

Univerzita Karlova
Filozofická fakulta
Ústav informačních studií a knihovnictví

Studijní program Informační studia a knihovnictví
Studijní obor: informační studia a knihovnictví

Ludmila Hercová

Citační analýza a komunikace vědeckých znalostí

Rigorózní práce

konzultant Prof. PhDr. Marie Königová, CSc.

Praha 2005

Prohlášení:

Prohlašuji, že jsem rigorózní práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje.

V Praze, 30. listopadu 2005

Lucie Kucová

podpis rigorózanta

Identifikační záznam

HERCOVÁ, Ludmila. *Citační analýza a komunikace vědeckých znalostí*. Praha, 2005. 214 s., 31 s. příl. Rigorózní práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví, 2005. Konzultant rigorózní práce Prof. PhDr. Marie Königová, CSc.

Abstrakt

Tématem práce je citační analýza, citace a vědecká komunikace v tištěném a elektronickém prostředí. Práce se zaměřuje na problematiku bibliometrie, infometrie, citační analýzu, definuje pojem impakt faktor. Stručně popisuje prostředí Web of knowledge. Práce se pojednává o kritice citací, autocitacích a odměňování ve vědě a zaměřuje se dále na citační rejstříky jako jsou Taiwan Humanities Citation Index, Scopus, Index Copernicus. V práci je zpracována oblast patentových citací a také problematika citační analýzy v elektronickém prostředí – webometrie.

Klíčová slova

citační analýza, citační rejstříky, citační indexy, bibliometrie, infometrie, webometrie, impakt faktor, immediacy index, cited half-life, citing half-life, Journal Citation Reports, Web of Knowledge, Web of Science, autocitace, Matthewův efekt, Taiwan Humanities Citation Index (THCI), Scopus, Index Copernicus, patentové citace

Obsah

1.	Předmluva	6
2.	Úvod.....	8
3.	Bibliometrie	9
3.1.	Vymezení pojmů bibliometrie, scientometrie a infometrie.....	9
3.2.	Bibliometrie.....	9
3.3.	Bibliometrické indikátory	10
3.4.	Bibliometrické zákony	12
3.4.1.	Lotkův zákon	12
3.4.2.	Bradfordův zákon	12
3.4.3.	Zipfův zákon	12
3.5.	Empirické zákony	12
3.5.1.	Některé případy hyperbolického rozšíření	13
3.5.2.	Weber-Fechnerovy psycho-fyzikální zákony	13
3.5.3.	Paretova rozložení.....	14
3.5.4.	Wilisova taxonomická rozložení.....	14
3.5.5.	Lotkova rozložení literární produktivity	14
3.5.6.	Bradfordův zákon literární produktivity	14
3.5.7.	Zipfův zákon	15
3.6.	Použití a omezení koeficientu Giniho.....	20
3.7.	Giniho poměr mezi populacemi	22
3.8.	Příklad: Bradfordovy originální sady dat	22
3.9.	Porovnání dvou měření	23
3.10.	Dvě Paretovy distribuce	23
4.	„Nová“ a „stará“ infometrie aneb infometrie se stále vyvíjí	29
4.1.	Racionálnost formule.....	31
4.2.	Konkrétní výpočet T/TT průměru μ^* ve funkci T/T průměr μ a Lotkovo α	35
5.	Budoucnost citačního indexování.....	39
6.	Citační analýza.....	41
6.1.	Vymezení pojmu citační analýza a využití citační analýzy.....	41
6.2.	Doba oběhu znalostí.....	42
7.	Impakt Faktor (IF; faktor vlivu).....	47
7.1.	Další indikátory počítané ISI	48
7.1.1.	Immediacy Index	48
7.1.2.	Cited Half-life	49
7.1.3.	Citing Half-life	49

7.2.	Citační rejstříky Institute for Scientific Information	49
7.2.1.	Journal Citation Reports (JCR).....	49
7.2.2.	Citační rejstříky	50
7.2.3.	Struktura SCI	51
8.	Web of Science (Web of Knowledge).....	52
8.1.	Web of Knowledge.....	52
9.	Kritika citační analýzy a autocitace	60
10.	Modely autocitací v publikovaném výzkumu	64
10.1.	Výzkumná spolupráce: modely spoluautorství	64
10.1.1.	Vůdci a následovníci: citační modely	65
11.	Užití a zneužití impakt faktoru	69
11.1.1.	Dovoluje impakt faktor hodnotit kvalitu jednotlivých autorů?	69
12.	Odměňování ve vědě	71
12.1.	Systém odměňování a fenomén „okupantů 41. křesla“	71
12.2.	Matthewův efekt.....	71
12.2.1.	Matthewův efekt v systému odměn.....	71
12.2.2.	Sociální a psychologické základy Matthewova efektu.....	72
12.3.	Hromadění výhod a nevýhod v rámci vědeckých institucí.....	73
12.4.	Od výzkumu struktury světové vědy k objevení Matthewova efektu pro země 73	
12.4.1.	Mikrostruktura Matthewova efektu pro země	74
12.4.2.	Matthewovy citace a Matthewovy časopisy jádra.....	75
12.4.3.	Typologie Matthewových časopisů jádra	79
13.	Webometrie.....	81
13.1.	Webometrie, bibliometrie a infometrie.....	83
13.2.	Cybermetrie	84
13.3.	Terminologie a diagramy webovských uzlů.....	85
13.4.	Základní terminologie webovského uzlu a diagramy.....	88
13.5.	Hypotéza shluku odkazů	97
13.6.	Rozklad obsahu podobnosti.....	100
13.7.	Různorodost obsahového rozkladu.....	100
13.8.	Hypotéza shluku odkazů	102
13.8.1.	Rozklad předpokládané relevance.....	104
13.8.2.	Implikace pro předmětné webovské crawlery	104
13.9.	Studie týkající se daného problému	107
13.10.	Výzkumné otázky	108
13.11.	Online nástroje pro sběr dat.....	108

13.12.	Kategorie pro klasifikování obsahu webovské stránky	109
13.12.1.	Klíčový faktor 1: impakt časopisu v rámci vědecké komunity.....	109
13.12.2.	Klíčový faktor 2: úroveň obsahu webovské stránky časopisu	110
13.12.3.	Klíčový faktor 3: věk webovské stránky časopisu	110
13.13.	Terminologie webovského prohledávání.....	112
13.14.	Obsahové prolézání versus prolézání odkazů	114
13.15.	Spolehlivost a platnost webovského prolézání.....	115
13.15.1.	Experiment.....	117
13.15.2.	Metoda.....	117
14.	Další citační rejstříky	121
14.1.	Taiwan Humanities Citation Index (THCI)	121
14.1.1.	Cíle projektu THCI:	121
14.1.2.	Vytvoření THCI	122
14.1.3.	Kritéria pro indexaci v THCI:	122
14.1.4.	THCI struktura.....	124
14.1.5.	Struktura THCI	125
14.1.6.	Politika a pravidla THCI	125
14.1.7.	Aplikace THCI.....	125
14.1.8.	Vyhledávací nástroje.....	126
14.1.9.	Těžkosti a problémy.....	130
14.1.10.	THCI projekt.....	130
15.	SCOPUS	131
15.1.	Scopus – vlastnosti a funkce.....	136
15.1.1.	Vyhledávání:	136
15.1.2.	Zobrazení výsledků.....	137
15.1.3.	Propojení na plný text	137
15.2.	Odkaz na vyhledávací historii	137
15.3.	Prohledávání citací a odkazů	138
15.4.	Export výsledků.....	138
15.5.	Pokrytí předmětu ve Scopusu	138
15.6.	Přidávání toho nejlepšího z webu:	138
15.7.	Odkazy k plným textům přes CrossRef a další	139
15.8.	Odkazy na plné texty přes analyzátor odkazů.....	139
15.9.	Jak obrazově založené odkazování pracuje	139
15.10.	Scopus – knihovní integrace	139
15.11.	Pokrytí oborů z oblasti umění a humanitních oborů	140

15.12.	Pokrytí Scopusu v oblasti sociálních věd	140
15.13.	Geografické rozložení publikací ve Scopusu	141
15.14.	Vyhledávání	141
15.14.1.	Sort by relevance (třídění podle relevance)	141
15.14.2.	Analyzátor odkazů	142
15.14.3.	Scopus.....	142
16.	Index Copernicus	143
16.1.	Hodnocení časopisů pro Index Copernicus.....	145
16.1.1.	Total-Basic Score (TBS – celkové základní skóre)	145
16.2.	Vědecká kvalita.....	145
16.2.1.	Vydavatelská kvalita	146
16.2.2.	Cirkulace/dostupnost	148
16.2.3.	Frekvence-pravidelnost-stabilita na trhu	148
16.2.4.	Technická kvalita	149
16.3.	Metodologie bodování Index Copernicus (Index Copernicus Score – ICS) celkové výsledky	149
17.	Patentové citace.....	150
17.1.	Přirozenost citací.....	151
17.2.	Užitečnost a relevance.....	151
17.3.	Jazyk patentových citací	153
17.4.	Budoucnost	154
17.5.	Předešlé patentové citace.....	158
Průměr.....		158
17.6.	Modely publikačních efektů patentových citací	161
17.7.	Přesahy znalostí: patentová „stopa“	166
17.8.	Geografická distribuce citací	168
17.9.	Struktura „zpoždění citací“ mezi citujícími a citovanými patenty	168
17.10.	Citované patenty v rámci jedné firmy/ v rámci jednoho oboru průmyslu ..	170
17.11.	Mohou být citace užitečné?.....	173
17.12.	Některé skutečnosti o citacích	174
17.13.	Tržní hodnota a patenty	176
17.14.	Kdy citace jsou informace?	177
17.15.	Zjištění	182
17.16.	PATCITES – vyhledává soubory s patentovými citacemi.....	199
18.	Průzkum evropských lékařských časopisů.....	203
19.	ZÁVĚR.....	205
	Slovník pojmů a zkratk	206

Seznam použité literatury	209
Evidence výpůjček	214
Příloha ke kapitole 12: Průzkum - Údaje z Journal Citation Reports.....	215
Počet časopisů evropských zemí v Journal Citation Reports	215
Časopisy Polska, České republiky, Slovenska a Maďarska v JCR	217

1. Předmluva

Téma citační analýzy a vědecké komunikace znalostí jsem zvolila jako pokračování diplomové práce obhájené v září 2003 na Ústavu informačních studií a knihovnictví. Předkládaná práce je v některých částech jejím rozšířením, aktualizací a zaměřuje se také na nové oblasti, které se tohoto tématu týkají. Tuto problematiku jsem se zájmem sledovala několik let. Z důvodu rozsáhlosti tématu bylo nutné se v některých oblastech zaměřit na dané téma jen stručně, neboť jednotlivé části by mohly být také zpracovány samostatně jako jednotlivé práce.

Jedná se o téma, o kterém se v posledních letech hodně píše v souvislosti s otázkou, zda je správné hodnotit instituce a vědce na základě citační analýzy a impakt faktoru. U nás tím více, že impakt faktor mnoho českých časopisů nemá, a tak pokud chtějí čeští autoři publikovat v prestižním časopise, musejí svou práci nabízet v zahraničních redakcích. S tím se také objevila např. v rámci lékařské komunity otázka, zda české časopisy mají své opodstatnění - zda je vydávat a pokud vydávat, tak v jakém jazyce.

Toto téma jsem si také vybrala, protože se problematikou citační analýzy na teoretické úrovni u nás hlouběji nikdo nezabývá, přestože je to téma je zajímavé a neustále se objevují nová a nová rozšíření jako např. webometrie.

Když otevřeme naše odborné časopisy, nalezneme obvykle stejný model – představení, co je impakt faktor, jak se počítá a k čemu slouží a případně nejlepší časopisy z daného oboru seřazené podle impakt faktoru. Ale již se nikdo příliš nezabývá např. výzkumem a možnými novými způsoby hodnocení časopisů, popřípadě vědců na základě jejich citovanosti. Určitou výjimkou je český citační rejstřík lékařské literatury, který je počítán sice pro české časopisy (zatím jen tři), ale po vzoru impakt faktoru Institute for Scientific Information (ISI). Tento rejstřík se bohužel nyní dále nerozvíjí a pokud se nenaleznou peníze na jeho další rozvoj, zbyde z něj jen torzo, což by byla samozřejmě škoda.

Cílem mé práce je stručný popis vzniku a vývoje citační analýzy, vymezení souvisejících pojmů a popis současného vývoje v této oblasti. Toto téma zahrnuje mimo jiné i aktivity spojené s vytvářením nových faktorů hodnocení vědy a také citační rejstříky - i jiné než rejstříky vytvářené Institute for Scientific Information. Také se soustředím na rozšíření citační analýzy na prostředí webu – webometrii a v jedné části se také zaměřím na patentové citace. Dalším cílem práce je i analýza současných problémů spojených s hodnocením vědeckých časopisů, institucí, vědců a zemí.

Metodou průzkumu jsem se zaměřila na evropské medicínské časopisy, na jejich zastoupení v polském citačním rejstříku Index Copernicus a na databázi časopisů excerpovaných Institute for Scientific Information Journal Citation Reports. Zkoumala jsem zastoupení jednotlivých zemí a zastoupení nejlepších evropských časopisů v daných rejstřících.

Pro svou práci jsem vyhledávala literaturu z databází a informačních produktů EBSCO, Science Direct, Medline, EMBASE, BiblioMedica, Web of Science a z jeho nadřazeného produktu Web of Knowledge. Hodně jsem sledovala časopis **Journal of the American Society for Information Science and Technology**, který v podstatě neustále přináší nové a nové články, které se daného tématu týkají. Témeř každý měsíc se objeví nový článek, který se daného tématu týká.

Problematika citační analýzy neboli studia citací, impakt faktoru a modelů citací v oblasti vědy je velmi široká a pokud by někdo chtěl tuto oblast obsáhnout celou,

jednalo by se o studium stovek až tisíců pramenů po dlouhou dobu a nikoliv v rozsahu rigorózní práce. Samozřejmě jsem nemohla obsáhnout celou škálu problémů a ani to nebylo účelem této práce. Především jsem se nezabývala (nebo jen v náznacích) problematikou citační analýzy v prostředí Internetu. Myslím, že toto téma by vydalo na samostatnou práci.

Ve světě se rozvíjí i další citační rejstříky, např. Tchajwanský citační rejstřík pro oblast sociálních a humanitních věd a také další, které ovšem ve své práci nezmiňuji, ani blíže nepopisuji. Problémem u některých může být i např. jazyková bariéra, specificky např. u čínských citačních rejstříků, případně i nedostatek informací a přístupu k daným rejstříkům. Namátkou uvedu např. polský citační rejstřík pro oblast sociálních věd (Polish Sociology Citation Index). Snahou tvůrců čínského citačního indexu (Chinese Science Citation Database, Chinese Scientific and Technical Papers and Citation Database; CSTPC) je mapovat a hodnotit rozsáhlou neanglickou vědeckou literaturu. Tyto rejstříky jsem nepopisovala z důvodů jazykových (u čínského citačního rejstříku), z důvodu jejich nedostupnosti, a u polského rejstříku též z důvodu jeho zaměření na oblast sociálních věd. Více nerozvádím citační rejstříky ISI. Popisují jen Science Citation Index. Další rejstříky Social Science Citation Index a Arts & Humanities Citation Index jen zmiňuji, protože oblast medicíny nesledují a také se jedná o velmi známé rejstříky, které není potřeba představovat. Případní zájemci si mohou potřebné informace snadno získat i díky přístupu řady našich knihoven do těchto rejstříků.

Chtěla bych poděkovat všem, kteří mi s touto prací pomáhali a podporovali mě při jejím vzniku. Můj velký dík patří vedoucí této práce profesorce Marii Königové, která nás své studenty s touto zajímavou problematikou seznámila již na začátku našeho studia knihovnictví a informační vědy na vysoké škole, i když v rámci jejích předmětů tato problematika nemohla být obsáhle probírána. Ale její přístup k přednáškám, kdy se nás snažila zasvětit i do pro nás ne vždy oblíbených matematických problémů nebo jiných témat, hlavně v počátcích studia neznámých problémů informační vědy způsobily, že snad každý student si ji velmi oblíbil. Především ti z nás, kteří navštěvovali seminář o kreativě se na jednotlivé hodiny těšili a tento seminář doporučovali ostatním.

Také bych ráda poděkovala mé rodině a příteli, kteří mi velmi pomohli s formální podobou a vydrželi se mnou těch několik měsíců, kdy jsem se konečně připravě této práce věnovala.

2. Úvod

Tématem této práce je citační analýza. Jedná se o dosti široké téma, které se dotýká jak knihovnictví a informační vědy, tak mnoha ostatních oborů a věd. Přirozeností vědců a výzkumníků je publikovat výsledky, kterých dosáhli a problémů, se kterými se potýkali. Neustále vycházejí nové a nové publikace, které se touto problematikou zabývají. Není jednoduché nalézt články týkající se této oblasti v české literatuře. Nevytvářejí se příliš databáze, ať již oborové nebo univerzální, které by zachycovaly podrobněji především časopiseckou literaturu, která na území České republiky vychází. Výjimkou je snad jedině databáze BiblioMedica vytvářená Národní lékařskou knihovnou. Tato databáze sleduje nejen české časopisy jako celek, nýbrž také jednotlivé články v těchto časopisech.

Je však problém přesně dané články dohledat, jelikož podle názvu se slovy "citační analýza", "impakt faktor" (případně dalšími), je článků jen několik a tak jsme odkázáni na předmětové vyhledání a ani to není nejpřesnější. Pokud vědec hledá, kdo se daným problémem u nás zabývá, nezbyvá mu než sledovat průběžně časopisy daného oboru nebo několik ročníků časopisů prolistovat.

3. Bibliometrie

3.1. Vymezení pojmů bibliometrie, scientometrie a infometrie

Před vymezením pojmu citační analýza je nutné jej zařadit do širších souvislostí. Nadřazeným pojmem pro pojem citační analýza je **bibliometrie** a nad ním stojí další nadřazený pojem, **infometrie**.

Infometrie je kvantitativní metoda publikačních a citačních modelů.

V SSSR v 30. letech 20. století byla nezávisle vytvořena metoda nazývaná jako naukomatika, která v západních zemích byla nahrazena třemi rozdílnými pojmy: **bibliometrie**, **scientometrie** a **infometrie**.

Bibliometrie se nejvíce zabývá studiem tištěné literatury.

Pojem **scientometrie** je používán pro studium rozvoje fyzikálních a biologických věd.

Infometrie je obecnější termín, který pokrývá scientometrii a bibliometrii.¹

3.2. Bibliometrie

Podle A. Pritcharda je cílem bibliometrie vnést světlo do procesu psané komunikace, nalézt směr a způsoby počítající a analyzující různé fasety psané komunikace. Bibliometrie je statistický popis literatury – skupin dokumentů, které mají mezi sebou určitý vztah. To dovoluje, aby literatura byla popsána a monitorována.

Existují dva typy bibliometrických studií.

- První typ je známý jako popisné studie, které charakterizují hlavní vlastnosti literatury nebo publikací. To znamená, že vytvářejí soubor informací.
- Druhý typ bibliometrických studií zkoumá chování a bibliografické vztahy mezi publikacemi. Ty jsou zkoumány pomocí citační analýzy. Lze je nazvat studiem prací, ve kterých autoři citují souhrn literatury nebo jednotlivé publikace. Tyto studie ukazují vztah mezi různými autory a souhrny literatury.²

Bibliometrické indikátory popisují vědy a technologie v oblasti produkce a toku znalostí na mezinárodní, národní, regionální, subregionální a institucionální úrovni.

Objektivním cílem bibliometrie a scientometrie je poskytovat tvůrcům veřejné politiky návody k použití aktuálně dostupných bibliometrických nástrojů. Existuje mnoho aplikací bibliometrie jako prostředku pro hodnocení programů vědy, technologie, ovlivňování chování soukromých a veřejných výzkumů nebo pro technologické monitorování a strategické rozhodování.

¹ BRIN, Beth; COCHRAN, Elissa, Access and ownership in the academic environment: one library's progress report.

GISVOLD, Sven-Erik, Citation analysis and journal impact factors – is the tail wagging the dog?.

SHERRI, Edwards, Citation analysis as a collection development tool : a bibliometric study of polymer science theses and dissertation.

KARLSSON, Fred, Linguistics in the light of citation analysis.

² THIESEN, Richard, A bibliometric study of Direction.

Bibliometrické metody slouží tedy třem hlavním funkcím - popisu, hodnocení a vědeckému a technologickému monitorování. Jako nástroj popisu poskytuje bibliometrie množství publikačních aktivit na úrovni zemí, provincií, měst nebo institucí a je užívána pro srovnávací analýzy produktivity vědy. Data mohou být poté použita pro odhad chování výzkumných celků. Bibliometrická data jsou také používána pro monitorování vědy a technologie, pro hodnocení vědeckého výstupu, pomáhají identifikovat oblasti vědy, které jsou rozvíjeny nebo které upadají.

