lze pomocí hyperlinku získat **úplný záznam**, který se vyznačuje bohatostí poskytovaných informací (viz obr. č. 5). Zahrnuje úplné bibliografické údaje, zpravidla vždy abstrakt, URL s uloženým úplným textem a kompletní citační informace. V přehledné tabulkové formě, která je generována korelačním generátorem, je uveden počet citací a webových stažení daného článku a průměrný počet citací a průměrný počet webových stažení všech dokumentů autorů daného článku. Grafická informace zobrazuje uvedené parametry v čase. V dalších sekcích úplného záznamu jsou k dispozici seznamy citací z daného článku, seznamy citací na daný článek a seznamy kocitací článků.

Systém Citebase Search nabízí uživatelům sofistikované rešeršní prostředí, poskytuje specifické bohaté citační služby týkající se volně dostupné vědecké literatury. Data z databáze lze využít k dalším čítným scientometrickým, bibliometrickým a informetrickým analýzám, byť jde zatím o experimentální provoz. V budoucí reálné verzi bude potřebné rozšířit obsah databáze o informace z dalších archivů.

3.2.7 Web Citation Index

V roce 2004 oznámila společnost Thomson, která produkuje a zpřístupňuje světoznámé citační rejstříky v systému WOS (viz část 3.2.1), vytvoření doplňkového citačního rejstříku nové generace s názvem **Web Citation Index™** (WCI). Nový rejstřík má zpracovávat vybranou kvalitní vědeckou literaturu volně dostupnou na WWW včetně citací. Záměr sledovat a zpracovávat tuto literaturu byl přijat již v roce 1998. Nový systém je budován ve spolupráci se společností NEC, která poskytla program pro automatické extrahování a zpracování citací **ACI** (Autonomous Citation Indexing), užívaný v systému CiteSeer (viz část 3.2.5). Společnost Thomson tento program adaptovala a v mnoha směrech rozšířila (zahrnuje například doplňkový program zajišťující sklizení metadat ze stanovených repozitářů podle protokolu OAI-PMH) [MARTELLO, 2006].

V rámci pilotního projektu se systém orientuje na vybrané institucionální repozitáře a mezinárodní předmětově orientované elektronické archivy (archiv arXiv.org aj.), jejichž obsah je volně dostupný na WWW. Na projektu spolupracuje sedm dalších institucí: Australská národní univerzita, Kalifornský technologický institut, Cornellská univerzita, Společnost Maxe Plancka, Monashská univerzita, Rochesterská univerzita a NASA. Jmenované instituce podávají připomínky k vyhledávacímu systému a jejich repozitáře byly určeny jako první k testování softwaru pro sklizení metadat.

Stejně jako základní rejstříky je i WCI budován v polyoborovém záběru. Předmětem zpracování je ale již **více typů elektronických zdrojů**. Jsou jimi především preprinty článků z časopisů nebo konferenčních sborníků, technické zprávy, disertační a diplomové práce a jiné typy šedé literatury. Jde o materiály vědeckého charakteru lokalizované zejména v otevřených institucionálních repozitářích a archivech elektronických tisků. Prozatím se systém orientuje na formáty PDF a PostScript. Dokumenty pro zpracování v systému WCI jsou předmětem **pečlivého výběru**. Na základě stanovených kritérií ho zajišťují **editoři** [MARTELLO, 2006]. Pozornost je zaměřena na elektronické archivy a repozitáře, které spravují a provozují významné autorizované vědecké instituce (univerzity, vědecké ústavy, národní laboratoře, vládní či mezinárodní agentury a společnosti aj.) Z výběru mohou být vyloučeny repozitáře, které nejsou kvalitně spravovány (zahrnují velké množství neplatných linků, nejsou dlouho aktualizované, doba odezvy je dlouhá apod.). Vyloučeny mohou být také repozitáře, které neposkytují kvalitní metadata, nemají alespoň základní předmětové třídění uložených materiálů nebo nezajišťují bezproblémový přístup k úplným textům. Dalším

zvažovaným kritériem může být frekvence nebo aktualizace nových přírůstků do repozitářů a jejich případná kontrola ze strany jejich moderátorů.