3.3. Bibliometrické indikátory

Bibliometrické indikátory poskytují pouze obraz vědeckého výstupu země. Status bibliometrie je popsán Godinem takto: "Může nastat doba, kdy skutečnost, že bibliometrické indikátory byly standardizovány, může omezit jejich užitečnost. Mají svá omezení, protože normálně zahrnují jen přírodní vědy, inženýrské obory a biomedicínské vědy. Existují ale také zřejmé lingvistické zákonitosti, které širokou měrou omezují pokrytí vědeckého výstupu publikovaného v angličtině. Konečně je nutné si uvědomit, že publikování reprezentuje jen jednu z aktivit výzkumníků. Navzdory takovému omezení jsou bibliometrické indikátory jedním z hlavních nástrojů pro měření výzkumného výstupu a z toho důvodu si představují významné místo ve vědeckých a technologických postupech."³

Vědecká komunikace je nadmíru důležitá pro porozumění vzniku a vývoje disciplín. Od vzniku organizovaných akademických aktivit se jednotlivé disciplíny staly sférami individuální a kolektivní aktivity ve vědě, technologii, humanitních a sociálních vědách. Společnosti, specializované výzkumné instituce, univerzitní obory a konference usilují o pokračující sociální formaci těchto aktivit. Publikační aktivity zahrnují jak monografie, tak časopisy. Jádrem vědeckého publikování však tvoří především časopisy. Časopisy představují významný prostředek komunikace mezi vědci a výzkumníky, kteří svými články na stránkách časopisů formují další vývoj vědních disciplín.

Jakákoliv publikace může být zkoumána bibliometricky, z článku časopisu, konferenčního materiálu, technické zprávy či monografie - tedy všude tam, kde najdeme dostatek bibliografických údajů. Často je to ve formě bibliografie, ale nejčastěji jsou to údaje stanovující primární data, která mají být zkoumána. Autoři, články nebo monografie, rok publikování a název časopisu jsou zahrnuty ve spektru informací, které mohou být dále zkoumány.⁴

První hlavní práce o rozšíření citační analýzy jako objektivního měření hodnoty časopisu byla napsána pány Grossem a Grossem v roce 1927. V této průkopnické studii autoři hledali objektivní metodu výběru časopisů pro hodnocení chemického výzkumu. Autoři se rozhodli dát do tabulky všechny citované odkazy v jednom ročníku *Journal of the American Chemical Society*. Celkový počet citací byl 3 633 a bylo zjištěno, že se nacházejí v 247 rozdílných časopisech.⁵

³ MCMILLAN, G. Steven; HAMILTON, Robert D., Using bibliometrics to measure firm knowledge : an analysis of the pharmaceutical industry.

⁴ DING, Ying; CHOWDHURY, Gobinda; FOO, Schubert, Incorporating the results of co-word analyses to increase search variety for information retrieval.

⁵ SHERRI, Edwards, Citation analysis as a collection development tool : a bibliometric study of polymer science theses and dissertation.
ROMANO, Claudio; RATNATUNGE, Janek, Citation analysis of the impact of journals on contemporary small enterprise research.

Koncept citačního indexování je jednoduchý. Vychází z toho, že hodnota informace je vymezena těmi, kdo informaci využívají. To představuje pravděpodobně lepší způsob měření kvality prací, než je např. měření kvality vlivu, např. na základě **impakt faktoru**. Nejširší vědecká populace v rámci vědecké komunity - to je kdokoliv, kdo užívá nebo cituje zdroje materiálů - vymezuje vliv dané myšlenky a jejího tvůrce.

V minulosti existovaly tři faktory, které vedly k rozvoji citačního indexování v 50. letech 20. století. Po druhé světové válce, kdy docházelo k velkému toku vládních peněz do oblasti výzkumu, začala výzkumná komunita přirozeně publikovat dokumenty prostřednictvím uznávaných kanálů vědecké časopisecké literatury. Objevila se potřeba nalezení nové metody indexování a vyhledávání, která by byla efektivnější z hlediska nákladů a účinnější, než tehdejší modely lidského indexování dokumentů v rámci vědeckých indexů.

Indexování vytvořené jedinci bylo propracované a hodnotné z hlediska hloubky daného předmětu. Takovéto indexování však bylo velmi nákladné a časově náročné - náklady se zvyšovaly přímoúměrně s počtem dokumentů.

Dr. Eugene Garfield, zakladatel **ISI (Institute for Scientific Information)**, se účastnil výzkumu, který se zabýval strojovým vytvářením indexů v polovině 50. a na počátku 60. let 20. století. Jeden z projektů, kterých se zúčastnil, Welch Medical Indexing Project na Johns Hopkins University v USA (byl financován tehdejší Army Medical Library – předchůdkyně National Library of Medicine), se zabýval rolí automatizace v organizování a vyhledávání lékařské literatury.

E. Garfield brzy pochopil, že přehled článků v časopisecké literatuře bude spoléhat na bibliografické citace, které odkazují čtenáře na originální publikovaný zdroj. Shromážděním těchto citací, jak Garfield věřil, by výzkumník mohl okamžitě získat přehled o konceptu vytvořeného jiným vědcem jako podklad myšlenky nebo metodologie založené na zdrojích, které vědec publikoval, konzultoval a citoval jako pertinentní ve své bibliografii. Termíny vyhledávání vybrané z citace mohly fungovat stejně dobře jako klíčová slova a deskriptory, které byly pečlivě vytvořené profesionálními indexátory.

Eugene Garfield vytvořil multidisciplinární citační index – **Science Citation Index**, publikovaný od roku 1963. Jeho dnešní webovsky založená verze – **Web of Science**, dnes již spíše známým pod názvem **Web of Knowledge** - pokrývá přibližně 5 600 časopisů vydávaných v rámci více než 150 vědeckých disciplín.⁶

Garfieldův úspěch je třeba přičíst vytvoření užitečného objektivního citačního indexu (respektive indexů) identifikujících příbuzné práce v publikované literatuře, která se nemusí vždy jevit jako pertinentní zdroj vědeckého publikování. Lze říci, že i přes mnohé výhrady, které budou ještě později prezentovány, je dnes brán jako nejspolehlivější zdroj v propojování myšlenek a mapování vztahů v rámci velkého množství disciplín.⁷

⁶ GARFIELD, Eugen, How it all began – with a loan from HFC.
GARFIELD, Eugen, Current Contents Ninth Anniversary. Essays of an Information Science.

⁷ HERCOVÁ, Jana, Dr. Eugen Garfield - 75. narozeniny.

3.4. Bibliometrické zákony

3.4.1. Lotkův zákon

Lotkův zákon popisuje, že největší počet publikací v jakémkoliv poli je produkován jen malým počtem vysoce produktivních autorů. Jestliže 100 autorů publikuje práce v konkrétní oblasti, pak 25 z nich bude publikovat dvě práce, 11 tři práce a 6 čtyři práce.⁸

$$a_n = K / n^r$$

kde:

a_n Je počet autorů s n publikacemi,

K je konstanta,

n je počet publikací,

r je konstanta.

3.4.2. Bradfordův zákon

Bradfordův zákon popisuje způsoby, jakými jsou práce distribuovány v časopisech.

Pokud časopisy rozdělíme do tří oblastí:

- 1) časopisy daného oboru,
- 2) časopisy příbuzných oborů,
- 3) ostatní časopisy,

zjistíme, že počet časopisů v těchto oblastech je úměrný vztahu $1 : n : n^2$.

3.4.3. Zipfův zákon

Zipfův zákon je velmi známý i mimo bibliometrii. Popisuje distribuci slov v textu. Některá slova bývají použita mnohokrát (např. tzv. stop slova), mnoho slov bývá použito jen jednou.⁹

Předmětem Zipfových výzkumů byla kniha *Odysseus* autora J. Joyce. Zipf zjistil, že v knize je 29 899 různých slov. Tato slova uspořádal podle frekvence (f) výskytu a slovo s hodnotou dostalo rank 1 (r), další slovo rank 2 atd. Když vynásobil rank každého slova s jeho frekvencí zjistil, že výsledné hodnoty jsou velmi blízké (c).

$$r * f = c$$

Této znalosti se mimo jiné využívá pro vytváření frekvenčních slovníků.

3.5. Empirické zákony

Sociální fenomén může mít mnoho rozličných příčin, které jsou samostatné nebo se vyskytují dohromady. Dohromady s jednoznačností sociálních jevů, což je činí neopakovatelné. Každodenní boj s těmito fakty nutí ekonomy, inženýry, administrátory, knihovníky a další lidi ve službách k opatrnému zacházení s těmito sociálními nároky.

⁸ Infometrics.

⁹ Infometrics.

Je velmi obtížné předvídat změny tak, aby bylo smysluplné vzít je v úvahu. Ale ani úspěšné proroctví negarantuje konečný úspěch. Kvalifikovaná předpověď vyžaduje pečlivé a rozsáhlé sledování. Nepožaduje jen dovednosti, trpělivost a vytrvalost, ale také inspiraci. Poté musí být data racionálně popsána. Popis také potřebuje inspiraci, protože způsob popisu je vědomě nebo nevědomě jen dílčím vysvětlením.

Vzácné události, objevující se nezávisle na sobě s konstantní rychlostí, jsou společné v mnoha jinak nesouvisejících aktivitách. Poissonova rozložení může být použito pro předvídaní. Jistá třída empirických popisných „zákonů“ byla dlouho známa jako vhodná jistému fenoménu v psychologii, meteorologii, lingvistice, ekonomice a pravděpodobně jakémukoliv dalšímu fenoménu s nezbytnými formálními vlastnostmi. Mezi nimi jsou jisté typy bibliografického chování bez většího efektu. Zájem o tuto oblast se objevuje asi s rostoucími požadavky na knihovny a příbuzné služby. Daleko více lidí pracuje v oblasti knihoven a dokumentaristiky než dříve. Více počítají, více čtou. Dříve nebylo obvyklé zabývat se touto oblastí z matematického hlediska.

Goffman a Newil aplikovali svůj "epidemický" model rozložení vědeckých myšlenek. Komunikace myšlenek pomocí zaznamenaného diskurzu je předkládána jako určující proces pro formálně ekvivalentní komunikaci, nemoci, mikroorganismy a signály šířené stroji. Informační vyhledávání poté odpovídá procesům rozšiřujícím „infekci“ díky kontaktům citlivějších osob s infekcí. Goffman a Newil se zabývali studiem růstu, úpadku a stability takovýchto procesů.

3.5.1. Některé případy hyperbolického rozšíření

Teorie Goffmana a Newila se zdá být jen bibliografickým modelem, který zahrnuje jednoznačný čas. Empirický vztah je mezi podnětem a reakcí sledovaný Weberem a mezi skenovanými časopisy a články vyhledanými Bradfordem.

Hyperbolická rozložení pro diskretní a spojitě proměnné se nacházejí v mnoha oblastech a jsou pojmenována různě dle Fechnera, Zipfa, Pareta, Bradforda, Wilise, Gerger-Mandelbrota a dalších. Jsou pojmenována ve spojitosti s jednotlivými vysvětleními pro jednotlivé manifestace tohoto typu chování, ne pojmenované pro typ chování samotný.

3.5.2. Weber-Fechnerovy psycho-fyzikální zákony

V roce 1846 Weber experimentálně zjistil, že zvýšení stimulu je úměrné počáteční hodnotě podnětu. To implikuje fakt, že podnět se musí zvyšovat v geometrické posloupnosti k výsledku aritmetického zvýšení v reakci.

Fechner (1860) to rozvinul jako podporu jeho filozofické pohledu na vztah mysl-tělo. Měl za to, že Weberovo pravidlo by platilo pro pododdíly lépe. V omezení to vede k:

$$R = k \log (S / s)$$

Kde:

R je počet právě pozorovatelným přírůstkům reakce nebo vjemu k dané hodnotě od počátku,

S je veličina podnět,

s je počáteční hodnota.

Fechner toto nazýval Weberovým zákonem. Formálně je stejný jako Bradfordovo rozložení.

3.5.3. Paretova rozložení

Ekonom Pareto (1897) ukázal, že rozložení bohatství v různých komunitách je reprezentováno podílem $P(u)$ z příjmů rovno k nebo převyšující danou hodnotu u jako obráceně úměrné ke stejné mocnině z u , exponent je větší než nula nebo roven jedné:

$$P(u) = k \cdot u^{-\theta} \quad (1 \leq u, 0 < \theta \leq 1) \\ = 1 \quad (0 \leq u < 1)$$

3.5.4. Wilisova taxonomická rozložení

Willis zjistil, že pro některé biologické rodiny počet genera $g(s)$ má přesně s druhů a mohly by být dány vztahem:

$$g(s) = k \cdot s^{-(1+\theta)} \quad (0 < \theta \leq 1)$$

Alternativně, vzato s jako spojitá proměnná, počet genera $G(s)$ mající alespoň s tříd je:

$$G(s) = K \cdot s^{-\theta}$$

k , a K jsou přiměřené konstanty.

Benoit Mandelbrot zpracovával jeho distribuci z mnoha úhlů pohledu v roce 1955.

Základem je příklad populace utvářené více obecnými distribucemi. Mají mnoho charakteristik z taxonomie a lingvistiky a také z dalších oborů. Je stabilní ve smyslu, že části jsou distribuovány stejným způsobem. Toto rozštěpení nebo shromáždění genera nebo tříd výsledku v populacích se řídí stejným zákonem a vztah mezi rodinami a generou je stejný jako mezi generou a třídami. Později Mandelbrot rozvinul charakteristiku stability nebo podobnosti sama sebe a aplikoval to v mnoha fenoménech, včetně obměn v cenách trhu a občasných chyby vzniklé v komunikačních systémech. Jeho studie jsou velmi relevantní pro bibliometrii.

3.5.5. Lotkova rozložení literární produktivity

Lotka zjistil, že počet autorů, kteří publikovali n prací v daném oboru bylo hrubě $1/n^2$ počtu autorů, kteří publikovali jen jednu práci. Pro všechny autory v daném oboru by jejich počet byl konečný a méně než $\pi^2 / 6$, přibližně 1.65 počet autorů s jedním článkem, dokonce jestliže by celkový výstup byl nekonečný. To je ovšem nepravděpodobné. Vztah podceňuje počet více plodných autorů, ale hodí se dobře pro méně plodné.

3.5.6. Bradfordův zákon literární produktivity

V roce 1934 Bradford publikoval studii rozptylu článků o specifickém tématu. Zjistil, že pokud časopisy rozdělíme do tří skupin, každá vytváří stejný počet článků o specifickém tématu, počet časopisů v každé skupině se zvyšuje geometricky. To znamená pro následnou oblast n článků, kdy člověk musí najít následně a , ab a ab^2 článků. Správný výběr základu a rozsahu může být redukován k Weber-Fechnerovu vztahu. Vztah může být vyjádřen mnoha způsoby a stejně tak jako diskrétní proměnné.

3.5.7. Zipfův zákon

Zipf rozvinul a rozšířil pole empirického vztahu pozorovaného Estoupem mezi řazením slov dle frekvence. "Slovo" je zde jakýkoliv řetězec písmen omezený mezerami a užívaný v textu. Pořadí slova, kde r je počet slov, zahrnující slovo samo, která mají alespoň stejnou frekvenci, $f(r)$ výskytu. Zipfův vztah je:

$$f(r) = k/r \text{ (k přibližně } 1/10)$$

Benoit Mandelbrot danou úvahu založil na informační teorii, získanou obecnější a lépe odpovídající formu. Mandelbrot vzal "úsilí" nebo "cenu" slov získaných v přenosu jako sekvenci vzor slova nebo fonému, oddělené mezerami nebo pauzami. Nastolil premisu, že cíl jazyka je dovolit přenos co nejvíce různých signálů (tj. "informace" ve specializovaném smyslu daného slova v informační teorii) jak jen to je možné s co nejmenším zpožděním, použil techniky pro porovnání kódů užití zprávy. Takto získal vztah mezi frekvencí slova $f(r)$ a pořadí slova, r jako:

$$f(r) = k(r+c)^{-\theta}$$

Konstanta c zlepšuje vhodnost pro běžná slova, která mají nízké r , nebo pokud vztah je interpretován jako Bradfordova rozložení, pro nejproduktivnější časopisy. Exponent zlepšuje vhodnost pro velké r , to je pro vzácnější slova nebo méně produktivní časopisy. Pro většinu přirozených jazyků θ je spíše větší než jedna, pro jazyky s vynucenými slovníky a pravidly užití to je méně než jedna. Perfektně aplikované indexování jazyka má tendenci zobrazovat se jako horizontální křivka s θ , protože každý termín by byl použit v průměru stejně často.

Parker-Rhodes a Joyce získali derivovaný Zipfův vztah. Jejich vztah je stejný jako Lotkuv pro počet autorů dané produktivity

$$n(u) = j \cdot u^{-2}$$

Derivovali toto přijetím, že slova ve slovníku jsou uložena a skenována v pořadí se snižující frekvencí, že toto je proporcionální k počtu skenovaných slov a že jazyk má zájem se rozvinout k danému maximu proměnných ("informací").

Hyperbolické rozložení vzniká z vlastností přetržitých kolekcí nebo pokusů kvantovat nepřetržitý fenomén. Jedinec musí zpracovat problém k začátku v jeho přesné přetržité formě. Aplikace výsledků, kontinuální forma je mnohem více vyhovující jako pravidlo a může být dohodnuta jako vhodná danému rozsahu s rozumnou přesností.

Omezení platnosti vztahů je část vztahu. Jestliže se mění v každé aplikaci, musí být považovány za přidané parametry.

V mnoha aplikacích měřítka a omezení jsou stanoveny ne fenoménem samotným, ale citlivostí, rozlišením a rozsahem metod pozorování nebo prezentování. Podobnost sama sobě charakterizuje Bradfordovu distribuci, protože jestliže rozdělíme dále jakoukoliv z originálních zón na další zóny, budou se řídit stejným zákonem, ačkoliv s rozdílným parametrem. Stromy stejného druhu mohou reprezentovat také stavění slov z písmen, čísel-slova z čísel, fráze ze slov a mezer, úroveň fyzického přístupu nebo administrativní přístup do knihoven, encyklopedií nebo počítačových skladišť; všechna čísla z nejlepších čísel; náhodné kroky a házení mincí a tak dále neomezeně. Není překvapující, že hyperbolická rozložení jsou tak rozšířená v kontextech

managementu a dokumentace, kde aktivity jsou podobné jako modely na rozdílných úrovních designu a potřeby.

Bradford např. rozdělil časopisy do tří zón rovných polí. Pokud se na to podíváme z jiného úhlu pohledu, pomůže nám to pochopit rozložení Normala Gausiana - které může být generováno díky binomické distribuci s celočíselnými exponenty, hyperbolicky z binomických distribucí z exponentů mající desetinné číslo.

Vickery vyčistil a rozšířil matematickou část Bradfordovy komunikace. Zejména ukázal, že jestliže vztah je udržován v obecné rovině, musí platit pro jakoukoliv podskupinu ve skupinách stejných výnosů, ne jen pro vysoké, střední a nízké vytěžované kvantity.

Leimkuhler podotknul:

$$F(x) = 1 + 0.185 \ln x$$

Leimkuhler zkoušel existující studie a příklady Bradfordova rozložení a derivoval různé formulace pro nespojitě a spojitě proměnné. Vickery zjistil, že základním parametrem je míra b_m mezi počtem zdrojů časopisů v následných zónách, kdy časopisy jsou seskupovány do m zón rovných výnosů, s časopisy ve snižujícím se pořadí dle výnosů. Vztah mezi těmito poměry pro rozdílné počty zón, m, n je:

$$b_n^n = b_m^m = b_2^2$$

To může lépe viditelné, pokud vezmeme v úvahu podskupiny mn zón. Míra b_m^m se ukazuje být stejná, když první dělíme na m zón a poté dále rozdělíme každou do n nebo do n zón a poté na další podskupinu do m .

Brookes přezkoumal literaturu od Bradforda a novější literaturu a formuloval distribuci přímo z úvah, že počet pertinentních článků dává výnos n pořadí časopisů přesahující a že $(n+1)$ tý z $b_1 n [n/n+1]$, kde b je poměr mezi po sobě jdoucími stejně-výnosnými zónami časopisů. Pro většinu zdokonalených podskupin musí být tento rozdíl výnosu jeden článek.

Goffman a Warren zjistili, že Bradfordova rozložení časopiseckých článků vyplývá z rovnováhy autor-editor spíše než předmět-časopis. Pro krátkost publikování jednoho masivního časopisu obsahující všechny články všech autorů existují dvě politiky pro oddělení jejich výstupu. Jeden je publikování časopisů tak, že každý využívá takového množství autorů, jak jen je to možné, což vede k omezení praktického publikování; např. *Nature*, *Science*, *Readers Digest*. Další politika je dána ekonomickým omezením. Těmto vydavatelským ekonomikám odpovídá politika autorů.

Fairthorne ve svých pracech upozorňuje na vypovídací schopnosti netechnické angličtiny, kde rozsah platnosti a informace o teoretických metodách nalézáme v dokumentaci, ve slovní zásobě a jejím kódování, pracujeme se shromažďováním a uspořádáním souborů, s přístupem k souborům, klávesnicím a policím s knihami.

Pokud vezmeme v úvahu abecedu n písmen, každé stejné délky nebo doby trvání nebo požadující stejné úsilí nebo peníze na vytištění nebo psaní, pak se má za to, že: za prvé, že všechny kombinace písmen jsou dovoleny a že budeme ignorovat mezery jako písmeno. To znamená, že slova budou uvažována jako izolované výskyty. (Ve skutečnosti mezery a uspořádání mezer jsou nejdůležitější typ písmen v částech posloupnosti). Existují n jednopísmenná slova, n^2 dvoupísmenná slova... n^r r slovná

slova a tak dále. Jejich cena je 1,2,... r... jednotek. Takto získáme důvěrné známou geometricko-aritmetickou shodu mezi počtem slov s jejich cenou, kterou pozorujeme tak často mezi zdroji a jejich výnosy.

Jestliže zde předpokládáme, že méně cenná slova jsou používána častěji, v izolaci, než cennější slova a že stejnou měrou cenná slova jsou používána stejně často, získáme Zipf-Bradfordovu hyperbolickou distribuci s jednotou exponentu, někdy nazývanou „harmonická“ rozložení“.

Costelo a Wall v roce 1959 předložili empirický vzorec vztahující se k počtu t z indexovaných termínů (klasifikace Uniterm), nezaměřené na určitou disciplínu a počet dokumentů n indexovaných pomocí nich.

$$t = 4,170 \log_{10}(n + 730) - 11,620$$

Wall podporoval normální logaritmickou distribuci (takové rozložení, kde logaritmus proměnné má normální rozdělení), spíše než hyperbolickou (z použití indexování termínů); kde hodnota logaritmu dokumentu na termín je rozptýlena okolo průměrné hodnota dle normálního Gaussova rozložení.

Wall zobecnil rozložení stejně tak jako logaritmické normální rozložení pro různé vztahy pro užití v indexovaných termínech. Také přizpůsobil procedury pro užití předpovídaných termínů jako funkci velikosti indexu. Zde jsou pouze empirické a mimořádně vypracované termíny užívající mnoha libovolných parametrů. Nicméně Wall citoval a užíval obsáhlá data z více indexů, slovníků a seznamů předmětových hesel v čase. V některých případech dostal Zipf-Bradfordův vztah mezi pořadím termínu a jeho ustanovenou frekvencí.

Zákony sami však vytvářejí určitá omezení zlepšení a metr pro kontrolu kvality. Neuposlechnutí zákonů je významnější než poslušnost, protože to indikuje, že podmínky se mění.

Tento souhrn sám podporuje Bradfordův závěr z jeho zákona, který říká, že pokud chceme více poznat náš předmět specializace, musíme jít vně. Nesmíme zapomínat na minulost (minulost je prospěšná předejhra), ale také nezůstávat stát na místě a hledat nové cesty.¹⁰

Fairthorne odkazuje na všudypřítomné zákony, které mají většinou „hyperbolické“ rozložení, (Zipfův základní zákon je hyperbolický, Lotkův ne).

Fairthorne si šímá také Hoffman-Newillova epidemického modelu jako časově závislého dostupného modelu a ptá se: Jaké jsou empirické „hyperbolické“ distribuce?

1. Weber-Fechnerovy psychofyzikální zákony
2. Paretova distribuce v ekonomii
3. Willisova taxonomická distribuce
4. Lotkův zákon o produktivitě autorů
5. Bradfordův zákon
6. Zipfův zákon a principy nejmenšího úsilí
7. Počáteční distribuce čísel (Benfordův zákon)
8. Práce Richardsonova a Mandelbrotova o důvodech války, délka drsných pobřeží a lomená dimenzionalita.

¹⁰ FAIRTHORNE, Robert A. Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction

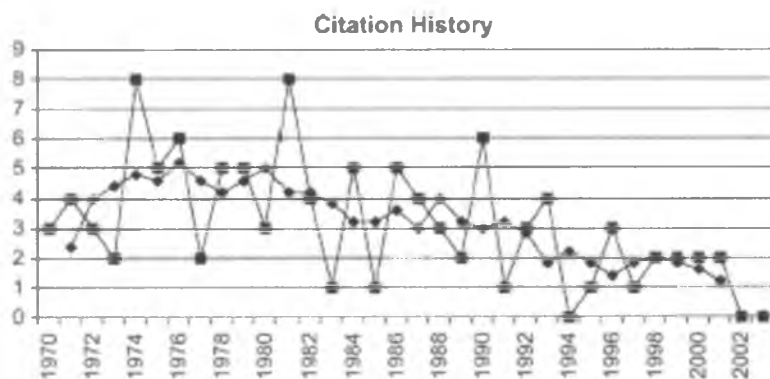
Slovo „empirické“ má velký význam. Fairthorne upozorňoval, že popis leží daleko od vysvětlení. Navíc konstatoval, že stejný sociální fenomén může mít mnoho rozdílných příčin. K tomu dodal, že necitlivost k příčinám ztěžuje vysvětlení, ale pomáhá předpovědi. Pro Fairthorneho je všudypřítomnost silných zákonů pouze důsledek vlastností směsi postavené na opakování a kombinování základních jednotek, abeceda je složená z písmen a mezer, slovníků slov a frází, fenoménů, týmů autorů, atd. závislých na konkrétní aplikaci.

Fairthorne byl mezi prvními (možná dokonce první), kdo si byl vědom, že pravidelnosti pozorované a studované Zipfem, Lotkou, Bradfordem, Mandelbrotem a Leimkuhlerem vycházejí ze stejného univerzálního principu. Fairthorne však nechal matematický důkaz těchto premis na dalších autorech.

Bookstein napsal a připomněl starou indiánskou pohádku o slepých mužích a slonovi. Pokud pohádku aplikujeme na naši problematiku, tak Zipf zkouší nohu slona, Bradford jeho chobot, Lotka jinou část jeho anatomie. To vše v současnosti náleží k standardnímu jádru znalostí v našem oboru, i když bibliometrické nebo infometrické zákony (lépe pravidelnosti, protože tyto fenomény nejsou zákony přírody) jsou od základu ekvivalentní.

Fairthorne uvedl příklad článku, kde pracovní název byl „Zipf unfastened“ (uvolněný Zipf). Však název, pod nímž byl článek publikován, byl „napěchován klíčovými slovy“ v takovém pořadí, aby přilákal pozornost širokého spektra potencionálních čtenářů. Tím upozornil na vztah mezi Bradfordovým zákonem a Mandelbrotovými fraktály a zdůraznil, že tím, že fenomény popsané bibliometrickými zákony jsou oddělené procesy, souvislé funkce mohou být použity jen k jejich popisu obecnými termíny, ale ne pro jejich vysvětlení. Tímto problémem se zabývá celá řada dalších autorů, včetně Egghe a používá další funkce jen pro modelování infometrického fenoménu.

Na **obrázku Roční počet citací** je vidět, že průměrná křivka má nejvyšší hodnotu v roce 1976 (ale rok 1980 je také vysoký). Obecně to znamená, že dochází ke zvýšení během prvních let, pak následuje pomalý pokles. Mějme na paměti, že nebylo myšleno, že citační křivky obecně mají pomalý pokles, jak je tomu zde, ale je to druh chování, které Ronald Rousseau předpokládá pro tzv. citační klasiku v „nepopulárním“ oboru jako je informační věda.



Roční počet citací (čtverce) a průměr za pět let (diamanty)

Jedná se o citační historii článku „Empirical hyperbolic distributions“ vydaného v časopise *Journal of Documentation* v roce 1969.¹¹

¹¹ ROUSSEAU, Donald. Robert Fairthorne and the empirical power laws

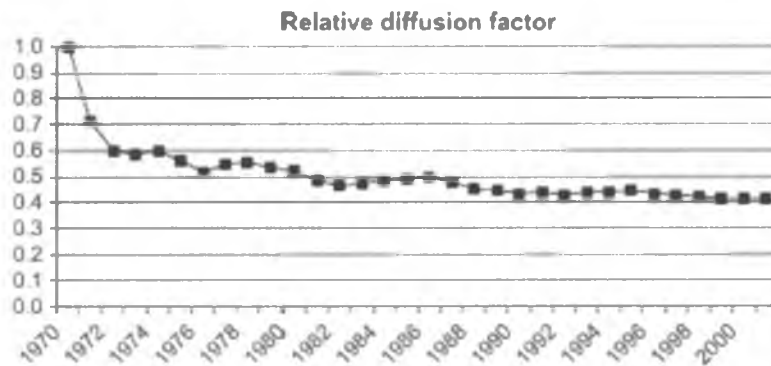
Povšimněme si rozptylu Fairthorneho článku v časopise *Journal of Documentation* v dalších časopisech. Relativní faktor rozptylu byl definován Rowlandesem v roce 2002 jako „měření mezioborového vnímání sady dokumentů, způsob shromažďující šíří vlivu dokumentů v literatuře“. V Rowlandsově originálním článku je tato sada dokumentů skupinou článků publikovaných v jednom konkrétním časopise (vedoucí k faktoru rozptylu časopisu; journal diffusion factor), ale definice je platná obecně a dokonce pro sadu obsahující i jen jeden článek. *N*-let diachronní relativní faktor rozptylu sady dokumentů publikovaných v roce *Y* je definován jako:

$$RDI_n(Y) = \frac{\sum_{j=0}^{n-1} U(Y+j, Y)}{\sum_{j=0}^{n-1} CIT(Y+j, Y)}$$

Počet unikátních, nových časopisů citujících alespoň jeden z dokumentů (z určité sady dokumentů) v roce $Y + j$ je označováno jako $U(Y + j, Y)$. Fráze „unikátní, nové“ znamená skutečnost, že časopis necitoval jeden z dokumentů (publikovaných v roce Y) během let $Y, Y+1, \dots, Y+j-1$, ale že citoval (v roce $Y + j$) alespoň jeden z dokumentů z dané sady dokumentů.

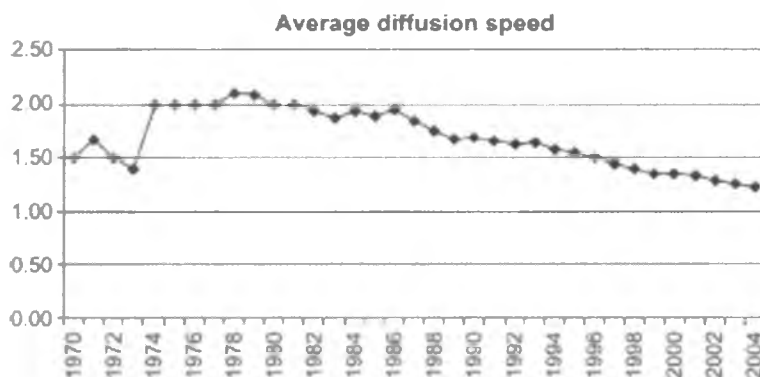
Symbol $CIT(Y+j, Y)$ značí počet citací získaných v roce $Y+j$ skupinou dokumentů publikovaných v roce Y . Dělení počtem citací vytvoří index relativního indikátoru. V tomto případě $Y=1969$ a počítá se RDI_n pro $n=1$ (1970, protože zde nejsou žádné citace v roce 1969) k $n=35$ (rok 2004).

Na obrázku To **Průměrná rychlost rozptylu** článku „Empirical hyperbolic distributions“ je průměrná rychlost rozptylu poněkud nepravidelná v prvním roce, ale poté se zvyšuje, dosahuje maxima v roce 1978, a pak pomalu klesá. Jak zde nejsou žádné citace v roce publikování, průměrná rychlost rozptylu není definována v tomto roce publikování.



Relativní faktory rozptylu pro článek Fairthorna za léta 1970-2004 ¹².

¹² ROUSSEAU, Donald. Robert Fairthorne and the empirical power laws



Průměrná rychlost rozptylu článku „Empirical hyperbolic distributions“¹³

Faithorneho „Empirical hyperbolic distribution“ (Bradford-Zipf-Mandelbrot) pro bibliometrický popis a předpověď je jeden ze základních článků časopisu *Journal of Documentation*. Ačkoliv jeho vliv, jak je měřeno citacemi, klesal v rámci let, může být spravedlivě nazýván skutečnou citační klasikou.¹⁴

Mnoho výzkumných prací v infometrii se týká nerovnosti aspektů. Koeficient Giniho je měření koncentrace/nerovnosti produktivity v jedné sadě dat. Problémem je měření relativní nerovnosti produktivity mezi dvěma sadami dat. Měření originálně předkládané Dagumem, analogické k Giniho koeficientu, je diskutováno a rozvíjeno Burrellem na základě teoretických a empirických ilustrací. Z toho bylo získáno standardizované měření – relativní koeficient koncentrace – založené na myšlence „relativního ekonomického nadbytku“, které také představil Dagum. Nové standardizované měření – paralelní koeficient koncentrace, v některých způsobech analogický korelaci koeficientu – je definován v práci Burrella.

Infometrie je proces populace zdrojů vytvářející položky v čase. K nim patří např. studie oběhu v knihovně a půjčování.

Aspekty koncentrace jako jeden ze zkoumaných parametrů v bibliometrii a infometrii se objevují např. u Lotky, kde demonstruje ohromující dominanci v termínech publikování relativně malého počtu autorů, Bradford hledá definici „jádra“ nejproduktivnějších časopisů a Trueswell se snaží najít definici „jádra“ nejvíce používaných textů v knihovní sbírce. Ačkoliv význam je v identifikování nejproduktivnějších zdrojů, je jasně potvrzeno, že existuje obecný stupeň nerovnosti mezi zdroji, měřené jejich produktivitou. Příští přirozený krok je definovat a poté měřit nerovnost s jakoukoliv jednotlivou sadou dat. Mezi prvními se tím zabýval Ravicharda Rao následovaný pracemi Eggheho a Rousseaua.

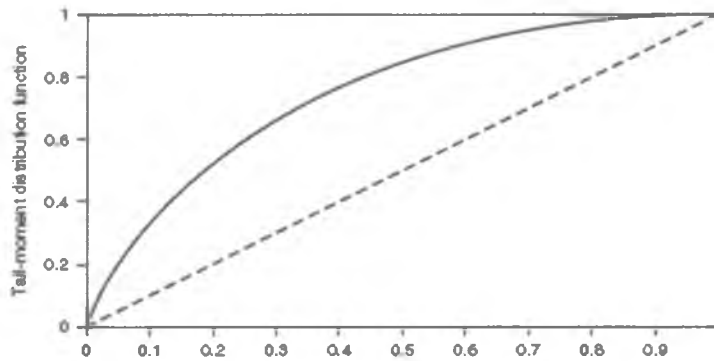
3.6. Použití a omezení koeficientu Giniho

V ekonomii je nerovnost v příjmech a bohatství vysvětlována pomocí tak zvané Lorenzovy křivky koncentrace. V infometrii, kde jsou zajímavá obvyklá ohniska nejproduktivnějších zdrojů, je ekvivalentní grafickou reprezentací Leimkuhlerova křivka. Rozdílnost mezi oběma konstrukcemi můžeme vysvětlit tím, že pro Lorenzovu křivku se řadí zdroje dle zvětšující se produktivity, zatímco u Leimkuhlerovi křivky jsou seřazeny

¹³ ROUSSEAU, Donald. Robert Fairthorne and the empirical power laws

¹⁴ ROUSSEAU, Donald. Robert Fairthorne and the empirical power laws

dle snižující se produktivity. V obou případech načrtneme poměr celkové produktivity vzhledem k souhrnnému poměru zdrojů takto:



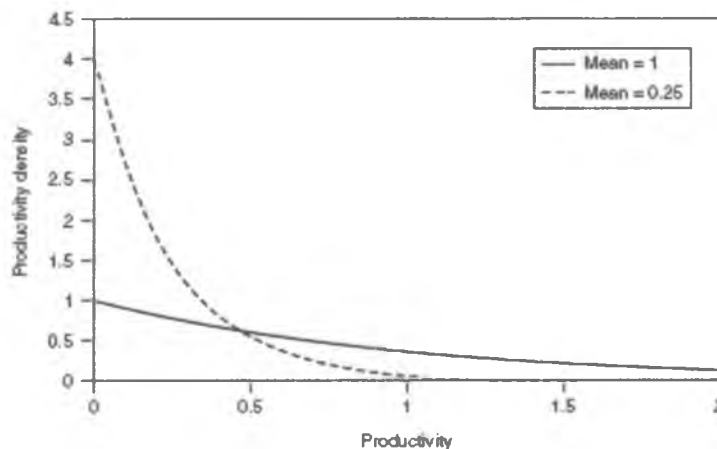
Příklad Leimkuhlerovy křivky – exponenciální uspořádání ¹⁵

Myšlenkou této definice je, že pokud se podíváme na každou dvojici zdrojů v populaci v řadě, zjistíme absolutní rozdíl mezi jejich produktivitou a poté průměr mimo všechny možné dvojice.

Grafická interpretace $0 \leq \gamma_x < 1$. Příklad $\gamma_x = 0$ odpovídá všem zdrojům rovnajícím se produkci; jestliže je zde právě konečný počet produktivních zdrojů, poté existuje omezení, že celkový počet (neproduktivních) zdrojů se stává nekonečný

$\gamma_x \rightarrow 1$. To je také jasné z grafické reprezentace, že Giniho koeficient respektuje Leimkuhlerovo uspořádání, to znamená, že jestliže Leimkuhlerova křivka z jedné produktivity rozložení dominuje, potom je Giniho koeficient větší.

Giniho koeficient je obvykle sledován jako jeden parametr. Jestliže tomu není tak, pak se měření nerovnosti nejlépe provádí ze všech sedmi „vhodných“ vlastností navrhnutých Daltonem pro takováto měření.



Dvě exponenciální rozmístění ¹⁶

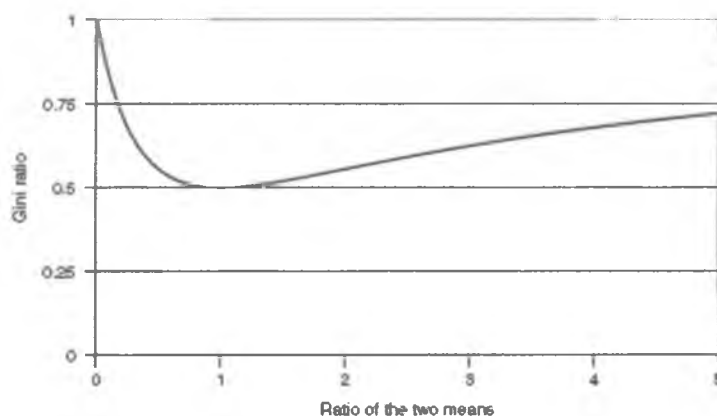
¹⁵ BURRELL, Quentin L. Measuring similarity of concentration between different infometric distributions

¹⁶ BURRELL, Quentin L. Measuring similarity of concentration between different infometric distributions

Exponenciální je rozsah parametru rodiny, Giniho koeficient pro každou populaci bude stejný. Burrell ukazuje, že Giniho koeficient pro exponenciální rozmístění je jedna polovina. Proto lze důvěřovat Leimkuhlerově křivce a standardnímu měření nerovnosti, které se rovná Giniho koeficientu, pokud nejsou nevhodná pro měření relativní nerovnosti mezi populacemi.

3.7. Giniho poměr mezi populacemi

Použijeme Giniho koeficient, abychom získali měření vše zahrnující nerovnosti infometrické produktivity mezi dvěma populacemi zdrojů. Myšlenka zkonstruování Giniho poměru mezi dvěma populacemi je přesně analogická pro Giniho koeficient pro jednu populaci. Posoudíme dvojici zdrojů, ale nyní jednu z každé populace, zjistíme absolutní rozdíl mezi jejich produktivitami a zjistíme průměr tohoto rozdílu u všech možných dvojic. Giniho poměr závisí jen na *relativním* významu produktivity dvou populací. Graf naznačuje unikátní minimum $\alpha = 1$, korespondující s dvěma distribucemi, které jsou stejné v případě, kdy Giniho poměr předpokládá hodnotu Giniho koeficientu jako jednu polovinu. Jestliže existuje rozdíl ve významech, poté se to odráží v hodnotě Giniho poměru, který bude větší než jedna polovina.



Giniho poměr pro dvě exponenciální rozmístění ¹⁷

V praktických infometrických aplikacích, kde porovnáváme dvě sady dat, to znamená, že počet zdrojů a počet položek bude konečný.

3.8. Příklad: Bradfordovy originální sady dat

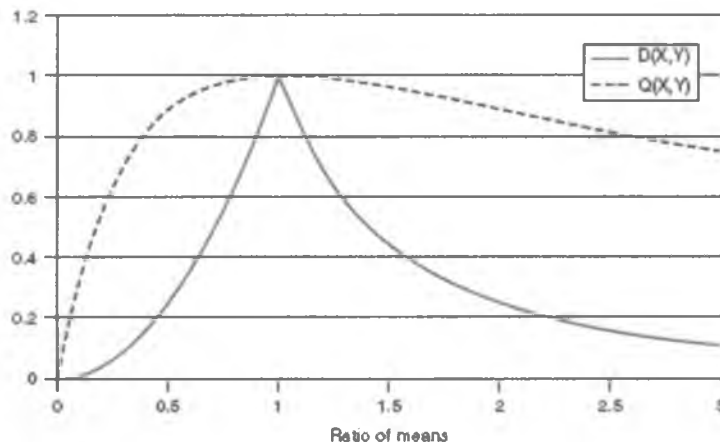
V pokusu identifikovat „jádro“ časopisů v rámci disciplíny prezentoval Bradford (právě) dvě sady dat pro ilustraci myšlenky, která později dosáhla status „zákon“ v infometrii. Bradfordova myšlenka je vyjádřena graficky jako lineární. Mnoho studií hledalo dokázání/ilustrování/nedokázání tohoto empirického pozorování užitím další sady dat (dobrý náčrt a další technickou analýzu představuje Egghe). V této originální práci Bradford prezentoval data udávající počet článků v časopisech. Applied Geophysics (AG) a Lubrication (Lub) identifikují takovéto články přes (přibližně) tříletou periodu.