The screenshot displays the 'Web Citation Index' interface. At the top, there is a navigation bar with buttons for 'WELCOME', 'HELP', 'GENERAL SEARCH', 'CITED REF SEARCH', 'SEARCH HISTORY', and 'ADVANCED SEARCH'. Below this is a 'Full Record' section for 'Record 59 of 8,733 (Set #10)'. The record details include:
Title: Black Holes on Cylinders Are Not Algebraically Special
Author(s): De Smet Pieter-Jan
Repository: arXiv.org e-Print Archive
Cited References: 18 **Times Cited:** 6
Associated Records: 1 (with a 'FIND RELATED RECORDS' button)
Abstract: We give a Petrov classification for five-dimensional metrics. We classify Ricci-flat metrics that are static, have an SO(3) isometry group and have Petrov type 2. We use this classification to look for the metric of a black hole on a cylinder, i.e. a black hole with asymptotic geometry four-dimensional Minkowski space times a circle. Although a black string wrapped around the S1 and the five-dimensional black hole are both algebraically special, it turns out that the black hole on a cylinder is not. Address after 1 September 2002: SUNY at Stony Brook, USA
WCI Number: 1007293
On the right side, there are several utility sections:
Output This Record: Includes a dropdown for 'Bibliographic Fields', buttons for 'PRINT', 'E-MAIL', and 'SAVE', an 'EXPORT TO REFERENCE SOFTWARE' button, and an 'ADD TO MARKED LIST' button (showing 0 records marked).
Create Citation Alert: Includes a 'CREATE CITATION ALERT' button and text: 'Receive e-mail alerts on future citations to this record. (Requires registration.)'
Additional Links: Includes a 'VIEW FULL TEXT' button.
View in Web of Science: Includes links for 'Cited References', 'Citing Articles', and 'Related Records'.
View record in: Includes links for 'Inspec' and 'Web of Science'.

Obr. č. 6: ukázka záznamu preprintu z archivu arXiv.org z testované databáze WCI [převzato z: ESPECHE, 2006]

Systém WCI byl do zkušebního (neveřejného) provozu uveden v listopadu 2005. Testování vyhledávání informací z pilotního rejstříku WCI v rešeršním systému WOS aktuálně stále probíhá (je ohlášeno až do října 2006). Testů se zúčastňují vybrané partnerské instituce z řady zemí. Systém nabízí stejné či podobné rešeršní funkce jako u základních citačních rejstříků (viz část 3.2.1) včetně uspořádání a zobrazování výsledků, jak vyplývá z prezentace pracovníka společnosti Thomson z letošního roku [ESPECHE, 2006]. Specifickou novinkou je zajištění **indexace a vyhledávání v úplných textech dokumentů** s tím, že texty ale nejsou v systému ukládány (nejde o digitální knihovnu). Systém zobrazuje jen záznam s abstraktem a hyperlink na úplný text lokalizovaný v původním repozitáři (View Full Text, viz obr. č. 6). Nabízena je i funkce rozšířeného citačního vyhledávání (podle autora, repozitáře a roku publikování) s možností následné bibliometrické analýzy. Podrobnější vyhodnocení systému bude možné až po jeho zveřejnění (komerčně) koncem roku 2006 nebo počátkem roku 2007.