Giniho poměr není standardizovaný, těžko říci, zda je velký nebo malý. Je otázkou, jaká má být hodnota pro poměr nebo jak by to mělo být interpretováno.

¹⁷ BURRELL, Quentin L. Measuring similarity of concentration between different infometric distributions

3.9. Porovnání dvou měření

Zajímavý je rozdílný způsob, jak dvě měření rozlišují mezi rozdílnými exponenciálními rozloženími.

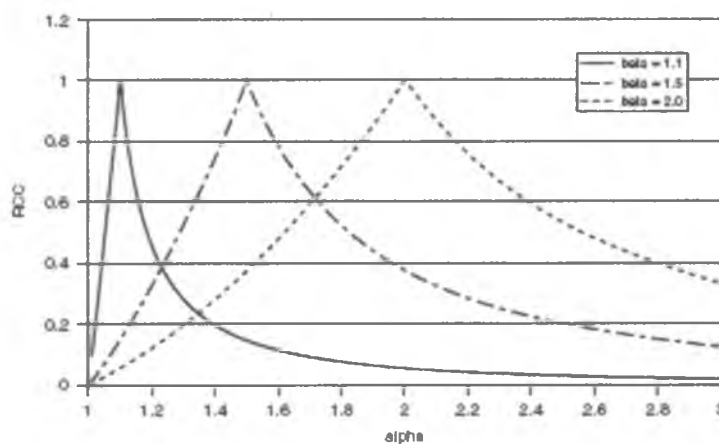


Porovnání dvou měření v exponenciálním případě¹⁸

3.10. Dvě Paretovy distribuce

Běžné delší kalkulace jsou vyjádřené pro Giniho poměr dvou Paretových rozložení a ukazují, že v tomto případě relativní koeficient koncentrace je dán:

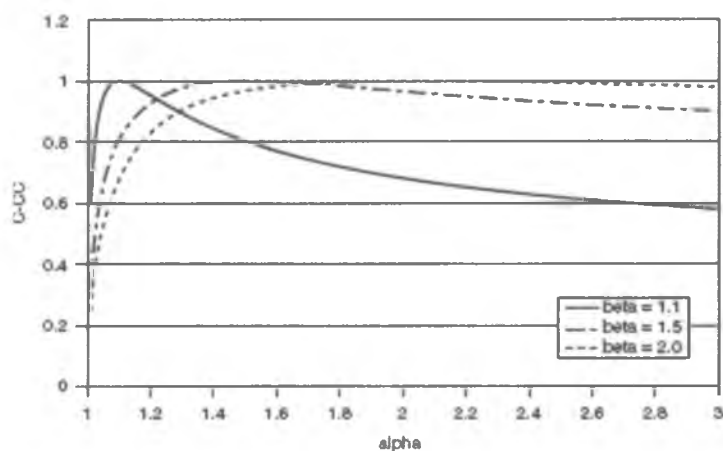
$$D(x,y) = \frac{\alpha(\alpha - 1)}{\beta(\beta - 1)}, \text{ kde } 1 < \alpha \leq \beta$$



(a) relativní koeficient koncentrace pro Paretovo rozložení.¹⁹

¹⁸ BURRELL, Quentin L. Measuring similarity of concentration between different infometric distributions

¹⁹ BURRELL, Quentin L. Measuring similarity of concentration between different infometric distributions



(b) koeficient paralelní koncentrace pro Paretovo rozložení.²⁰

Dvě infometrická rozložení, obě získaná z Giniho (poměrného) poměru. Jedno z nich, relativní koeficient koncentrace, je velmi citlivý na rozdíly ve významu dvou rozložení; koeficient paralelní koncentrace na druhé straně je mnohem více robustní a odráží koncentrace napříč neporušenými rozloženími.²¹

Sociální vědy se liší navzájem stejně tak jako se liší od přírodních věd v publikování a citačních modelech. Zatímco různé způsoby organizování dat vytvářejí rozdílné výsledky, disciplíny jsou konzistentní v sadách dat. Když porovnáme sociologii/antropologii s dalšími disciplínami, ukazují se podstatné rozdíly mezi jednotlivými univerzitami, kde se dané obory studují.

Jak Merton říkal, problémem je dosáhnout měření alternativních způsobů, jak jsou znalosti zpřístupňovány a užívány. Jestliže snadno měřitelné a přístupné indikátory jsou používány k měření výzkumného výkonu, poté je seznam dostupných indikátorů omezen. Takový seznam zahrnuje publikace (knihy, časopisecké články) stejně tak další viditelné produkty vědy (např. hudební kompozici, architektonickou kompletaci). Tyto produkty odrážejí úsilí a výstup, ale nejsou pravděpodobně užitečné, jestliže se zaměříme na kvalitu výstupu. Můžeme argumentovat tím, že člověk zná kvalitu, když ji vidí. Takový argument ignoruje vnitřní disciplinární konflikt (konflikt ve vztahu disciplíny k disciplíně), což ovlivňuje posouzení kvality.

Disciplíny jsou organizovány v přírodní vědy, které mají obvykle velmi zřetelnou pozici a sociální a humanitní vědy jako sociologie. Někteří vědci také navrhnou, že tyto vědy jsou „nevědami“. Toto rozvrstvení je založeno na tom, jak strukturovány znalosti v každé disciplíně jsou a zvláště, jak organizace a soudržnost každé disciplíny napomáhá vývoji znalostí v rámci oboru. Obecně přírodní hard vědy jsou viděny jako lépe organizované a strukturované než sociální vědy a jsou proto efektivnější v rozvíjení a vytváření znalostí.

Centrální hodnota vědeckého úsilí je chápána jako racionalita a univerzálnost. U více zralých disciplín se věří, že demonstrují silnou tendenci k vytváření hodnotového systému. Hierarchie věd je založena nejen na komplexnosti studovaného fenoménu, ke kterému disciplína rozvíjí paradigma nebo sdílí teoretickou strukturu a metodologické přístupy, o nichž existuje vysoký stupeň konsensu.

20 BURRELL, Quentin L. Measuring similarity of concentration between different infometric distributions

21 BURRELL, Quentin L. Measuring similarity of concentration between different infometric distributions

Sociální a humanitní vědy jsou vnímány jako disciplíny, které projevují méně konsensu při vytváření základních bloků. Fuchs přidává, že přírodní vědy jsou většími adepty v kontrolování a chránění zdrojů ve větší centrální hierarchii, zatímco sociální vědy mají menší, slabší a rozprostřenější zdroje. Tyto rozdíly mohou být evidentní v počtu prací akademie, kde musíme uvažovat parametry jako jsou učení a výzkum ve vztahu student-učitel.

Studie *The validity of publication and citation counts for sociology and other selected disciplines* se zaměřila na hodnocení v molekulární biologii, která se jeví jako rozrůstající se disciplína manifestující rychlý růst na mnoha australských univerzitách. Dále sociální disciplíny – politická věda a administrace a sociologie/antropologie, psychologie, protože to je disciplína, která stojí na rozhraní sociálních a přírodních věd a jeví se být ve fázi růstu. Konečně byla vybrána také filozofie.

Citace je subjektivní rozhodnutí vytvořené autorem v určitém bodě psaní článku s úmyslem citovat práci jiného autora. Takovéto posouzení je ovlivněno dostupností materiálu a autorovým povědomím o tomto materiálu, což je střídavě ovlivnitelné jejich sítí informací v rámci daného oboru. Studie ukazují, že lidé ze stejné sítě mají tendenci často citovat každou práci dalšího vědce v daném oboru a že existují silné geografické a regionální tendence autorů citovat každou další práci. Další studie ukazují, že jistá jména v rámci oborů přitahují více citační zájem než jiná. Existuje také zjevná tendence pro články ve stejných časopisech jádra být častěji citovány než ty, které jsou obsaženy v okrajových časopisech, ačkoliv některé studie zjistily, že články jsou citovány primárně z rozhodnutí autora a ne pouze pro dříve zmíněné faktory.

Některé typy akademických prací nejsou adekvátně reprezentovány v citačních databázích. Např. monografie/knihy a práce publikované vládními agenturami se neobjevují tak často jako primární zdroje nebo jako citované práce. Práce tohoto typu pak budou podhodnoceny ve všech ISI publikacích a výpočtech citací. Sociální a humanitní vědy publikují tyto práce častěji než přírodní a fyzikální vědy.

Většinu vysoce impaktových disciplín najdeme v oblasti přírodních věd nebo medicíny. Psychologie má druhý nejvyšší impakt. Sociologie/antropologie je na dalším stupni, pak následuje filozofie a politické vědy (v této studii). Výsledky studie postihují hlavní rozdíly v úrovni publikování a citačním poměru vlivu těchto disciplín a posuzuje rozdíly mezi obory dané tradicí citační praxe, délkou článků a počtem lidí pracujících a publikujících v oboru.

Velké změny jsou v posledních letech v biologii a genetice, ale naopak např. ve filozofii jsou zaznamenány jen malé změny v impakt faktoru v čase.

Politické vědy mají podobný model. V letech 1981-99 byl zaznamenán mírný nárůst citací na počet univerzit, ale nárůst není konzistentní.

Psychologie má konzistentnější nárůst v čase, ačkoliv se tento poměr mění instituci od instituce.

U sociologie/antropologie je model nekonzistentní. Některé univerzity mají nárůst, jiné mají stabilně vysokou úroveň v citacích v čase. Citační model v sociologii se liší mezi univerzitami více než v jiných disciplínách. To může odrážet rozdílné tradice v tomto oboru, které existují na různých univerzitách.

Je možné porovnat citační poměry, abychom viděli, zda se některé disciplíny projevují efektivněji na každé ze sledovaných univerzit. Žádná katedra filozofie nevyčnívá jako nejsilnější. Je to dáno relativně malým počtem zahrnutých citací, rozdíly mohou dobře odrážet roční kolísání. Mnohé z těchto pravidel najdeme i u politických věd a veřejné administrativy. Žádná instituce nestojí mimo, žádná instituce nevykazuje abnormální

nárůst. Existují jen malá vybočení u některých univerzit, což je dáno jejich relativně malým počtem. U psychologie existují některé instituce, které dosahují lepší úrovně.

V sociologii/antropologii jsou závěry nejasné, dané relativně malým počtem zahrnutých subjektů.

Existují tři hlavní zdroje relevantních dat pro vymezení výzkumné produktivity a kvality výkonu výzkumu: publikace, citace a ústavní afiliace. Všechny tyto zdroje jsou problematické a všechny zahrnují významný riziko chyb a/nebo špatné klasifikace. Otázka, co se stalo s monografiemi a jak je možné spočítat monografie v kontextu vědecké produktivity, je problematické. Ne všechny monografie představují příspěvek do fondu znalostí. Někteří vydavatelé publikují monografie jen tehdy, pokud autoři poskytnou příspěvek. Další vydavatelé využívají nízkonákladové formy publikací, což zahrnuje drobné tisky a kde také není nutné realizovat rozsáhlý proces recenzí. Další monografie jsou publikovány na komerční bázi a nejsou zamýšleny jako vědecké publikace. Existuje potřeba rozvíjet lepší způsoby posouzení výzkumného příspěvku monografií. Ať už bude vybrána jakákoliv metoda (shromáždění recenzí monografií nebo něco jiného), je jasné, že současný status monografií není efektivně začleněn do procesu ISI.²²

Mnoho vědců shledává práci významnou jen na základě toho, jak často je citována ve vědecké literatuře. Tak se ale může stát, že citační analýza v rukách neodborníků bude velmi nebezpečným nástrojem.

Po dobu čtyř desetiletí ISI (Institute for Scientific Information) skenoval seznamy citací dříve publikovaných prací. Výsledná databáze byla aktuálně vyvíjena pro informační vyhledávání – dovolující výzkumníkům provádět rychlé vyhledávání literatury a identifikovat jednotlivé vědce pracující na konkrétních tématech.

Protože vládní a jiné agentury podporující výzkum nemají jiné důvěryhodné nástroje pro měření výzkumné kvality, používají se často citační data ISI jako pomocníci při hodnocení kvality. Vychází se z toho, že významné práce budou citovány častěji. Jako obecné pravidlo je to rozumný předpoklad. Ale aplikované slepě, bez ohledu na vývoj disciplín, jejich rozdílné postavení ve vědě a omezení v uspořádání dat, mohou přinášet zcela zavádějící výsledky.

Prioritní postavení ISI ztěžuje použití dat dalších nezávislých bibliometrických výzkumníků. Výsledkem dominantního postavení ISI je, že si některé skupiny stěžují na zvyšování cen a omezení využívání jiných dat, kvůli vynucované citační analýze.

Mnoho bibliometrických výzkumníků se podílí na tomto trendu a požadují, aby nezávislé skupiny pomáhaly s kontrolou kvality citačních dat. Další varují, že ISI představuje produkty, které jsou zneužívány.

Za příklad zneužití mohou sloužit podle expertů impakt faktory časopisu ISI – měření průměrného počtu citací shromažďovaných na práce, které každý časopis obsahuje. Vydavatelé rádi používají těchto dat, operují s výškou impakt faktorů v propagačních materiálech a knihovníkům jsou doporučovány jako vhodné pomůcky při rozhodování, které časopisy předplatit.

Ale zneužití impakt faktorů časopisů jde mnohem dále, používá se k hodnocení jednotlivých institucí, oddělení a vědců. Nejspravedlivější by v různých vědeckých pracech bylo přímé počítání citací. To je sice z dat ISI možné, ale získat ISI nebo

²² NAJMAN, Jake M.; HEWITT, Belina. The validity of publication and citation counts for sociology and other selected disciplines

nezávislou skupinu pro provedení analýzy je nákladné a časově náročné. Proto se jako levnější a příjemná alternativa nabízí, podívat se na záznamy vědeckých publikací a hodnocení kvality jejich výstupů na základě impakt faktorů časopisů, v kterých se jejich práce objevily. To je „ubohá citační analýza jedince“, říká van Raan.

Setříděný seznam časopisů v konkrétní oblasti se zdá vyjadřovat dobrou korelaci mezi impakt faktorem časopisu a vnímáním jeho kvality. Ale je to začátek porovnávání mezi obory – proti čemuž varuje i ISI – a výsledky se rychle stávají bezvýznamné: výzkumníci z matematiky zřídka citují více než jeden nebo dva zdroje, zatímco typická práce v molekulární biologii zahrnuje tucty odkazů. To je důvod široké variace v impakt faktorech, dokonce mezi porovnatelnými časopisy plní úkol rozdílné disciplíny, jak upozorňuje Per Seglen, výzkumník z oblasti rakoviny, který se zabývá bibliometrií v norském Radium Hospital v Oslu. Autoři mají sklon si oblíbit časopis, který převážně publikuje přehledové články a ty pak bývají citovány častěji.

Další problémy jsou méně zřejmé, ale když jsou použity na hodnocení jednotlivých vědců, stávají se závažnějšími. Seglen poukázal na to, že okolo 15 % článků v typickém časopise získává polovinu citací. To znamená, že typická práce v časopise s vysokým impakt faktorem nemůže ve skutečnosti být citována mnohem častěji než průměrná práce v hůře řazeném časopise. „To je obecná korelace mezi počtem citací na článek a impaktem časopisu, ale to je jeden způsob vztahu“. Seglen říká: „Časopis nepomáhá článku, pomáhá celé prostředí kolem“.

Např. ve Finsku vládní podpora univerzitních nemocnic je částečně založena na publikování s klouzající mírou odpovídající impakt faktoru časopisů, ve kterých výzkumníci publikují jejich práce.

U impakt faktoru časopisů čítatel v rovnici může zahrnovat citace na články, které se neobjeví v seznamu citujících položek – obecně jsou omezené na originální výzkumné práce a přehledové články. Dle Henka Moeda, výzkumníka v van Raanově centru v Leidenu, to může vytvořit impakt faktor, který je až do 40 % v některých případech příliš vysoký. K porozumění proč, je nezbytné vědět, jak ISI počítá citace.

Mezi živé problémy patří rozvoj procedur kontroly citací u prací vytvářených konsorciem. Citace na takovéto práce se objevují jako podhodnocené ISI. To se stalo předmětem zájmu v lednu (2002), kdy časopis Nature zkoumal podezřele nízký počet citací za poslední rok u prací zabývajících se lidskou genetikou vytvářené Human Genome Sequencing Consortium.

Největším zdroje chyb ale je, že vědci dělají chyby, když citují další práci. Raději použijí příspěvek ke známé „lži. Chyba v citování znamená, že citační statistika může chybně umístit zásluhu jinde. A jako citační počty se plazí takové chyby z práce do práce.

Pro velmi vysoce citované práce není neobvyklé zjistit varianty chyb jako je např. chyba v čísle ročníku časopisu. Variace v adresách také mohou způsobit spoustu chyb, zvláště v národních, regionálních a institucionálních porovnáních.

ISI aktuálně pracuje s použitím přesnější kontroly v databázi, což je hodnotná aktivita. V současné době ISI si drží efektivní monopol v oblasti citačních dat – ačkoliv výhodu v informační technologii mohou nést další alternativní databáze.

Např. CrossRef je neziskové konsorcium, kde je zapsáno 99 vydavatelů, kteří dovolují uživatelům kliknout přímo z citace na práci online s možností získat abstrakt nebo plný text asi 3.9 milionu výzkumných článků z více jak 5600 časopisů.

Pro získání této možnosti je nutný software, manuální systémy nebo kombinace obou, které mohou přesně identifikovat všechny citace – které budou zapsány v různých formátech a jsou často naplněné chybami. Takové systémy musí také identifikovat

elementy jako je jméno autora a počáteční stránka. Každá citace je poté porovnána s těmi v centrální databázi, aby byl nalezen identifikátor – URL pro webovskou stránku, kde se práce nalézají nebo nalézt unikátní digitální identifikátor objektu (DOI).

Ale jednou propojené práce a citace jsou spojeny pomocí odkazů, to je relativně jednoduché pro aplikaci konvenčního algoritmu, aby bylo možné vytvořit citační indexy, impakt faktory a další indexy a řazení citovaných výzkumníků a článků, jak to ISI dělá po léta.

CrossRef nedělá žádný pokrok v tomto směru, ale další skupiny ano. Mezi pionýry jsou vývojáři ResearchIndexu v NEC Research Institute v Princetonu, v New Jersey. Tento nástroj je vystavěn bez lidské intervence, digitální knihovna obsahuje více jak 300 000 prací z oblasti počítačové vědy nashromážděné na webu. ResearchIndex není bezchybný, ale v testech odkazy ke článkům dosahují okolo 95% přesnosti pro časopisy se standardní citační praxí, což je výkon blízky 100%. A vedoucí projektu Steve Lawrence argumentuje, že systém by mohl být rozšířen, aby pokrýval další publikace, více než databáze ISI, ale je to velmi nákladné z hlediska manuálních vstupů.

Plně automatizované systémy citačního indexování jsou rozšiřovány. Open Citation Project je vytvářen pro elektronické archivy preprintů a NASA Astrophysics Data System (ADS) má jednu ze sbírek 3 miliony abstraktů. Günter Eichhorn z Smithsonia Astrophysical Observatory v Cambridge, v Massachusetts, projektový vědec pro ADS říká, že uživatelé stále více užívají systém pro vytváření jejich vlastních citačních statistik.

Ke zlepšení výsledků kontroly používá např. společnost Parity Computing of San Diego hybridní přístup. Má automatizovanou většinu funkcí, ale také zaměstnává lidi na klíčových pozicích pro zlepšení kontroly kvality. Systém dosahuje přes 99% přesnosti. Společnost nedávno implementovala citační index pro archiv literatury za 50 let společnosti Association for Computing Machinery, největší světovou počítačovou společnost. Zamýšlí vytvořit u takovýchto citačních indexů tvorbu impakt faktorů a další citační statistiky.

Ovšem i když je používána automatizace, stále takovéto archivy představují vysoké náklady spojené se získáváním a organizováním velkého počtu časopisecké literatury v použitelné formě a pro zajištění kontroly kvality.²³

²³ ADAM, David. The counting house. News feature.

4. „Nová“ a „stará“ infometrie aneb infometrie se stále vyvíjí

Type/token-taken infometrie je nová část infometrie, která studuje položky. Položky jsou zde chápány jako objekty, které jsou vytvářeny zdroji (např. časopisy vytvářející články, autoři vytvářející práce atd.) V lingvistice je zdroj také nazýván type (např. slovo) a položka token (např. použití slov v textech). V infometrii types (slova) se objevují často, např. v informačním vyhledávání. Relativní použití těchto výskytů bude vyšší než jejich relativní výskyt samotný, proto pojmenování type/token-taken infometrie. Článek Lea Eggheho studuje frekvenci rozložení type/token-taken infometrie: Začíná se od jedné type/token infometrie (to je vztah zdroj-položka). Byl studován průměrný počet μ^* položek užitých v type/token-taken infometrii a porovnán s klasickým počtem μ v type/token infometrii. Ukazuje se, že μ^* je vždy \geq a že μ je navýšená funkce μ . Metoda je prezentována aktuálně počítající μ^* z μ a dané α , což je exponent v Lotkově frekvenci rozložení type/token infometrie.