3.3 Souhrn

Vybrané citační systémy, analyzované výše v textu, poskytují koncovým uživatelům užitečné a zajímavé **výstupy a služby**. Nejnovější systémy mají moderně řešená rozhraní pro vyhledávání (Scopus aj.) a nové nástroje pro navigaci z výsledků rešerší. Některé systémy nabízejí webové linky ze záznamů citované literatury (WOS, Scopus, Citebase Search), jejich největší počet ale zatím vykazuje systém WOS, i když linky zatím nevedou na volně dostupné verze dokumentů uložené v elektronických repozitářích (tyto vazby jsou právě zaváděny v novém systému Web Citation Index). U některých systémů zatím chybí nástroje pro analýzu výsledných rešerší (Google Scholar). Seznamy kocitujících dokumentů (dokumenty, které sdílejí podobné seznamy citací jako výchozí dokument) zajišťuje služba WOS a Scopus. Efektivní seznamy kocitovaných dokumentů (tj. dokumentů, které byly společně citovány aktuálním a nějakým jiným dokumentem) zajišťuje služba Citebase Search.

Rozbor systémů zároveň ukazuje, že v současnosti dochází k velké **multiplicitě při získávání a propojování citací** stejných titulů literatury včetně literatury volně dostupné na WWW. Citační systémy poskytují podobné, ale zároveň limitované služby. Není jasné, jak by vlastně měl být budován **optimální citační rejstřík pro volně dostupnou literaturu**. Systémy Citebase Search a CiteSeer jsou založeny téměř výlučně na elektronických tiscích archivovaných samotnými autory. Záznamy citované literatury jsou v tomto případě k dispozici v **nestrukturované formě**, a proto musejí tyto citační systémy zavádět určité algoritmy k rozkladu výchozích citačních textů a k vytvoření jednotné formy záznamu citovaného dokumentu, která bude vhodná k automatickému propojování na dokumenty nebo jiné záznamy. U systému Google Scholar není postup získávání a propojování citací veřejnosti znám, z provozu systému lze však usuzovat, že má k dispozici také vlastní algoritmus k rozkladu nestrukturovaných citací z úplných textů a jejich propojování. Citační rejstříky v systému WOS a Scopus využívají jak **strukturovaných dat** od vydavatelů, tak automatického rozkladu citací. U systému WOS je navíc vkládán velký podíl lidské práce.

4 Citační propojování a strukturování citací

Bibliografické údaje, které popisují citovaný dokument (jeho název, autory, název časopisu aj.), bývají propojovány na tento dokument tak, že dochází k jejich vyhledávání v nějaké bibliografické databázi. Tímto způsobem se vytvářejí propojení mezi citujícími a citovanými dokumenty. **Citační propojení** mohou být využita k vytvoření citačního rejstříku, který uživateli umožní zjistit dokumenty, které daný dokument citovaly.

Metadatové záznamy plnotextových dokumentů z volně dostupných elektronických archivů, které vyhovují protokolu OAI-PMH, by měly v optimálním případě zahrnovat seznamy citovaných dokumentů, byť to uvedený protokol nepožaduje [HARDY, 2005]. Některé archivy vstřícně (například archiv E-LIS pro knihovní a informační vědu) dovolují autorům, aby seznam záznamů použité literatury sami nakopírovali do webového formuláře pro ukládání metadat (ruční způsob není ovšem nikterak pohodlný). Software, zajišťující chod archivu, pak může podporovat kontrolu přesnosti záznamů citovaných dokumentů prostřednictvím nějaké vybrané síťové služby s bibliografickou nebo citační databází.

V systémech citačního propojování, které jsou využívány v komerčních digitálních knihovnách, musejí být metadata přesná a úplná, aby celý systém spolehlivě fungoval. V praxi se přesto vyskytují chyby. V komerční sféře vývoj proto vyústil k budování kolaborativních sítí, v jejichž rámci záznamy citací vytvářejí profesionálové (viz například systém CrossRef - část 3.2.3). Na WWW jsou ale již k dispozici také nové volně dostupné služby (**CiteuLike**, <http://www.citeulike.org> nebo **Connotea**, <http://www.connotea.org>), které

zajišťují kolaborativní budování databází záznamů, jež by měly v budoucnu odbourat ruční tvorbu a ukládání záznamů citací.

Autorům dnes pomáhají vytvářet záznamy citací také profesionální nástroje, jako jsou systémy **EndNote**, **ProCite** nebo nový systém **RefWorks** společnosti CSA (je nabízen i českým klientům). Některé webové citační služby řeší nedostatečné strukturování citací pomocí specifických algoritmů (CiteSeer, Google Scholar, Citebase Search aj.). Těmto službám by měly v současnosti pomoci nové přístupy pro **formalizaci citačních informací**. Slibné experimenty jsou realizovány pomocí metadatových formátů a schémat XML pro přenos citačních dat v síťovém prostředí (s využitím protokolu HTTP). Ke dvěma nejdůležitějším přístupům patří standard **OpenURL** a doporučení **Směrnice pro kódování bibliografických citačních informací** Metadatové iniciativy Dublin Core (DCMI). V obou přístupech jde o efektivní vytváření citačního propojování tak, že k citačním údajům je přidána specifická adresa URL.

V případě užití nového standardu **OpenURL** pro automatické kontextově citlivé propojování informací (standard USA Z39.88-2004) jde o velmi efektivní a perspektivní směr i pro citační propojování. OpenURL je nástrojem pro jednoznačné strukturování a přenos citačních metadat a identifikátorů za účelem jejich propojení s daným dokumentem. Standard definuje jednoznačnou syntax pro vytvoření souboru metadat a/nebo identifikátoru o nějakém informačním objektu. Uživatel, který při vyhledávání informací nachází záznam citovaného dokumentu, nabídne propojovací služba (link server) na základě identifikátoru OpenURL přístup k nejvhodnější elektronické kopii daného dokumentu. Podstatnými momenty fungování této technologie je tedy jednak existence propojovací služby (její adresa je součástí struktury celého identifikátoru OpenURL), jednak přítomnost identifikátoru OpenURL v bibliografických záznamech, které musí zajistit jejich producent. Technologie OpenURL je již běžně uplatněna v komerčních systémech a zdá se, že nic vážného nebrání jejímu nasazení také v oblasti systémů pro rozšiřování informačních materiálů s otevřeným přístupem.

DCMI v rámci svých čtených aktivit navrhla jiný způsob strukturování citačních informací. Na rozdíl od standardu OpenURL jde o strukturování ručního charakteru, které ale z druhé strany může být využito také v rámci OpenURL nebo jiných aplikací. Jde o doporučení **Směrnice pro kódování bibliografických citačních informací** (Guidelines for Encoding Bibliographic Citation Information), jejíž aktuální verze pochází z června 2005 [APPS, 2005].

Standard Dublin Core, který zahrnuje 15 základních prvků pro popis libovolného informačního zdroje, doposud nespécifikoval popis citace bibliografického zdroje. Speciální Pracovní komise DCMI pro citování proto připravila výše zmíněnou směrnici, která má povahu doporučení. Citační informaci (článek z časopisu, statě z konferenčního sborníku a popřípadě jiných materiálů) je doporučeno zapisovat v rámci kvalifikovaného Dublin Core. Jádro údajů o citovaném dokumentu je konkrétně doporučeno zapisovat v rámci jemnějšího prvku '**dcterms:bibliographicCitation**', který patří k základnímu prvku '**dc:identifier**'. Vlastní hodnota tohoto citačního prvku by měla obsahovat nezbytné minimum údajů (zapsaných podle syntaxe zvoleného citačního stylu), které citovaný dokument jednoznačně identifikují. Jinou možností pro zápis citací je využití specifického jemnějšího prvku '**dcterms:references**', který patří k základnímu prvku '**dc:relation**'. Pro potřeby strojového zpracování však doporučení předpokládá zápis všech údajů podle standardu **OpenURL**.