Duální přístup infometrie je v současné době velmi dobře známý a také formy základních klasických infometrických zákonů. Příkladem je bibliografické skládání časopisů, které obsahují články na jisté téma. Články jsou brány jako zdroje, které „produkuje“ články, obecně nazývané položky. Použitím terminologie jdeme zpět k originální formulaci Bradforda. Budeme se tedy zabývat tím, co může být nazýváno zdroj/položka infometrie. Časopisy/články nejsou jen interpretacemi konceptu zdroj/položka. Lotka popsals vztah autor/práce, který potvrzuje jeho zákon frekvence

$$f(n) = \frac{D}{n^\alpha}$$

kde $f(n)$ označuje počet autorů s n pracemi $n = 1, 2, 3, \dots$, $\alpha = 2$, ale obecně můžeme použít „jakoukoliv“ hodnotu s vyšší než 1.

Zde byl omezen Lotka-type zákon frekvence, ačkoliv některé výsledky (takové jako Teorem III.1) jsou obecně platné. Důvodem je, že nový infometrický model (Type/Token-Taken) je nejlépe (první) aplikovatelný pro nejjednodušší možnou funkci f (z důvodu vypočítání). Sílu zákona představuje velmi jednoduchá funkce, pokud ji porovnáme s dalšími frekvencemi funkcí, které byly využity v infometrii v poslední době. To je široce užívaný zákon a má historické kořeny. Funkce s volnou mírou vlastnictví vyjadřuje, že pokud n je násobená konstantou (řekněme C , jiné „měřítko“ pro proměnnou n), máme stále až ke konstantě stejnou funkci f . Konečně, jak je ukázáno v mnoha publikacích, Lotkův zákon může být vzat v úvahu jako základ pro infometrii s nelotkovskou frekvencí funkce f (nebo alespoň jako první přiblížení). Také z důvodů porovnání infometrie Type/token s Type/token-taken infometrií je nezbytné, aby byla použita jednoduchá konkrétní funkce (jako silná funkce). Další příklady type/token infometrie (infometrie zdroje/položky) jsou:

1. klíčová slova (mezi nimiž hledáme klíče nebo N-gramy) a jejich výskyt v databázi
2. články a jejich odkazy, které dávají synchronní citační analýzu nebo citace získají (diachronní citační analýza).
3. knihy v knihovně a jejich půjčování v jistém čase
4. služby (např. oběh v knihovně nebo informační pult), jejich trvání (měřeno např. v sekundách)
5. slova a jejich výskyt v textu (např. kniha)

Poslední příklad vychází z kvantitativní lingvistiky. Zdroje (slova) jsou zde nazývána „types“ („typy“) a výskyt těchto slov je nazýván „Tokens“ (projevy). Mluvíme tudíž o lingvistice type/token (typ/projev). Tak by byly všechny zmíněné příklady duální infometrie (to je vztahy zdroj-položka) nazývány type/token infometrií (infometrie typ/projev). Tato terminologie je výhodná k popsání cíle článku: úvod toho, co se nazývá type/token-taken infometrics (infometrie typ/projev).

Problém s infometrií type/token (T/T) je, že jen popisuje zdroje a položky v těchto zdrojích. Jinými slovy – zdroje a jejich produkce. T/T infometrie (důležitá sama o sobě) nestuduje použití těchto položek v takové podobě, jak se s nimi setkávají informační profesionálové nebo další uživatelé.

1. vyhledávání klíčů (např. N-gramy) je používáno katalogizátory v knihovnické síti - to znamená, jak rychle je zjištěno, že byla přijata nová knížka (protože musí být katalogizována) a je již v databázi. Rozložení výskytu (projev) N-gramů (typů) v katalogu bude rozdílné od rozložení výskytu těchto N-gramů vyzkoušených katalogizátorem. Čím více N-gram výskytu v katalogu, tím více N-gram bude také používáno katalogizátorem.

To povede k většímu průměru zobrazení délek, než jak by bylo předpokládáno z výskytů N-gramů v databázi samotné. Katalogizátoři používají („token“ výskyt (projev) N-gramů (typu)+ (příklad Type/token-taken (T/TT) infometrie).

2. zevšeobecnění v příkladu: každý uživatel vyhledávající informace z databáze, užívající klíčové slovo použije více „populární“ (to je více užívané) klíčové slovo než méně populární a ta najdou více záznamů v databázi. Porovnání těchto uživatelů a počtu záznamů na vyhledávání bude delší než průměrný počet dokumentů obsahující toto klíčové slovo (a samozřejmě, stejně pro jejich rozložení, z nichž tyto průměry budou odvozeny). To je jiný příklad T/TT infometrie.

3. Některé knihy v knihovně jsou půjčovány často, ale mnoho jich je půjčováno jen zřídka. Uživatel bude hledat populární knihy častěji než jiné. Tak se v průměru bude uživatel setkávat s více půjčovanými knihami než s průměrným počtem výpůjček na knihu. To je také příklad T/TT infometrie.

4. Doba služby u pultu. Při dlouhé době služeb se s vyšší pravděpodobností setkáme s osobou, která hledá služby. Budeme se zajímat o průměr, kdy se služba s takovou osobou setká a o průměr doby služby (doba čekání) než o to, co vyplývá z předpokládaného průměru doby služby jako takové. Opět je to příklad T/TT infometrie.

T/TT infometrie nikdy nebyla studována. Egghe se pokusil najít nějaký příklad v literatuře a našel popis Rousseau o pozitivním zesílení (také efekt užití položek – zde nazývané dostupnosti, ale jen v kontextu efektu Lorenzových křivek -to je koncentrační teorie- a ne v kontextu obecných infometrických zákonů).

T/TT infometrické studie zákonů infometrie jako zkoušené uživatelem je v tomto smyslu důležitější než T/T infometrie.

Není jednoduché vymezit funkce $\mu^* = \mu^*(\alpha, \mu)$.

Ve čtvrté sekci (numerická metoda) bude vytvořeno tak, jak je vysvětleno a explicitní obraz $\mu^*(\alpha, \mu)$ je prezentován. Je zde neuzavřený pro další zkoumání problém ne-lotkovského T/TT infometrického vývoje (proto, nahrazení α dalšími (jinými) parametry (s)).

Začneme s Lotkovou T/T infometrií, kde adoptujeme pokračující uspořádání, jak bylo také u Eggheho. Označíme A jako celkový počet položek a T jako celkový počet zdrojů.

Klasická Lotkova T/T infometrie je dána Lotkovým zákonem (pokračující uspořádání):
pro $\alpha > 1$

$$f(j) = \frac{D}{j^\alpha}$$

pro $j \in [1, \rho(A)]$. Zde $f(j)$ označuje hustotu zdrojů (typů) s hustotou položek j (projevů) (porovnané s diskretním uspořádáním. Poznámka, že definicí :

$$\int_1^{\rho(A)} f(j) dj = T$$

a

$$\int_1^{\rho(A)} jf(j) dj = A.$$

To je jen formalismus duální teorie infometrie, jak ji rozvíjí Egghe. Funkce f popisuje výskyt položek ve zdrojích; jinak stanovené, vztah typ-projev, T/T infometrie. Poznámka, že průměrný počet položek na zdroj (projevy na typ) je dán:

$$\mu = \frac{A}{T} = \frac{\int_1^{\rho(A)} jf(j) dj}{\int_1^{\rho(A)} f(j) dj}.$$

Infometrie type/token-taken je získána nahrazením funkcí f funkcí g :

$$g(j) = jf(j)$$

$$j \in [1, \rho(A)]$$

popisující projev hustoty místo typu hustoty: $g(j)$ označující hustotu projevu v typech s hustotou projevu j . V diskretním uspořádání to by znamenalo (nahrazení $j \in \mathbb{N}$): zlomek položek, který náleží zdrojům s n položek.

4.1. Racionálnost formule

Formule $g(j) = jf(j)$ je základ pro infometrii typu type/token-taken. Jdeme zde zpět k uspořádání myšlenek, k příkladu (dané v předešlé části), který se týká vyhledávání klíčů, které používají katalogizátoři. Toto vysvětlení může být použito pro jakoukoliv aplikaci. Tak předpokládejme, že f je frekvence zákona vyhledávání klíčů v katalogu, $f(n)$ je počet vyhledání klíčů, jejichž výskyt $n \in \mathbb{N}$ krát v katalogu. Nyní katalogizátor používá tyto vyhledávací klíče a pak postupujeme na třetí úroveň („taken“ úroveň). Každou dobu katalogizátor chce vytvořit v katalogu popis nové knihy, vyhledávací klíč této knihy je zase zkontrolován v již existujícím katalogu (vyjádřené v termínech vyhledávacích klíčů). Položky jsou zde používány jako zdroje (proto jméno „taken“):

v katalogu máme, že $f(n)$ je počet použitých klíčů, které se objeví n krát. Jinými slovy, třetí úroveň infometrie má funkci g (formule $g(j) = jf(j)$) jako funkce frekvence a není f .

Průměrné užití μ^* je získáno jako parametr označující pravděpodobnost těchto projevů (v typu hustoty projevu j). Takový typ hustoty projevu je nazýván j . Proto průměrný počet μ^* užitých projevů na ρ je roven

$$\mu^* = \int_1^{\rho(A)} jP^*(j) dj$$

což vede k formuli

$$\mu^* = \frac{\int_1^{\rho(A)} jg(j) dj}{\int_1^{\rho(A)} g(j) dj} = \frac{\int_1^{\rho(A)} j^2 f(j) dj}{\int_1^{\rho(A)} jf(j) dj}$$

Předešlé rovnice jsou základem pro type/token-taken infometrii (T/TT infometrie). Ta měří plnění průměrného počtu položek na zdroj uživatelů, spíše než průměrný počet μ položek na zdroj ve smyslu výskytu.

Tuto sekci uzavřeme konkrétní formulí pro výchozí rozložení f a g . Protože rovnice

$$\int_1^{\rho(A)} f(j) dj = T$$

implikuje (protože $\alpha > 1$)

$$\int_1^{\rho(A)} f(j) dj = \frac{D}{\alpha - 1} \left(1 - \frac{1}{\rho(A)^{\alpha-1}} \right) = T.$$

máme za to, že pravděpodobnost rozložení f je dané

$$P(j) = \frac{f(j)}{\int_1^{\rho(A)} f(k) dk} = \frac{1 - \alpha}{\rho(A)^{1-\alpha} - 1} \frac{1}{j^\alpha}.$$

protože

$$A = \int_1^{\rho(A)} g(j) dj = \int_1^{\rho(A)} jf(j) dj = \frac{D}{2 - \alpha} (\rho(A)^{2-\alpha} - 1)$$

jestliže $\alpha \neq 2$ a

$$A = \int_1^{\rho(A)} g(j) dj = \int_1^{\rho(A)} jf(j) dj = D \ln \rho(A)$$

a jestliže $\alpha = 2$ máme pro pravděpodobnost rozložení P^* g následující formuli

$$P^*(j) = \frac{g(j)}{\int_1^{\rho(A)} g(k) dk} = \frac{jf(j)}{\int_1^{\rho(A)} kf(k) dk}$$

jestliže $\alpha \neq 2$

$$P^*(j) = \frac{\frac{D}{j^{\alpha-1}}}{\frac{D}{2-\alpha} (\rho(A)^{2-\alpha} - 1)}$$

jestliže $\alpha = 2$

$$P^*(j) = \frac{\frac{D}{j^{\alpha-1}}}{D \ln \rho(A)}$$

Proto

$$P^*(j) = \frac{2-\alpha}{\rho(A)^{2-\alpha} - 1} \frac{1}{j^{\alpha-1}}$$

jestliže $\alpha \neq 2$

$$P^*(j) = \frac{\ln \rho(A)}{j}$$

jestliže $\alpha = 2$ pro všechna $\alpha > 1$.

T/T infometrie popisuje vztah zdroj-položka jako takový, zatímco infometrie T/TT popisuje vztah zdroj-položka tak, jak to je vyjádřeno uživateli (informačními profesionály stejně tak jako těmi, kteří hledají informace). Je jasné, že další upřesnění tohoto modelu je důležité.

T/T a T/TT infometrie byly porovnávány užitím významných T/T a T/TT eventuelně průměrů μ a μ^* průměrů s následujícími výsledky:

Teorém III.1: $\mu^* \geq \mu$ ve všech případech.

Důkaz: jeden důkaz může být dán vyjádřením, že rozdíl rozložení P je pozitivní:

$$\int_1^{\rho(A)} (j - \mu)^2 P(j) dj \geq 0.$$

Jiný jednoduchý důkaz, navržený jedním ze soudců je jako následující: vyjádření, že příští kvadratická forma (v x) je vždy pozitivní:

$$0 \leq \int_1^{\rho(A)} (xj + 1)^2 f(j) dj$$

$$= x^2 \int_1^{\rho(A)} j^2 f(j) dj + 2x \int_1^{\rho(A)} j f(j) dj + T.$$

tudíž, rozlišení této rovnice musí být negativní, poskytující $\mu^* \geq \mu$.

Výsledek teorému III.1 může být hodnocen v infometrických termínech jako následující: průměrný počet položek na zdroj sledovaný uživatelem je větší než reálný existující průměrný počet položek na zdroj. Katalogizátor najde větší zobrazené délky (na průměr), než průměrný počet v databázi (negativní konstanta). Knihovní uživatel, vyhledávající informace z databáze, užívající klíčová slova, najde více položek než by mohlo být předpokládáno z databáze samotné (dle IR hodnotících kritérií „zpětného vyvolání“ a „přesnosti“). To by mohlo být interpretováno jako pozitivní respektive negativní prohlášení. Další dva příklady dávají negativní interpretaci: uživatel najde více půjčených knih (proto nedostupné), než by předpokládal a stejně to platí pro čekací dobu u pultu!

Hodnocení přesných hodnot μ a μ^* .

$$\mu = \frac{A}{T} = \frac{1 - \alpha \rho(A)^{2-\alpha} - 1}{2 - \alpha \rho(A)^{1-\alpha} - 1}$$

jestliže $\alpha \neq 2$ (jako vždy $\alpha > 1$). Jestliže $\alpha = 2$ máme:

$$\mu = \frac{A}{T} = \frac{\ln \rho(A)}{1 - \frac{1}{\rho(A)}}$$

μ^* hodnotíme:

$$\int_1^{\rho(A)} j^2 f(j) dj = \int_1^{\rho(A)} \frac{D}{j^{\alpha-2}} dj.$$

Předpokládáme $\alpha \neq 2$, $\alpha \neq 3$, pak se rovnice rovná:

$$\int_1^{\rho(A)} j^2 f(j) dj = \frac{D}{3 - \alpha} (\rho(A)^{3-\alpha} - 1)$$

Jestliže $\alpha = 2$ máme:

$$\int_1^{\rho(A)} j^2 f(j) dj = D(\rho(A) - 1)$$

Jestliže $\alpha = 3$

$$\int_1^{\rho(A)} j^2 f(j) dj = D \ln \rho(A).$$

To dává pro μ^* : jestliže $\alpha \neq 2, \alpha \neq 3$

$$\mu^* = \frac{2 - \alpha}{3 - \alpha} \cdot \frac{\rho(A)^{3-\alpha} - 1}{\rho(A)^{2-\alpha} - 1}$$

Jestliže $\alpha = 2$, nacházíme, že:

$$\mu^* = \frac{\rho(A) - 1}{\ln \rho(A)}$$

a konečně, jestliže $\alpha = 3$ máme:

$$\mu^* = \frac{\ln \rho(A)}{1 - \frac{1}{\rho(A)}}$$

Z formule pro μ a μ^* vidíme, že přímý vztah mezi μ a μ^* není derivovaný (protože

$$\mu = \frac{A}{T} = \frac{1 - \alpha}{2 - \alpha} \frac{\rho(A)^{2-\alpha} - 1}{\rho(A)^{1-\alpha} - 1}$$

nemůže být vyřešeno přesně pro $p(A)$ —to může být vyřešeno numericky a to bude téma další sekce, udávající konkrétní hodnoty pro μ^* ve funkci μ a α).

Nicméně jsme z dřívějšího schopni dokázat následující výsledek:

Teorém III.2: Pro každé α fixované máme, že μ^* je zvyšující se funkce μ .

Aby došlo k objasnění z následující sekce, můžeme dohadovat následující, ale nejsme schopni nalézt důkaz pro to:

Konjektura III.3: pro každé fixované μ máme, že μ^* je zvyšující se funkce α . Také konstatujeme následující:

Problém III.4: v teorii T/TT infometrii byl užíván Lotkův zákon jako základní zákon frekvence. Vývoj T/TT infometrické teorie je pro další akceptovatelné frekvence rozložení (pokračující nebo ne) takový jako Poissonovo rozložení. Negativní binominální distribuce, logaritmické normální rozložení ...

4.2. Konkrétní výpočet T/TT průměru μ^* ve funkci T/T průměr μ a Lotkovo α

Vypočítání μ^* ve funkci μ a α požadující řešení následující rovnice v $x = p(A)$ jestliže $\alpha \neq 2$ viz rovnice dříve.

$$x^{2-\alpha} \left(\frac{1}{\mu} \frac{\alpha - 1}{\alpha - 2} \right) - x^{1-\alpha} + 1 - \frac{1}{\mu} \frac{\alpha - 1}{\alpha - 2} = 0$$

a jestliže $\alpha \neq 2$

$$\ln x + \frac{\mu}{x} - \mu = 0.$$

Jedno x je nalezeno, máme poté μ^* užívající formuli: jestliže $\alpha \neq 2$, $\alpha \neq 3$

$$\mu^* = \frac{2 - \alpha x^{3-\alpha} - 1}{3 - \alpha x^{2-\alpha} - 1}$$

jestliže $\alpha = 2$

$$\mu^* = \frac{x - 1}{\ln x}$$

jestliže $\alpha = 3$

$$\mu^* = \frac{\ln x}{1 - \frac{1}{x}}$$

- Z předešlých rovnic je jasné, že vymezení přímého funkčního vztahu $\mu^* = \mu^*(\mu, \alpha)$ není možné: dané α a μ , můžeme jen rozluštit rovnice

$$x^{2-\alpha} \left(\frac{1 - \alpha - 1}{\mu - \alpha - 2} \right) - x^{1-\alpha} + 1 - \frac{1 - \alpha - 1}{\mu - \alpha - 2} = 0$$

a

$$\ln x + \frac{\mu}{x} - \mu = 0.$$

numericky. α a μ nemohou být plně nezávislé na sobě; existuje zde dost svobody, ale následující omezení musejí být respektována:

$$A = \int_1^{\mu(A)} j f(j) dj$$

$$= \int_1^{\mu(A)} \frac{D}{j^{\alpha-1}} dj$$

$$< \int_1^{\infty} \frac{D}{j^{\alpha-1}} dj.$$

Poslední rovnice

	$\alpha = 1.5$	$\alpha = 2$	$\alpha = 2.5$	$\alpha = 3$	$\alpha = 3.5$
$\mu = 1.2$	1.2148	1.2151	1.2157	1.2164	1.2178
$\mu = 1.5$	1.5838	1.5984	1.6177	1.6479	1.7010
$\mu = 2$	2.3317	2.4612	2.7320	—	—
$\mu = 2.5$	3.2492	3.7279	5.8541	—	—
$\mu = 3$	4.3333	5.5996	—	—	—
$\mu = 3.5$	5.5894	8.3955	—	—	—

Tabulka : hodnoty μ^* ve funkce μ a α ²⁴

Poznámka, že $\mu^* > \mu$ se zvyšuje v μ a μ . Poznámka také, že rozdíl $\mu^* - \mu$ se zvyšuje s μ . Další hodnoty μ^* ve funkce dané μ s a α s mohou být vypočítány podobně.

$$D \left[\frac{j^{2-\alpha}}{2-\alpha} \right]_1^\infty$$

Jestliže $\alpha \neq 2$ a $D \ln j \Big|_1^\infty$ jestliže $\alpha = 2$. Proto máme ∞ jestliže $\alpha \leq 2$,

$$\frac{D}{\alpha - 2}$$

jestliže $\alpha > 2$.

Jestliže přibližně T (protože $\alpha > 1$ a $p(A)$ je větší)

$$T = \int_1^{p(A)} f(j) dj$$

$$\approx \int_1^\infty f(j) dj$$

$$= \frac{D}{\alpha - 1}$$

zjistili jsme kombinací těchto dvou předešlých rovnic, že neexistují žádná omezení pro μ , dané $\alpha \geq 2$, ale že jestliže $\alpha < 2$ musíme mít, že

$$A < T \frac{\alpha - 1}{\alpha - 2},$$

proto

$$\mu < \frac{\alpha - 1}{\alpha - 2}.$$

²⁴ EGGHE, Leo. Type/token-taken informetrics

Pro tabulku μ^* ve funkci z μ a α (pro důkaz této nerovnosti, bez užití aproximace pro T) máme následující omezení:

$$\alpha = 2.5 \Rightarrow \mu < 3$$

$$\alpha = 3 \Rightarrow \mu < 2$$

$$\alpha = 3.5 \Rightarrow \mu < 1.6667$$

To vysvětluje prázdnou část v tabulce. Tabulka je získána vyřešením rovnic

$$x^{2-\alpha} \left(\frac{1-\alpha-1}{\mu} \frac{\alpha-1}{\alpha-2} \right) - x^{1-\alpha} + 1 - \frac{1-\alpha-1}{\mu} \frac{\alpha-1}{\alpha-2} = 0$$

a

$$\ln x + \frac{\mu}{x} - \mu = 0$$

pro x a poté dosazení x v rovnicích

$$\mu^* = \frac{2 - \alpha x^{3-\alpha} - 1}{3 - \alpha x^{2-\alpha} - 1}, \quad \mu^* = \frac{x-1}{\ln x} \quad \text{a} \quad \mu^* = \frac{\ln x}{1 - \frac{1}{x}}$$

Další informace viz EGGHE, Leo. Type/token-taken infometrics.²⁵

²⁵ EGGHE, Leo. Type/token-taken infometrics

5. Budoucnost citačního indexování

Garfield řekl, že pokud všechny indexované a abstrahované služby přestanou být předmětem zájmu podnikání (budou nahrazeny strojovým indexováním), knihovny budou stále potřebovat přístup k jejich existujícím souborům literatury, která stále není pokryta elektronickými službami. Národní citační indexy jako čínské SCI nebo brazilské SCI by byly dostupné pro další regiony. V budoucnu by mechanické překládací systémy uspokojily neanglické uživatele tituly v jejich vlastních jazycích. Garfield by také rád viděl citační indexování zdrojů knih a dizertací, ale to je téma pro budoucí diskuzi.