5 Závěr

Nárůst volně dostupných vědeckých dokumentů ukládaných do stále se rozrůstajícího velkého počtu centralizovaných elektronických archivů nebo institucionálních repozitářů pokračuje v globálním rámci rychlým tempem. Význam této literatury je značný až do té míry, že se

staly také předmětem zpracování významných komerčních citačních služeb WOS a Scopus. Jde však o služby drahé, dostupné jen na základě předplatného, které navíc zpracovávají jen vybranou množinu těchto volně dostupných informačních materiálů. Je proto potěšitelné, že vznikly a úspěšně se rozvíjejí nové, moderně koncipované volně dostupné citační služby, které těží z toho, že zpracovávané materiály včetně metadat jsou v jen digitální formě, že jsou uloženy v dobře organizovaných repozitářích (digitálních skladech) a že jsou volně dostupné. Zahrnutí úplných textů do zpracování v rámci těchto citačních služeb poskytuje, jak ukazuje analýza, mnohem větší možnosti při výstupním zpracování všech uložených informací. Konečnou cílovou službou jsou scientometrické, bibliometrické a informetrické analýzy, které budou pomáhat odhalovat vliv volně dostupných materiálů na vývoj vědy.

Hlavním problémem, kterým je nutné se v současnosti zabývat, je optimalizace tvorby záznamů citací, jejich strukturace pro automatické strojové zpracování a automatické propojování na samotné dokumenty. V současné době se záznamy citací získávají v systémech citačních služeb především pomocí specifických algoritmů metodou extrakce a rozkladu přímo z úplných textů, což není optimální (mohou nastat problémy s identifikací a chybovost může celkově velká). Jde o různorodé ručně zapisované záznamy v podobě nestrukturovaných textů a je úkolem programu, aby dokázal automaticky sjednotit různé podoby citací na stejný dokument.

Je tedy nezbytné, aby se v dalším období odborníci zaměřili na problematiku standardizace citačních záznamů, které jsou ukládány v elektronických archivech. V rámci zdokonalování softwarů, které zajišťují chod archivů či repozitářů, by měly být nasazeny specifické moduly, které by v blízké budoucnosti umožnily automatickou extrakci a reformátování citací (dle stanoveného standardu) z textů bezprostředně po jejich uložení v procesu auto-archivace tak, aby je jejich autoři měli možnost ve strojovém formátu zkontrolovat a případně okamžitě opravit. Znamenalo by to možnost jejich automatického sklizení a dalšího bezproblémového zpracování v databázích citačních rejstříků. Doporučovaným standardem pro strukturování citací je v současnosti standard OpenURL.

Použité informační zdroje

- APPS, A. 2005. *Guideliness for Encoding Bibliographic Citation Information in Dublin Core Metadata* [online]. Dublin : Dublin Core Metadata Initiative, Citation Working Group, 2005-06-13 [cit. 2006-04-20]. Dostupný z WWW: <<http://dublincore.org/documents/dc-citation-guidelines/>>.
- BRAND, A. 2004. CrossRef and the research experience. *Learned publishing*. 2004, vol. 17, no. 3, s. 225-230. doi:10.1087/095315104323159658, <http://dx.doi.org/10.1087/095315104323159658>
- BRATKOVÁ, E. 2001. Citace odborné literatury jako nástroj rozvoje služeb a integrace digitálních knihoven. In *AKP 2001 : Automatizace knihovnických procesů - 8 : Sborník z 8. ročníku semináře pořádaného ve dnech 24.-25. dubna 2001 v Liberci*. Praha : ČVUT, Výpočetní a informační centrum 2001, s. 109-120. Dostupný také z WWW: <<http://knihovny.cvut.cz/akp/clanky/12.pdf>>. ISBN 80-01-02-366-4.
- BRODY, T. 2003. Citebase Search : Autonomous Citation Database for e-Print Archives. In *Conference on Worldwide Coherent Workforce, Satisfied Users - New Services For Scientific Information - 17-19 September 2003, Oldenburg, Germany*. 2003. Dostupný také z ECS: <<http://eprints.ecs.soton.ac.uk/10677/>>.
- ESPECHE, N. 2006. *ISI Web of Knowledge : A Commitment to Innovation* [online prezentace]. Helsinky (Finsko), May 2006 [2006-05-2]. Dostupný z WWW <http://www.lib.helsinki.fi/finelib/koulutus/ISI_Finland%20Presentation2006-final%5B1%5D.ppt>.