Celý world wide web, který je protkán miliardami odkazů (technický ekvivalent citovaným odkazům) mezi digitálními dokumenty, je založen na stejném konceptu jako tištěné časopisy a články.

Research Index (dříve Cite-Seer) a CiteBase jsou pracující prototypy pro dokumenty volně dostupné přes web. Tyto modely jsou schopné eliminovat nebo redukovat nejvíce nákladnou část citačního indexování – lidskou laboratoř - náklady, které jsou velmi vysoké u tradičních citačních indexů.

Research Index znamená nepochybně velký pokrok – zvláště pokud bude rozšířen tak, aby pokrýval celou oblast vědy a technologií. Je však závislý na časopisech a dalších materiálech, které jsou dostupné volně na webu. To je významné omezení, které může být překonáno, jestliže komerční vydavatelé zpřístupní jejich plné texty těm, kdo spravují nové databáze vytvořené daným softwarem. Nicméně existují problémy dokonce u materiálů, které jsou volně dostupné.

Jiným aspektem myšlenky volného přístupu k indexům je otázka vzdělání uživatelů. V komerčním světě to představuje část marketingového úsilí. Jestliže Research Index bude obchodován tímto způsobem, bude mít mnohem větší podporu.

Henk Moed v komentáři v časopise Nature podotkl, že 7 procent chyb je založeno na užívání automatických procedur při porovnávání ISI zdrojových dat se souborem citovaných odkazů. Jsou to chyby ve stránkování k citacím z dalších než prvních stránek, jak je časté např. v chemii. Nejcitovanější práce v historii vědy od Olivera H. Lowryho a spolupracovníků byla citována 270 000 krát. (další tři nebo čtyři práce byly citovány více jak 100 000 krát). Dalších 10 000 citací v Lowryho práci obsahuje varianty a chyby kolem 4 procent. Člověk se může zeptat, kdo by chtěl vyhledat tak mnoho záznamů, ale v kombinovaném vyhledávání s klíčovými slovy nebo dalšími citovanými pracemi, jako je kocitačním vyhledávání, je to přiměřené a možné.

V oblasti umění a humanitních věd by bylo možné předpokládat mnohem nižší počet článků z časopisů, ale při kontrolování citovaných knih musí člověk vzít v úvahu různé citace odkazující na zkrácené tituly knih, stejně tak jako citace jednotlivých stránek.

Myšlenka online **National Citation Facility** (Národní citační databázi) pro autory k ověřování jejich citací, může být uskutečněna na základě spolupráce s CrossRef, který má okolo 10 milionů identifikátorů digitálních objektů (kromě tradičních elementů bibliografických dat), které mohou být zpřístupněny zdarma (jednou za čas) kýmkoliv. Moderní ekvivalent pro „Forever dictionary“ (slovník navždy), o kterém bylo informováno v jedné z Garfieldových esejí, je okolo 33.5 milionu položek, což je dokonce více než 32.8 milionu z let 1945-2003 zdrojových položek v SCI/SSCI/AHCI.

Rozšíření CrossRef je velmi výhodné, ale je možné pochybovat, že to bude to samé jako National Citation Facility. DOI CrossRef nalezne plný text jednoho článku. ISI také

používá DOI a společně se systémem Webfeat umožní integrovat bez problémů nepoškozený soubor databází v každé instituci.

Člověk by neměl opomenout velikost procesu korekce chyb. Mnozí z nás ani netuší, jak mnoho chyb je skutečně opraveno. ISI nedávno rozšířilo proces editování, a to tak, že v roce 2004 bude dokonce vykonáno daleko více oprav než dosud.

Na stránkách <http://eugengarfield.org> a <http://garfield.library.upenn.edu/histcomp> najdeme databáze HistCite. Ty umožňují uživateli jednoduchou cestou identifikovat zdrojovou literaturu vyhledávanou ve Web of Knowledge označeném seznamu a umožňují seřadit materiály v čistém chronologickém pořadí, což Web of Science nyní dělá přibližně v opačném chronologickém pořadí. HistCite také pomáhá uživateli editovat variace v citovaných odkazech. Také vytváří historiografii (mapy) klíčové literatury, aby pomohl uživateli vizualizovat historii pro hodnocené téma a vytvořit si přehled vývoje předmětu.²⁶

²⁶ JACSO, Peter. The future of citation indexing

6. Citační analýza

6.1. Vymezení pojmu citační analýza a využití citační analýzy

Citační analýza je obecný termín pro techniky, které mají dlouhou historii v bibliometrických studiích vědecké komunikace. Citace mohou odhalit formální komunikační modely. Metody citační analýzy jsou nevtrhavé a mohou být vysoce spolehlivé. Existují ovšem další způsoby hodnocení vlivu časopisů. Vydavatelé mohou získat informace o využití časopisů sledováním počtu předplatitelů - u elektronických časopisů počítáním shlédnutí článků. Tato data poskytují užitečné ukazatele o časopisech, ale jsou nedostatečná pro sledování skutečného vlivu článků časopisů ve vědě, výzkumu a tvorbě znalostí.²⁷

Všeobecně není akceptována teorie, která vysvětluje, proč autoři citují další práce. Brooks identifikoval několik motivací, jedna z nich demonstruje důležitost znalostí práce v dané oblasti. Některými dalšími jsou: a) úcta (hold dřívějšímu profesorovi, vydavateli časopisu nebo dalšímu vědci); b) oslnění diváka (víra, že počet citací zvýší kvalitu jinak nevyvíkajícímu článku; c) aktualizace; d) ozdoba.²⁸

Garfield charakterizoval základní podstatu citační analýzy jako „... velmi obecné měření úrovně příspěvku jednotlivce k rozšíření vědeckých poznatků...“ To by mohlo být nazýváno vlivem **jednotlivce** nebo individuální práce. Práce, která je opětovně citována, je užitečná pro vědeckou komunitu. Její prospěšnost je vysoká a je relevantní pro další výzkumníky. Jedinec může cítit pokušení charakterizovat takovou práci jako „významnou“ nebo „důležitou“, ale přesný význam těchto atributů není jednoduché specifikovat. Citace jako takové nejsou indikátory kvality a významu.

Nízká hodnota citačních indikátorů nedokazuje, že práce nebo jedinec jsou „nevýznamní“. Nízká hodnota může být sice jedním důvodem pro nízkou citační míru, ale mohla by znamenat také jiné negativní důvody (např. mimořádné práce průkopníků, které odporovaly premisám normální vědy, by mohly být zamítnuty a zapomenuty, protože svým způsobem předběhly dobu). Nízká citační míra by mohla být přiměřená absenci výzkumné komunity v konkrétním vzácném nebo specializovaném oboru. Není nikdo okolo, kdo by se mohl zajímat o danou věc a mohl tak odkazovat na naše skvělé myšlenky. Navíc se ukazuje, že výzkumníci z malých a vzdálených zemí musejí překonávat mnoho překážek, jestliže se chtějí stát významnými vedoucími mezinárodními vědci oproti vědcům ve velkých zemích.

Při vymezování citační míry je nutné vzít v úvahu, jaké publikace jsou indexované na prvním místě. Výzkumníci, kteří hlavně publikují jejich výsledky v neindexovaných publikacích, nevyhnutelně získají málo citací. Ale také kulturní a jazykové omezení humanitních věd jsou vážným problémem. Většina „malých“ kultur a jazyků nejsou pokryty citačními rejstříky.

V praxi jsou často vědecké výsledky studií publikovány v příslušném jazyce v časopisech nebo knihách, které nejsou sledovány v indexech. Což je problém hlavně

²⁷ HARTER, Stephen P., The Impact of electronic journals on scholarly communication : a citation analysis.

²⁸ DOMINICK, Joseph R., Citation analysis of the Journal of Broadcasting & Bectric Media : another perspective.

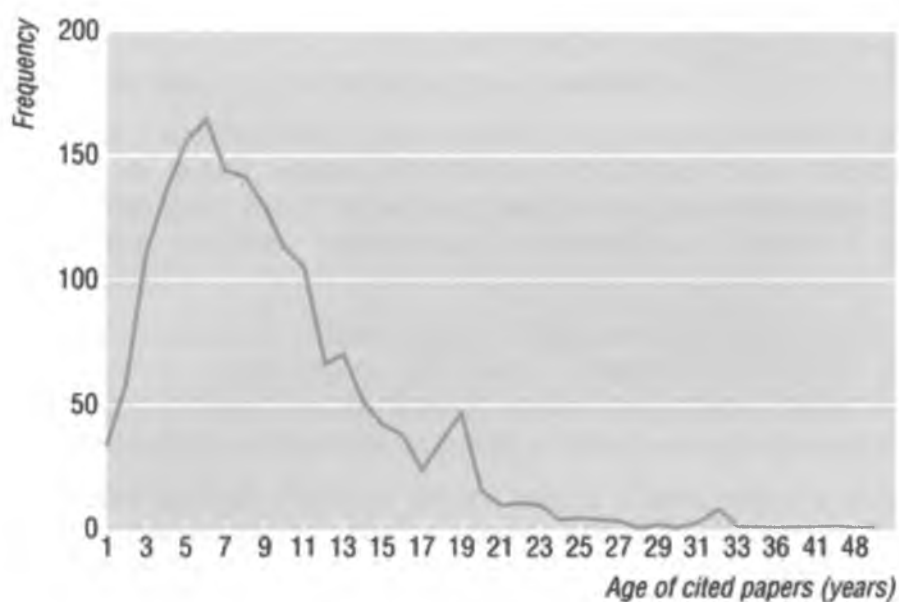
u typicky kulturních nebo regionálních institucí. Jestliže se takové disciplíny stanou zájmem správné bibliometrické analýzy, musejí být analyzovány odděleně.²⁹

Celkově nashromážděné citace ke všem článkům publikovaných časopisem jsou indikátorem vlivu daného časopisu na relevantní disciplínu (disciplíny).

Institute for Scientific Information (ISI) publikuje tři citační indexy, které umožňují vyhledávání článků podle jmen autorů nebo názvu článku. „Citace“ je v ISI interpretována jako unikátní odkaz na práce v citujícím článku. Citující článek (nazývaný „source article“) může zahrnovat několik rozdílných citací na článek ve stejném citovaném časopise. Každý z nich je započítán jako separátní citace. Počet citujících článků musí být menší nebo stejný než počet citací k práci.³⁰

6.2. Doba oběhu znalostí

„**Doba oběhu znalostí**“ je doba, která uplyne, než je práce citována v např. klinických návodech. Průměrná doba oběhu znalosti je pro všechny práce z oblasti přírodních věd 8 let. Z toho je 25% prací starších více než 10 let a 4% více než 25 let.



Doba oběhu znalostí citovaných v klinických návodech³¹

Je zajímavé pozorovat potřebnou dobu pro proniknutí základního výzkumu do klinické praxe. Průměrná doba potřebná pro to, aby práce byla citována, činí tři roky. Průměrná doba mezi publikováním práce a jejím zahrnutím v příručce je 8 let. Práce musí překonat 3 hranice generací.

Práce vznikají na úrovni 1. Poté jsou citovány pracemi o úroveň vyšší. Trvá 17 let, než se poznatky z práce vzniklé na úrovni 1 objeví na úrovni 4 ($8 + (3 \times 3) = 17$). V souvislosti s tím vyvstává otázka, jaká je optimální doba k tomu, aby výzkum byl plně

²⁹ KARLSSON, Fred, Linguistics in the light of citation analysis.

³⁰ HARTER, Stephen P., The Impact of electronic journals on scholarly communication: a citation analysis.

³¹ GRANT, Jonathan [et al.], Evaluating „payback“ on biomedical research from papers cited in clinical guidelines: applied bibliometric study.

ohodnocen a byl zaveden do praxe a zda tento proces je možno urychlit a zda je to potřebné.

Citační analýza pomáhá autorům identifikovat časopisy relevantní pro jejich výzkum a hledat časopisy, ve kterých by mohli publikovat. To umožňuje vydavatelům časopisů odhadnout efektivitu vydavatelské politiky a umožňuje sledovat postavení jejich časopisů.

Citační analýza má však svá omezení a neměla by být užívána jako jediný zdroj informací, podle nichž porovnáváme a hodnotíme publikace. Impakt faktory nejsou statisticky reprezentativní pro individuální články časopisů a mají špatný vztah s aktuální citací jednotlivých článků. Přehledy a dlouhé články získávají více citací, zatímco krátké, jako případové zprávy, jsou často méně citovány. Databáze **Journal Citation Reports ISI** dává přednost časopisům v angličtině, které mají vyšší impakt faktor než časopisy v jiných jazycích. Často jsou zahrnovány také autocitace.

Medicínské časopisy se širším publikem a větším oběhem (např. *British Medical Journal* a *Lancet*), dosahují vyššího impakt faktoru, než menší specializované časopisy.

V citacích také existuje vysoká chybovost - např. v časopisech otolaryngologie až 37%. Citační analýza by proto měla být doplňována dalšími metodami hodnocení časopisů, např. sledováním přehledů a zjišťováním názoru specialistů.³²

Disciplína se může z vnějšku zdát jako složitá, ale citace původních prací nabízejí objektivní přístup. A právě citace sledované v průběhu času mohou nabídnout přesnější měření. Sledování citací může také nastítnit vliv jedné disciplíny na druhou. Je možné takto nalézt počátek originálních myšlenek a působení publikací na akademickou komunitu.

Zmapování důležitých publikací v čase může odhalit modely, které se ne vždy snadno jeví mezi množstvím současných publikací. Zmapování těchto pravidelně a mezinárodně citovaných textů může odhalit hlavní vstupy do disciplíny. Porovnání vstupů a výstupů z disciplíny odhalí její podstatu a může pomoci nám porozumět jí.³³

Příklad využití citační analýzy můžeme čerpat také z oblasti práva a právní literatury. Citace používané v zákonodárství obvykle hledají podklad v argumentu nebo analýze. Pokud tedy někdo odkazuje ve svém prohlášení na určitý „zákon“, mělo by být umožněno čtenáři či posluchači poznat daný zákon v originální formě.

Můžeme se také setkat se zneužitím citací ve dvou formách:

- citování práva, které není aplikovatelné na daný případ,
- citování dokumentu, který není právním dokumentem.³⁴

Pokud se budeme podrobněji zabývat otázkou vícenásobného autorství, budeme se zabývat otázkou spolupráce autorů. Spolupráce autorů (často dvou autorů) má dlouhou a plodnou historii. V nedávných letech, nicméně, pozorujeme, že spolupráce autorů se zvyšuje od počtu dvou autorů ke třem, čtyřem a někdy i více autorům - jak je možné vidět např. v lékařských časopisech a časopisech zabývajících se veřejným zdravím.

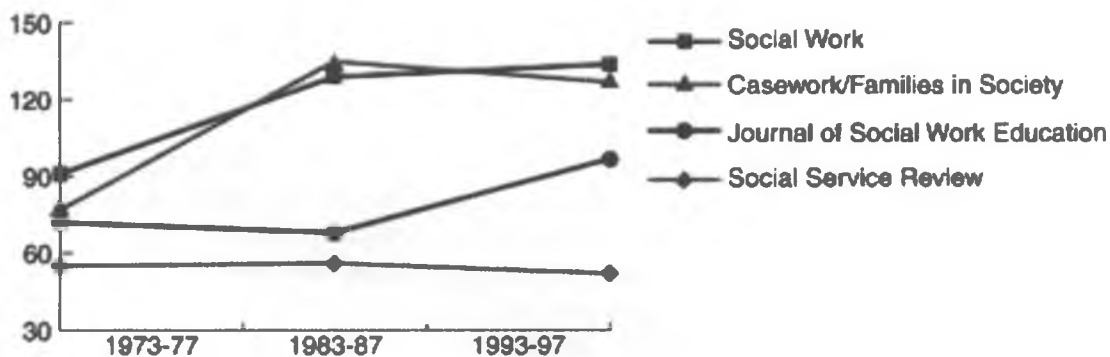
³² GRANT, Jonathan [et al.], Evaluating „payback“ on biomedical research from papers cited in clinical guidelines: applied bibliometric study.

³³ KATZ, J. Sylvan; HICKS, Diana, How much is a collaboration worth? A calibrated bibliometric model. HARTER, Stephen P., The Impact of Electronic Journals on scholarly communication: a citation analysis.

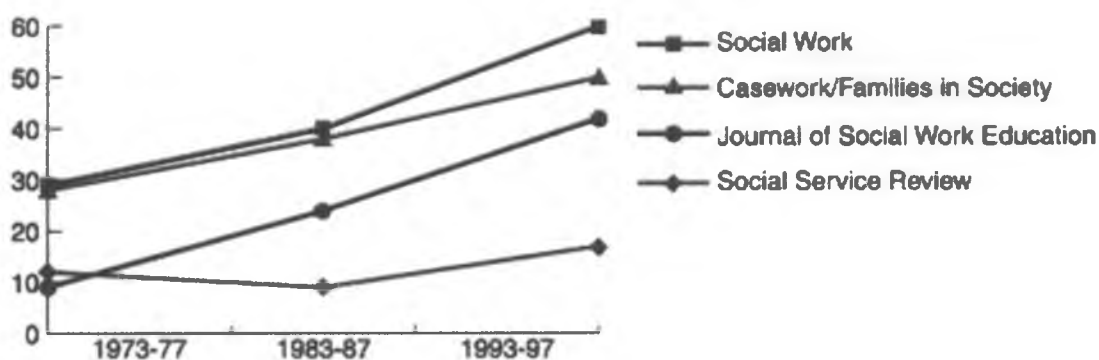
³⁴ GALLAGHER, Hill, Misuse of citations is a dangerous game.

Institute for Scientific Information ve Philadelphii zjistil na základě analýzy 4 000 vědeckých časopisů, že počet prací s 50 a více autory se zvýšil ze 49 v roce 1981 ke 407 v roce 1994 a počet článků se 100 a více autory vzrostl z 1 k 182 za stejné časové období. V roce 1993 *The New England Journal of Medicine* publikoval výsledky velkého klinického výzkumu, ve kterém bylo na seznamu 972 autorů.

Obrázek **Počet článků se dvěma a více autory** zobrazuje počty článků s vícenásobným autorství (2 nebo více autorů), zatímco obrázek **Počet článků se třemi a více autory** zobrazuje počty článků se třemi a více autory. Ačkoliv sledujeme velmi malý počet článků s více než 4 autory v těchto časopisech (jen 30 z 2 769), existuje velké zvýšení v počtu dvoj- a troj-autorství článků.



Počet článků se dvěma a více autory ³⁵



Počet článků se třemi a více autory ³⁶

Jeden z hlavních důvodů velkého nárůstu vícenásobného autorství článků je důraz na produktivitu vědců. Tento důraz je dokládán velkým množstvím literatury.

Vědci sami pochybují o své schopnosti vytvářet práce nezávisle a jsou vystaveni tlaku častěji publikovat. Tím jsou „blokováni“ od výzkumné práce. Spolupráce může představovat pro ně potřebný popud. Zvyšuje se pravděpodobnost citování článku

³⁵ GELMAN, Sheldon R.; GIBELMAN, Margaret, A quest for citation? An analysis of and commentary on the trend toward multiple authorship.

³⁶ GELMAN, Sheldon R.; GIBELMAN, Margaret, A quest for citation? An analysis of and commentary on the trend toward multiple authorship.

v životopise, úřady a rady zabývající se udělováním titulů mohou lépe zaznamenat jejich práci a mohou takto zvýšit možnou pozornost k jejich práci.³⁷

Studie v informační vědě odhalují, že většina citací má tendenci odkazovat na články, které jsou "nové". Budd zjistil, že průměrný věk citovaných materiálů je 10,29 let. To naznačuje, že rychle se vyvíjející obory mají tendenci citovat práce, které jsou relativně nové. Články starší než 20 let nejsou pravděpodobně nikdy citovány.³⁸

+V oblasti farmaceutického výzkumu je počet autocitací větší v technologické oblasti než ve vědě. Velké prestižní univerzity jsou častěji citovány organizacemi ve vědecké literatuře, které navazují na předešlé studie a získávají tak podporu pro pokračující veřejné financování tohoto typu institucí. To umožňuje lepší užití veřejných znalostí. Spolupráce a partnerství mezi vědci a veřejnými výzkumníky by mělo být podporováno. Avšak spolupráce průmyslu a univerzit je komplikovaná a má mnoho smíšených výsledků. Veřejné instituce nejsou obzvláště citované v technologické oblasti, ale soukromé firmy ano.³⁹

Patentová analýza je založena na průzkumu citačních vztahů mezi rozdílnými patenty, mezi patenty a vědeckou literaturou. Jestliže vynálezce žádá o patent, musí dokázat, že jeho vynález je nový, použitelný a nejedná se jen o průměrnou znalost. Ten, kdo žádá o patent, na základě citací vysvětluje, jak nový patent zlepšuje dřívější vynálezy. Společnosti s velkým počtem citovaných patentů pravděpodobně ovládají technologie, které jsou ústřední pro rozvoj v dané oblasti průmyslu. Ne všechny důležité patenty jsou však často citovány a ani každý často citovaný patent není tak důležitý.