- HARDY, R.; OPPENHEIM, Ch.; BRODY, T. aj. 2005. *Open Access Citation Information* [online]. September 2005 [cit. 2006-04-20]. Final Report - Extended Version. JISC Scholarly Communication Group. Ve formátu DOC. Dostupný také z ECS: <<http://eprints.ecs.soton.ac.uk/11536/>>.
- HARNAD, S. 2006. Publish or Perish - Self-archive to Flourish : The green Route to Open Access [online]. ERCIM News. January 2006, no. 64 [cit. 2006-04-20]. Dostupný z WWW: <http://www.ercim.org/publication/Ercim_News/enw64/harnad.html>.
- HARNAD, S.; BRODY, T. 2004. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine* [online]. June 2004, vol. 10, no. 6 [cit. 2006-04-20]. doi:10.1045/june2004-harnad, <http://dx.doi.org/10.1045/june2004-harnad>
- JACSÓ, P. 2004a. CiteBaseSearch, Institute of Physics Archive, and Google's Index to Scholarly Archive. *Online*. September/October 2004, vol. 28, no. 5, s. 57-60. Dostupný také z: WWW: <<http://projects.ics.hawaii.edu/~jacso/PDFs/jacso-CiteBase-InsPhysics-GS-28-5.pdf>>. ISSN 0146-5422.
- JACSÓ, P. 2004b. *Web of Science Citation Indexes* [online]. August 2004 [cit. 2006-04-28]. Gale – Reference Reviews. Péter's Digital Reference Shelf. Dostupný z WWW: <<http://www.gale.com/>>.
- JACSÓ, P. 2005a. As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current science*. November 2005, vol. 89, no. 9, s. 1537-1547. Dostupný také z WWW: <<http://www.ias.ac.in/currensci/nov102005/1537.pdf>>. ISSN 0011-3891.
- JACSÓ, P. 2005b. *CiteSeer* [online]. November 2005 [cit. 2006-04-28]. Gale – Reference Reviews. Péter's Digital Reference Shelf. Dostupný z WWW: <<http://reviews.gale.com/index.php/digital-reference-shelf/2005/11/citeseer/>>.
- LAGOZE, C. aj. 2004. *The Open Archives Initiative Protocol for Metadata Harvesting* [online]. Protocol Version 2.0 of 2002-06-14, Document version 2004/10/12T15:31:00Z. 2004 [cit. 2006-04-20]. Dostupný z WWW: <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>.
- MARTELLO, A. 2006. *Selection of Content for the Web Citation Index™ : Institutional Repositories and Subject-specific Archives* [online]. [2004, aktualiz. 2006] [cit. 2006-04-28]. Dostupný z WWW: <<http://scientific.thomson.com/free/essays/selectionofmaterial/wci-selection/>>.
- NEUHAUS, Ch. aj. 2006. The Depth and Breadth of Google Scholar : An Empirical Study. *Portal : libraries and the academy* [online]. April 2006, vol. 6, no. 2, s. 127-141 [cit. 2006-04-20]. Dostupný z WWW: <http://muse.jhu.edu/journals/portal_libraries_and_the_academy/toc/pla6.2.html> ISSN 1530-7131
- SUBER, P. 2006. *Open Access Overview : Focusing on open access to peer-reviewed research articles and their preprints* [online]. First put online, June 21, 2004, Last revised March 10, 2006 [cit. 2006-04-20]. Dostupný z WWW: <<http://www.earlham.edu/~peters/fos/overview.htm>>.