Citační vztahy mezi patenty také indikují rychlost, se kterou se společnosti inovují. Společnosti, které citují relativně nové patenty, jsou pravděpodobně rychlejší v inovování než společnosti, které citují starší patenty.⁴⁰

Stigler a Freidland naznačili, že citační analýza trpí tzv. iluzí, která vznikla vymizením známých předků. Úspěšný příspěvek se může stát tak rozšířeným, že jeho autorství je z počátku známo, ale později je zapomenuto.

Další otázkou je, jak je důležité vědět některé informace, které se vztahují k vědcům. Jestliže vědec získá Nobelovu cenu, průměrný student pravděpodobně porozumí důležitosti myšlenek, které jsou příčinou získání takovéto ceny. Je však sporné, zda je pro průměrného studenta důležité vědět, s kterou významnou univerzitou je nebo byl vědec ve spojení. Uvažuje se, že tyto informace jsou uváděny spíše více ve prospěch přednášejícího, který vytváří daný text, než pro studenta, který text čte.⁴¹

Ve všech akademických disciplínách vznikají výzkumné skupiny, které se zaměřují na společné otázky stejnými způsoby. V rámci těchto skupin mohou být pojmy přisuzované jednomu výzkumníkovi použity dalšími pro výzkum, rozšíření nebo

³⁷ GELMAN, Sheldon R.; GIBELMAN, Margaret, A quest for citation? An analysis of and commentary on the trend toward multiple authorship.

³⁸ REDMAN, Arnold L.; MANAKYAN, Herman; TANNER, John R. , The ranking of real estate journals: a citation analysis approach.

³⁹ MCMILLAN, G. Steven; HAMILTON, Robert D., III, Using bibliometrics to measure firm knowledge : an analysis of the pharmaceutical industry.

⁴⁰ BREITZMAN, Anthony ; THOMAS, Patric, Using patent citation analysis to target/value M & A candidates.

⁴¹ HOAAS, David J.; MADIGAN, Lauren J., Citation analysis of economists in principles of economics textbooks.

zdokonalení. Práce jednoho výzkumníka navazuje na další na základě citace, na základě historické vazby. Když jeden vědec cituje dřívější práci jiného, citace se stávají určitým způsobem dokumentování historie. Recenzované akademické časopisy hrají stále větší roli v rozšiřování vědeckých informací, zvláště v posledním desetiletí. Kromě toho, kategorie časopisů, v nichž jsou články publikovány, mohou ovlivnit zvýšení prestiže jedince, mohou pomoci k získání určitého místa a zvýšení platu .⁴²

Další informace k citační analýze lze kromě řady článků a publikací nalézt v diplomové práci, na kterou tato práce rigorozní navazuje.

⁴² ROMANO, Claudio; RATNATUNGE, Janek, Citation analysis of the impact of journals on contemporary small enterprise research.

7. Impakt Faktor (IF; faktor vlivu)

Impakt faktor (IF) vyjadřuje citovanost časopisu. Vyjadřuje, kolikrát je článek uveřejněný v daném časopise průměrně citován. Vypočítá se jako podíl mezi celkovým počtem citací v daném roce vztahujících se ke všem článkům publikovaných v časopisu v předchozích dvou letech, a celkovým počtem článků publikovaných v časopisu v předchozích dvou letech.

Jestliže časopis publikuje jen několik článků každý rok, nemůže ovlivnit mnoho dalších výzkumníků a spisovatelů a nemůže být citován velmi často. Impakt faktor normalizuje počet citací a nerozlišuje mezi časopisy podle počtu článků publikovaných v daném roce nebo jejich délkou života (respektive jejich citovanosti). Neměří příliš vliv časopisu, ale vliv typického článku objevujícího se v časopise. Časopis může mít vysoký impakt faktor, ale protože publikuje jen několik článků každý rok nebo právě začíná s publikováním, může mít malý celkový vliv na vědeckou komunitu.⁴³

IF za rok 2001 vypočítáme ze vztahu:

$$IF = \frac{\text{Počet citací v r. 2002 na články časopisu publikované v roce 2000 a 2001}}{\text{Počet článků publikovaných v časopisech v letech 2000 a 2001}}$$

Obecně se impakt faktor vypočítá:

$$IF = \frac{\text{počet citací v roce X na články časopisu publikované v roce X-1 a X-2}}{\text{počet článků publikovaných v časopisu v letech X-1 + X-2}}$$

Impakt faktor není konstanta, ale každým rokem může nabývat jiné hodnoty. Dnes je považován za objektivní kritérium kvality vědeckých časopisů a úspěšnost vědecké práce je hodnocena podle počtu publikací v časopisech s co nejvyšším IF.

Impakt faktor se vztahuje pouze k časopisům. Není proto možné počítat IF pro jednotlivé autory. Pokud více autorů publikuje jeden článek v impaktovaném časopisu, není možné jim tuto hodnotu rozdělit na části podle množství práce věnované publikaci.

Nejvyšší hodnoty vykazují časopisy víceoborové. IF je hodnota, která pomáhá zjistit postavení daného časopisu mezi ostatními excerpovanými časopisy, a je ukazatelem kvality v případě rozhodování o výběru časopisu, ve kterém by vědecký pracovník mohl publikovat.

IF však nemůžeme brát jako absolutní měřítko hodnocení kvality vědeckých časopisů. K výběru časopisů pro excerpci SCI hraje hlavní roli jazyk, ve kterém je časopis vydáván. Velmi významná výhrada se týká kvality citací.

Na druhé straně existují pracovníci, a nemusí to být zrovna vědci, kteří vypracovali určitou metodiku výzkumné práce, kterou využívají vědci v různých oborech a bez níž se neobejdou. Ti potom dotyčného pracovníka ve svých pracích, kde tuto metodiku použili, musí citovat. Tito pracovníci potom mohou mít ročně stovky citací.

Při hodnocení vědeckých pracovníků by citační analýza měla být pouze jedním z dílčích hledisek. Mělo by se přihlížet k tomu, jaký přínos měly práce hodnocených

⁴³ HARTER, Stephen P. The impact of electronic journals on scholarly communication: a citation analysis.

pracovníků pro naši vědu. Pak se může stát, že na vlastní badatelskou činnost vědci nebudou mít dostatek času, ale budou se zoufale snažit za každou cenu publikovat v impaktovaných časopisech, aby zhodnotili svou předchozí práci. Dosavadní vědní politika pak povede k tomu, že význační vědci přestanou publikovat v našich časopisech, čímž se jejich kvalita opět sníží.

Přes všechny uvedené výhrady neexistuje lepší metoda hodnocení vědeckých prací než impakt faktor. Proto je zcela pochopitelné, že v současné době je největší ctižádostí vědeckých pracovníků publikování v časopisech s co nejvyšším IF, ve kterých je možnost nejvyšší citovanosti.⁴⁴

Impakt faktor byl původně vytvořen jako výzkumný bibliografický nástroj pro vyhledávání překrývajících se výzkumů. Využívalo se toho pro hodnocení vědců, kteří pracovali relativně izolováni od svých kolegů s podobnými zájmy. Později byl rozšířen jako výzkumný nástroj pro oblast sociálních věd a v současné době je používám jako parametr pro hodnocení kvality vědecké práce vědců či skupin vědců.⁴⁵

Impakt faktory časopisů mohou pomoci nasměrovat čtenáře k těm časopisům, které mají významnou váhu publikovaných dat, které jsou pravidelně citovány a jsou tedy užitečné a mající vysokou informační hodnotu. Vypočítání přizpůsobeného impakt faktoru má malý efekt na řazení časopisů. Je tomu tak protože ne všechny časopisy z daných oborů zahrnují pravidelné rubriky, které zvyšují impakt faktor (např. dopisy, vydavatelské články o kontroverzních tématech nebo pozvání na zapojení se do diskuse).⁴⁶

Pro vědce i knihovníky je dlouhodobý vliv (např. deset let) významnější než krátkodobý vliv časopisu (dva roky). Některé časopisy jako *Cell*, *The New England Journal of Medicine*, *Proceedings of the National Academy of Sciences*, *Nature*, a *Science* vždy mají velký vliv, a to v jakékoliv době (dva, sedm či patnáct let). Vliv jiných časopisů však významně stoupá nebo klesá.

Stinnou stránkou „tradičního“ impakt faktoru je skutečnost, že citace nejsou vážené. Všechny citace jsou započítány jako stejně důležité, bez ohledu na citující časopis.⁴⁷

7.1. Další indikátory počítané ISI

7.1.1. Immediacy Index

Immediacy Index neboli "**Index bezprostředního vlivu**" (objevuje se i termín "Garfieldův Index") udává průměrný počet citací na články časopisu v roce, ve kterém byly publikovány.

$$II = \frac{\text{Počet citací v roce } X}{\text{Počet publikovaných článků v roce } X}$$

⁴⁴ PITTEROVÁ, Květa, Využití citačních rejstříků a impact factor.

⁴⁵ OPTHOF, Tobias, Sense and nonsense about the impact factor.
AMIN, M.; MABE, M., Impact factors: use and abuse.

⁴⁶ ROY, D. [et al.], Citation analysis of otorhinolaryngology journals.

⁴⁷ ROSSEAU, Ronald, Journal evaluation: technical and practical issues.

Immediacy index udává, jak rychle je článek po vydání v časopise citován. Můžeme předpokládat, že elektronické časopisy mají vyšší immediacy index, protože rychlost publikování je u nich vyšší.

Příklad: Immediacy index za rok 2002 je průměr počtu citací v roce 2002 ku počtu článků publikovaných v roce 2002.

7.1.2. Cited Half-life

"Poločas citovanosti" udává po kolika letech (ve vztahu k aktuálnímu roku) se objeví 50% všech citací na články daného časopisu v citačních rejstřících. Napomáhá při hodnocení stáří citovaných článků. Uvádí se pouze u časopisů, které byly alespoň 100 a vícekrát citovány.

7.1.3. Citing Half-life

"Poločas citování" napomáhá při hodnocení stáří většiny článků, jež cituje daný časopis. Uvádí se pouze u časopisů, které citují 100 nebo více dokumentů.

V databázi Journal Citation Reports se setkáme s termíny:

Source Date - informace o počtu tzv. "review" (přehledové články) a "non-review" článků, o počtu referencí v těchto člancích a o průměrném počtu referencí připadajících na jeden článek časopisu.

Cited Journal - abecední přehled časopisů, které citovaly vyhledaný titul.

Citing Journal - abecední přehled časopisů, které citoval vyhledaný titul.

Impact Factor Trend Graph - grafické zobrazení hodnoty IF v posledních pěti letech.⁴⁸

7.2. Citační rejstříky Institute for Scientific Information

7.2.1. Journal Citation Reports (JCR)

Journal Citation Reports je samostatná část SCI (Science Citation Index), a SSCI (Social Science Citation Index), ve které jsou obsaženy výsledky scientometrických analýz všech časopisů a publikací excerpovaných pro tyto rejstříky.

Journal Citation Reports vznikl v roce 1973 v papírové podobě. Později začal být distribuován na CD-ROM, v nedávné době i přes webovské rozhraní (Web of Knowledge). České vědecké veřejnosti je databáze přístupna online v rámci programu MŠMT "Informační zdroje pro výzkum a vývoj"

Údaje o časopisech, které JCR poskytuje, mohou posloužit:

- knihovníkům při akvizici a profilaci fondů,
- výzkumníkům a vědcům při hodnocení a porovnávání časopisů,
- vydavatelům při sledování postavení vlastních či konkurenčních časopisů,
- informačním analytikům při studiu citačních vazeb uvnitř i vně vědeckých disciplín.

⁴⁸ HARTER, Stephen P., The impact of electronic journals on scholarly communication: a citation analysis.
KALCU, Petra, Journal Citation Reports .

JCR obsahuje dvě řady:

- Science Edition (časopisy korespondující s databází Science Citation Index)
- Social Sciences Edition (časopisy korespondující s databází Social Sciences Citation Index)

Záznam časopisu obsahuje:

- ISSN
- Impakt Faktor
- Immediacy Index
- Articles (počet článků publikovaných v daném časopise v daném roce)
- Cited Half-life
- Citing Half-life
- Source Date
- Cited Journal
- Citing Journal
- Impact Factor Trend Graph ⁴⁹

7.2.2. Citační rejstříky

ISI spustila Current Contents v 50. letech 20. století. Na začátku 60. let spustila **Science Citation Index (SCI)**. SCI zahrnuje nejen úplnou bibliografickou informaci o každém článku v časopise (včetně jména autora a adresy, názvu článku, jménu časopisu a stránkování), ale také citace k dané práci nebo práce citující danou práci. Citation Index nám umožňuje nejen zjistit co kdo publikoval, ale také, jaké další práce daná práce cituje. Má tradiční funkce rejstříku. Rejstřík funguje na základě myšlenky, že jestliže znám práci, která se vztahuje k předmětu mého zájmu, pak další práce, které citují danou publikaci, mohou být pro mě také důležité. ⁵⁰

V ISI po dlouhém výzkumu zjistili, že většina každoročně publikovaných významných informací se nachází v relativně malém počtu časopisů. Mezi nimi je jakési jádro nejkvalitnějších. Směrem od jádra se kvalita snižuje.

Pro zpracování do SCI je vybíráno kolem 3 500 titulů časopisů a na 200 monografií z celého světa. Tento výběr se každoročně obměňuje podle impakt faktoru jednotlivých časopisů. ISI každý rok zkoumá a hodnotí asi 2 000 nových časopisů. Do báze bývá z tohoto počtu zařazeno jen 10-12%.

ISI přijímá a zpracovává pouze časopisy, které mají aspoň anglická abstrakta a klíčová slova, pravidelnou periodicitu a citovanost. Tato kritéria řada našich časopisů nespĺňuje. Kromě toho české časopisy vycházejí v relativně nízkých nákladech a do knihoven v cizině se dostávají v mizivém počtu. To se následně projevuje na citovanosti článků z časopisů naší produkce.

⁴⁹ KALCU, Petra, Journal Citation Reports.

⁵⁰ GARFIELD, Eugene, Citation analysis of sports medicine research, 1981-1996 productivity, impact and influence of nations, institutions and researchers.

7.2.3. Struktura SCI

SCI je tvořen jednotlivými citačními rejstříky, které vycházejí jak v tištěné podobě, tak na kompaktních discích. Jejich názvy jsou: *Citation Index*, *Source Index*, *Corporate Index*, *Permuterm Subject Index* a *Journal Citation Reports* - doplňkem je *Guide* (průvodce bází).

V rejstříku **Citation Index** je abecedně řazený seznam jmen citovaných autorů v daném roce, doplněný základními bibliografickými údaji jejich citovaných prací. Zároveň je pod každou citovanou prací uveden autor a zkráceně bibliografický popis citujícího pramene.

Source Index obsahuje abecedně řazený seznam jmen citujících autorů v daném roce s plnými bibliografickými údaji o citující práci a adresami pracovišť jejich autorů. Jsou zde uvedeni i spoluautoři citující práce, opět abecedně řazení. Dále je zde možné vyhledat autory a jejich práce, které byly v běžném roce publikovány v časopisech excerpovaných pro SCI.

Corporate Index zahrnuje abecední seznam názvů pracovišť autorů zařazených v Source Index.

Permuterm Subject Index je zpracován podle klíčových slov obsažených v titulech dokumentů. Klíčová slova, popř. jejich spojení jsou opět řazena abecedně.⁵¹

Dalšími rejstříky vytvářenými ISI jsou **Social Science Citation Index (SSCI)** a **Arts & Humanities Citation Index (AHCI)**. SSCI vznikl v roce 1974 a zaměřuje se na časopisy z oblasti sociálních věd (např. sociologie, psychologie nebo ekonomie). AHCI vznikl v roce 1980 a zaměřuje se na oblast humanitních věd. Tyto rejstříky mají stejné součásti jako SCI. Zde se jimi nebudu blíže zabývat, protože neobsahují časopisy z oblasti medicíny. Jsou stejně jako SCI přístupné v ČR v rámci webovského rozhraní Web of Science.⁵²

⁵¹ ROY, D. [et al.], Citation analysis of otorhinolaryngology journals.

⁵² PITTEROVÁ, Květa, Využití citačních rejstříků a impact factor.

8. Web of Science (Web of Knowledge)

Web of Science, (někdy zkracované jako WoS) vzniklo v roce 1997. Je webovským rozhraním do tří databází obsahujících bibliografické údaje z celého světa v oblasti výzkumu a vývoje. Skládá se z databází **Science Citation Index**, **Social Science Citation Index** a **Arts and Humanities Index**. V současné době je součástí většího celku **Web of Knowledge**, který bude popsán později.

V naší republice se v minulosti využívaly dva základní bibliografické zdroje - *Current Contents* a *Science Citation Index* (ať už v klasické papírové podobě, na disketách či CD ROM). Web of Science spojuje přednosti Science Citation Index (poskytuje údaje o citovanosti článků) se sledováním obsahu více jak 8 000 recenzovaných časopisů (týdně je aktualizováno více než 20 000 záznamů). Jde o databázi, tzn. že poskytuje veškeré možnosti prohledávání podle volby uživatele.⁵³

Trendem v poslední době jdou zcela nové způsoby citování, jejichž základem jsou jednoznačné identifikátory publikací, zejména pak elektronických. Ty vytvářejí za autora citaci použitého pramene, v případě potřeby přímo s hypertextovým propojením.

Z bibliografických a citačních databází je možné pohodlné získávání plných textů přímo z databází, depozitářů digitálních knihoven či archivů elektronických tisků nebo jednoduše z prostoru volného webu. Je možné získat dokument v elektronické podobě. Požadavkem uživatelů je účelné propojení informací v podobě hypertextových odkazů, které by umožnilo jejich co nejrychlejší a nejefektivnější využití. Požadujeme také možnost přímo získat text dokumentu (a to nejlépe v digitální formě a pomocí hyperodkazu) na základě vyhledané bibliografické informace, získané v bibliografické nebo citační databázi.

Angažovanou institucí v propojování informačních zdrojů v elektronickém prostředí na komerční bázi je také ISI. ISI se snaží o vybudování integrovaného informačního celku, kterým míří také k aplikaci citačních vazeb, které jak známo, jsou v oblasti vědy a výzkumu vazbami nejcennějšími.

Strategie samotného ISI směřuje také k budování automatických hypertextových vazeb na vybrané a žádané externí zdroje plných textů. Budování propojovaných systémů, v nichž bude integrována báze SCI jako celek nebo nějaká její část, představuje jeden ze základních směrů v rozvoji komplexních informačních systémů, v nichž jsou aplikovány citační vazby.

8.1. Web of Knowledge

V České republice bylo v červenci a srpnu 2003 zkušebně přístupné rozhraní **Web of Knowledge**, jehož je Web of Science součástí. Takto byl umožněn přístup do několika dalších databází např. INSPEC, Current Contents, či do rejstříků SCI, SSCI a AHCI (ve vyšší verzi vyhledávacího programu) a také přístup do Journal Citation Reports a Essential Science Indicators (ESI). A v současné době, knihovny a jiné instituce, které mají přístup do databází ISI vstupují do nich díky rozhraní Web of Knowledge.

⁵³ Web of Science. ISI. Timeline.

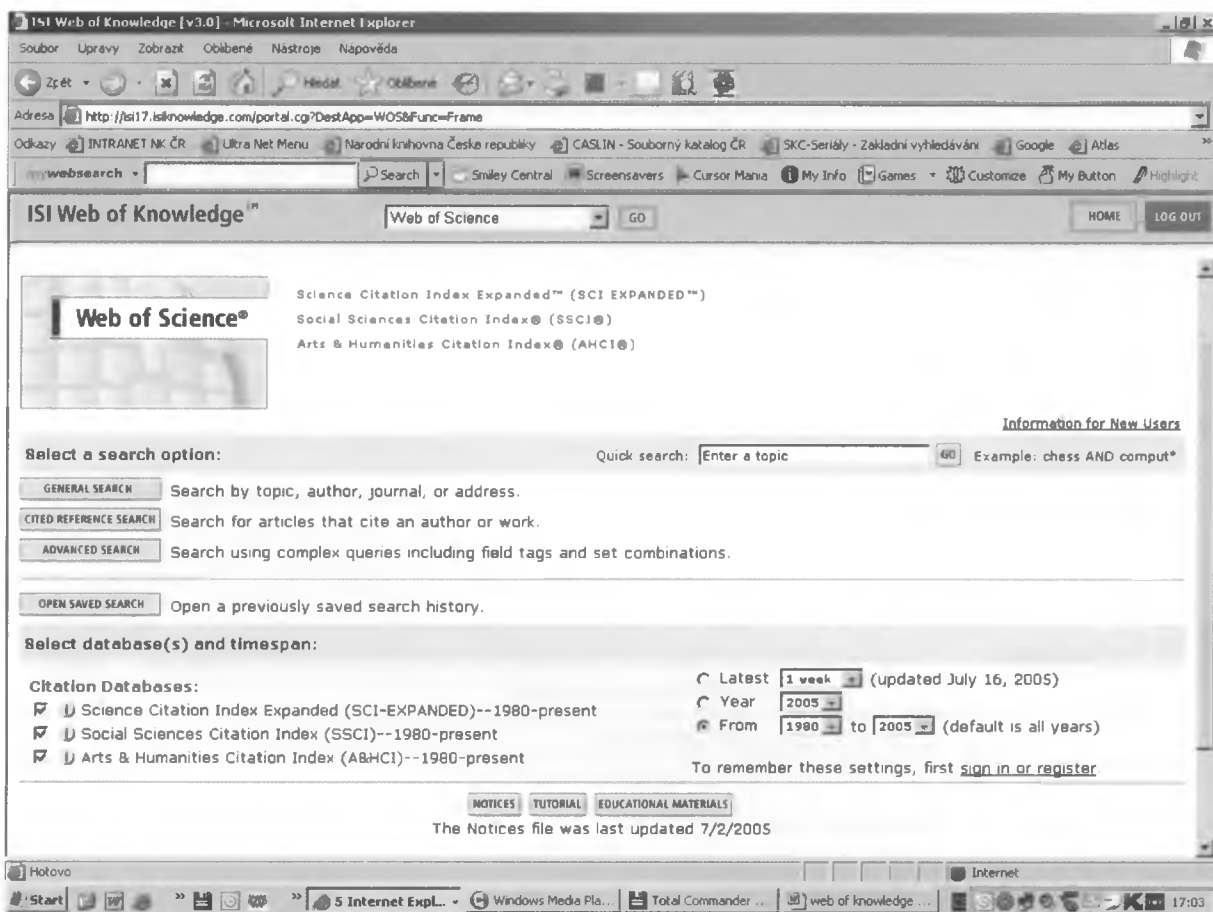


Hlavní stránka ISI Essential Science Indicators ⁵⁴

Webovská stránka **Web of Knowledge** a konkrétně **In cites** (která je další zpřístupňovanou částí Web of Knowledge) poskytuje komentáře a analýzu některých nejvlivnějších vědeckých objevů v naší době slovy vědců, kteří vytvářejí tyto objevy. Stránka je vytvořena pro doplnění publikací, citací a řazení citací na práci a dalších dat charakterizované v ISI Essential Science Indicators. ISI Essential Science Indicators je kompilace dat za 10 let výstupů a statistik zaměřující se na výzkumníky, instituce, země a časopisy. Tak zahrnuje seznamy vysoce citovaných a tak zvaných „hot“ prací pro 22 disciplín, stejně tak jako základní hranici citačních statistik měření, zda práce je citována nad nebo pod průměrem.

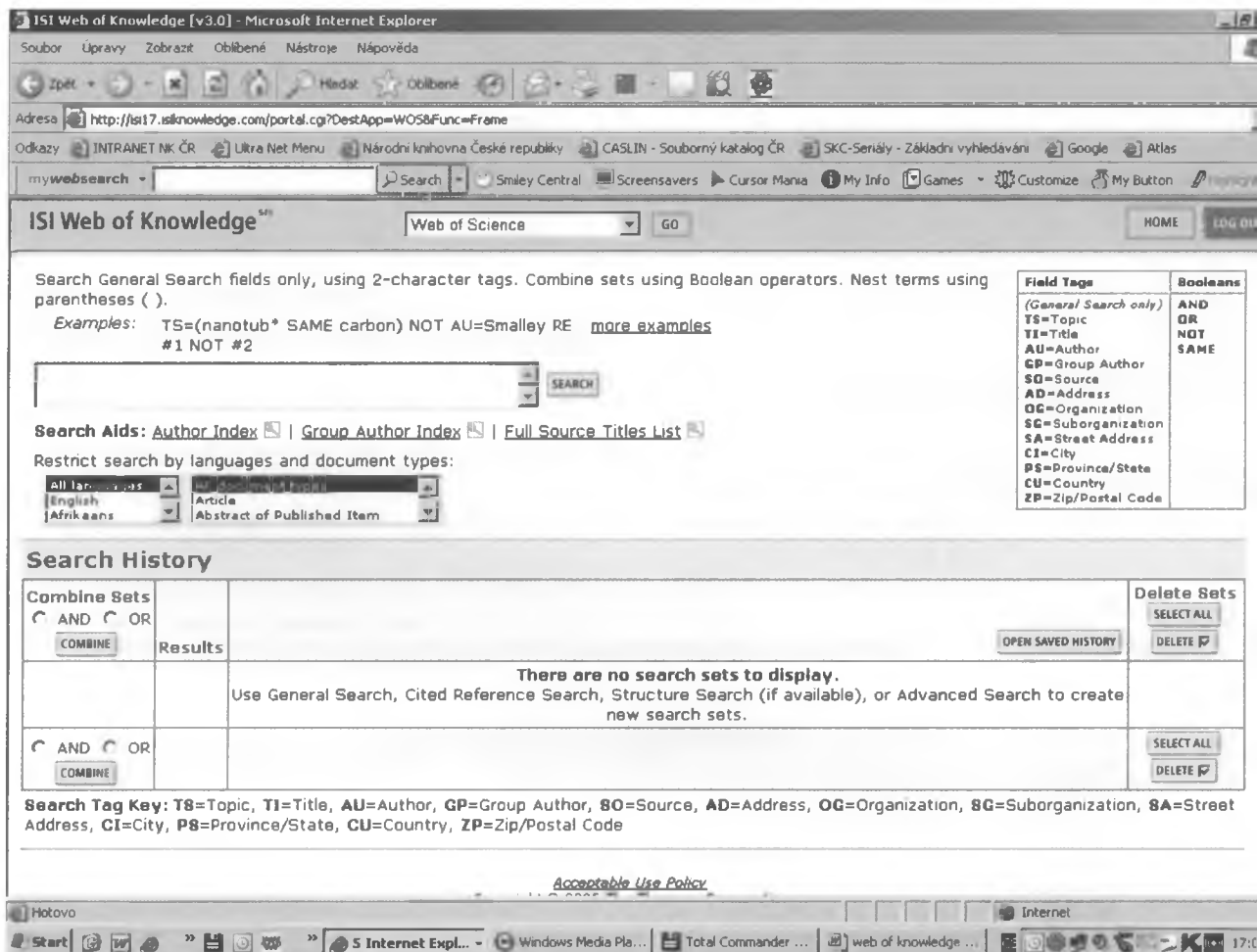
In cites identifikuje "top" výzkumné příspěvky a příspěvovatele, užívající citační data. Komentáře tvoří lidé jako byl Robert K. Merton, bývalý profesor Columbia University (zemřel na začátku roku 2003) a další sociologové vědy, chválící tyto komentáře ve stylu Citation Classic.

⁵⁴ www.isinet.com



Podoba Web of Knowledge ISI dle stavu 21.7.2005 ⁵⁵

⁵⁵ Získané z webovských stránek Web of Knowledge [http:// isiknowledge.com](http://isiknowledge.com)



Pokročilé vyhledávání WOK⁵⁶

⁵⁶ Získané z webovských stránek Web of Knowledge

ISI Web of Knowledge [v3.0] - Microsoft Internet Explorer

Soubor Upravy Zobrazit Obíbené Nástroje Nápověda

Adresa http://isi7.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame

ISI Web of KnowledgeSM Web of Science GO HOME

Web of Science®

WELCOME HELP GENERAL SEARCH CITED REF SEARCH SEARCH HISTORY ADVANCED SEARCH

Search Results -- Summary

TS=webometrics
DocType=All document types; Language=All languages; Databases=SCI-EXPANDED, SSCI, A&HCI; Timespan=1980-2005

Search within results: Enter a topic SEARCH

31 results found (Set #2) Go to Page: 1 of 4 GO

Records 1 -- 10 Show 10 per page

Use the checkboxes to select individual records for marking, then click Submit Marks to add them to the Marked List.

1. Li XM, Thelwall M, Wilkinson D, et al.
National and international university departmental Web site interlinking. Part 2: Link patterns
SCIENTOMETRICS 64 (2): 187-208 AUG 2005
Times Cited: 1
2. Chu H
Taxonomy of inlinked Web entities: What does it imply for webometric research?
LIBRARY & INFORMATION SCIENCE RESEARCH 27 (1): 8-27 2005
Times Cited: 0
3. Thelwall M, Vaughan L, Bjerneborn L
Webometrics
ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY 39: 81-135 2005
Times Cited: 0
4. Davenport E, Snyder HW

Sort by: Latest date SORT

Mark: [0 articles marked]

Selected records

All records on this page

Records to

SUBMIT MARKS

You can print, save, export, e-mail, and order records after adding them to the Marked List. (The list can hold 500 records.)

Analyze Results: ANALYZE

Hotovo

Start 5 Internet... Windows Me... Total Comma... 2 Microsoft... Petr - Messa...

Výsledky vyhledávání⁵⁷

⁵⁷ Získané z webovských stránek Web of Knowledge

ISI Web of Knowledge [v3.0] - Microsoft Internet Explorer

Soubor Upravy Zobrazit Obilbené Nástroje Nápověda

Adresa <http://isi17.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame>

Odkazy INTRANET NK ČR Ultra Net Menu Národní knihovna České republiky CASLIN - Souborný katalog ČR SKC-Serialy - Základní vyhledávání Google Atlas

my.websearch Search Smiley Central Screensavers Cursor Mana My Info Games Customize My Button

ISI Web of KnowledgeSM Web of Science GO HOME

Full Record

Record 1 of 31 (Set #2) SUMMARY

Title: National and international university departmental Web site interlinking. Part 2: Link patterns
Author(s): Li XM, Thelwall M, Wilkinson D, Musgrove P
Source: SCIENTOMETRICS 64 (2): 187-208 AUG 2005
Document Type: Article
Language: English

Cited References: 21 **Times Cited:** 1 FIND RELATED RECORDS

Abstract: Although many link patterns have been identified at the university level, departmental interlinking has been relatively ignored. Universities are multidisciplinary by nature and various disciplines may employ the Web differently, thus patterns identified at the university level may hide subject differences. Departments are typically subject-oriented, and departmental interlinking may therefore illustrate interesting disciplinary linking patterns, perhaps relating to informal scholarly communication. The aim of this paper is to identify whether and how link patterns differ along country and disciplinary lines between similar disciplines and similar countries. Physics, Chemistry and Biology departments in Australia, Canada and the UK have been chosen. In order to get a holistic picture of departments Web use profiles and link patterns, five different perspectives are identified and compared for each set of departments. Differences in link patterns are identified along both national and disciplinary lines, and are found to reflect offline phenomena. Along national lines, a likely explanation for the difference is that countries with better research performances make more general use of the Web; and, with respect to international peer interlinking, countries that share more scholarly communication tend to interlink more with each other. Along disciplinary lines, it seems that departments from disciplines which are more willing to distribute their research outputs tend to make more general use of the Web, and also interlink more with their national and international peers.

KeyWords Plus: CO-AUTHORSHIP; IMPACT FACTOR; WEBOMETRICS

Addresses: Li XM (reprint author), Wolverhampton Univ, Sch Comp & Informat Technol, 35-49 Lichfield St, Wolverhampton, WV1 1EL England

MARK (0 articles marked)
 (Save, Export, E-mail, Order, Print)

Create Citation Alert
 CREATE CITATION ALERT
 Receive e-mail alerts on future citations to this record. (Requires registration.)

View record in Journal Citation Reports

Hotovo Start S Internet... Windows Me... Total Comma... 2 Microsoft... Petr - Messa...

Konkrétní záznam s počty citovaných a citujících dokumentů ⁵⁸

⁵⁸ Získané z webovských stránek Web of Knowledge

ISI Web of Knowledge [v3.0] - Microsoft Internet Explorer

Soubor Upravy Zobrazit Oblíbené Nástroje nápověda

Adresa <http://isi17.isiknowledge.com/portal.cgi?DestApp=WOS8Func=Frame>

Odkazy INTRANET NK ČR Ultra Net Menu Národní knihovna České republiky CASLIN - Souborný katalog ČR SKC-Seriály - Základní vyhledávání Google Atlas

websearch Search Smiley Central Screensavers Cursor Mana My Info Games Customize My Button

ISI Web of Knowledge™ Web of Science GO HOME

Citing Articles--Summary

National and international university departmental Web site interlinking. Part 2: Link patterns
 LI XM, THELWALL M, WILKINSON D, et al.
SCIENTOMETRICS
 64 (2): 187-208 AUG 2005

These documents in the database cite the above record:

1 result found Go to Page: 1 of 1 GO
 Records 1 -- 1 Show 10 per page

Use the checkboxes to select individual records for marking, then click Submit Marks to add them to the Marked List.

1. Li XM, Thelwall M, Wilkinson D, et al.
National and international university departmental Web site interlinking. Part 1: Validation of departmental link analysis
 SCIENTOMETRICS 64 (2): 151-185 AUG 2005
 Times Cited: 1

Mark: [0 records marked]
 Selected records
 All records on this page
 Records [] to []
 SUBMIT MARKS
 You can print, save, export, e-mail, and order records after adding them to the Marked List. (The list can hold 500 records.)

Analyze Results:
 ANALYZE
 View rankings and histograms of the authors, journals, etc. for this set of records. (Up to 2,000 records at a time.)

SUBMIT MARKS

Hotovo Internet

Start 5 Internet... Windows Me... Total Comma... Z Microsoft ... Petr - Messa...

Záznam dokumentu, který předešlý dokument cituje (tzn. Citující dokument) ⁵⁹

⁵⁹ Získané z webovských stránek Web of Knowledge

ISI Web of Knowledge [v3.0] Microsoft Internet Explorer

Soubor Úpravy Zobrazit Obilbené Nástroje Nápověda

Adresa <http://isi17.isiknowledge.com/portal.cgi?DestApp=WOS8Func=Frame>

mywebsearch Search Smiley Central Screensavers Cursor Mania My Info Games Customize My Button

ISI Web of KnowledgeSM Web of Science GO HOME

MANAGEMENT SCIENCE 51 (5): 756-770 MAY 2005
Times Cited: 0

2. Fung MK
Are knowledge spillovers driving the convergence of productivity among firms?
ECONOMICA 72 (286): 287-305 MAY 2005
Times Cited: 0

3. Dahlin K, Taylor M, Fichman M
Today's Edisons or weekend hobbyists: technical merit and success of inventions by independent inventors
RESEARCH POLICY 33 (8): 1167-1183 OCT 2004
Times Cited: 0

4. Almeida P, Phene A
Subsidiaries and knowledge creation: The influence of the MNC and host country on innovation
STRATEGIC MANAGEMENT JOURNAL 25 (8-9): 847-864 AUG-SEP 2004
Times Cited: 0
[VIEW FULL TEXT](#)

5. Chen CM, Hicks D
Tracing knowledge diffusion
SCIENTOMETRICS 59 (2): 199-211 2004
Times Cited: 1

6. Acosta M, Coronado D
Science-technology flows in Spanish regions - An analysis of scientific citations in patents
RESEARCH POLICY 32 (10): 1783-1803 DEC 2003
Times Cited: 2

7. Huang MH, Chiang LY, Chen DZ
Constructing a patent citation map using bibliographic coupling: A study of Taiwan's high-tech companies
SCIENTOMETRICS 58 (3): 489-506 2003

Mark: [0 articles marked]
 Selected records
 All records on this page
 Records [] to []
[SUBMIT MARKS](#)
You can print, save, export, e-mail, and order records after adding them to the Marked List. (The list can hold 500 records.)

Analyze Results:
[ANALYZE](#)
View rankings and histograms of the authors, journals, etc. for this set of records. (Up to 2,000 records at a time.)

Start Internet Explorer Windows Media Player Total Commander Microsoft Office

Vyhledané dokumenty dle dotazu s možností získat plný text u dokumentu, ke kterému má daná instituce přístup⁶⁰

⁶⁰ Získané z webovských stránek Web of Knowledge

9. Kritika citační analýzy a autocitace

Citační analýza je kritizována pro množství důvodů. Někdy se uvádí, že studie citační analýzy nerozlišují mezi pozitivním a negativním hodnocením citovaných prací.

Studie spoléhající na citační indexy Institute for Scientific Information mají některé zvláštní problémy. Tyto citační indexy indexují (zvláště v některých oblastech) jen malé množství časopisů. Citace v neindexovaných časopisech v těchto studiích chybějí. U publikací s více autory se nelze spolehnout, že všichni autoři (zvláště ti, kteří nefigurují na prvním místě), budou uvedeni a citovanost bude k jejich jménu připsána.

Určitý problém tvoří autocitace. Některé analýzy vylučují tyto citace, další je běžně započítávají. Ani jedna cesta není plně uspokojivá. Vyloučení autocitací riskuje nedocenění vlivu profilu autorů odpovědných za daný počet publikací zkoumaných ve studii. Zahrnutí autocitací je zase nebezpečné z důvodu přeceňování vlivu vědců, kteří jsou zamilovaní do citování svých vlastních prací.⁶¹

Distribuce citací je nespravedlivá k mladým vědcům při hledání nezbytných součástí jejich práce skrze zveličování citací a dalších forem lichocení.

Citace mohou být zbytečné, jestliže obsahují nedostatečná data. Neadekvátní citační data mohou být sledována jako nepřiměřené nebo rušící „lepidlo“ intelektuálních diskurzů.

Problémem je někdy také nedostatečná znalost dané oblasti, kterou autor cituje.⁶²

Autocitace jsou již dlouho ve středu zájmu v infometrii a odhalují publikační chování jednotlivců a jejich vztah v rámci akademických sítí. Výzkum Kena Hylanda vytváří zajímavý pohled. Obvykle se má za to, že autocitace je neutrální forma neinformující na rozdíl od citací o dalších pracech nebo se považuje za určitý druh akademického egoismu. Zkoumání autocitací v širším kontextu jako zmínění sama sebe může znamenat část obsáhlejší řečnické strategie pro zdůraznění osobního příspěvku pisatele k části výzkumu a potvrzení jeho nebo její znalostí, výzkumné důvěryhodnosti a v širších souvislostech příslušnosti k určité disciplíně. Tyto významy nejsou snadno odhalitelné kvantitativními bibliometrickými metodami a požadují pečlivou textovou analýzu a rozhovory s vědci.

Potřeba jedince mít potvrzení jeho práce a být citován dalšími je ve skutečnosti stále více hodnocená komodita v dnešním vášnivě soutěživém akademickém světě. Rivalové se snaží být lepší co do kvality a významnosti jejich práce než jejich profesionální kolegové. Získávají tak nejen ceny, ale také granty, funkce, majetek a podporu.

Užití autocitace je způsob, jak tvůrci mohou podpořit svou profesionální důvěryhodnost. Bibliometrické indikátory jsou používány ke zkoumání modelů autocitací od prvních prací Garfielda, ale kromě toho také poskytují informaci o modelech čtení. Říkají však málo o tom, jak tvůrce hledá podporu jeho reputace v jeho psaní. Existuje velmi málo článků, které neobsahují žádné autocitace a studie ukazují, že autoři mají tendenci citovat jejich vlastní práci častěji než další práce. To je částečně proto, že pokud autoři mají dlouhou historii aktivit v dané oblasti, snaží se i publikovat mimo jejich oblast specializace nebo jsou sami často citováni. Zatímco někteří nahlížejí na nadměrné

⁶¹ WRIGHT, Richard; MILLER, J. Mitchell, The most-cited scholars and works in corrections. MCMILLAN, G. Steven; HAMILTON, Robert D., Using bibliometrics to measure firm knowledge : an analysis of the pharmaceutical industry.

⁶² KOEHLER, Wallace; COLIMNIST, Guest, Nightmares in citation analysis.

autocitace jako na pochybnou formu podpory sama sebe, mnozí z výzkumu Hylanda nevidí žádný rozdíl mezi motivací pro autocitace a citace dalších autorů, jednoduše jde o návaznost současné práce na dřívější relevantní práci.

Článek má rozdílné postoje a argumenty v tom, že autocitace je užitečnější v širším kontextu zmínění sama sebe u autorů.

Zatímco sociolingvistické přístupy nehrají hlavní roli v informační vědě, detailní studium lingvistického výběru autorů může odhalit důležité aspekty znalostí a modelů, které charakterizují oborové vztahy. Způsoby a míra, jak vědci odkazují na sama sebe a jejich práci nabízí významný pohled na jejich roli ve výzkumném procesu a epistemologické a sociální priority jejich disciplín. Komunikace je hlavní mechanismus pro vytváření znalosti a vytvoření reputace. Práce je posuzována jako příspěvek ke konkrétnímu oboru publikem kolegů, kteří ji mohou potenciálně využít. Jestliže editoři rozhodují a čtenáři časopisu na to pohlížejí jako na originální a významné, dovolují, aby to bylo publikováno, citováno v jejich vlastní práci a rozvíjeno dále, poté autor získává odměnu za uznání a úspěch v kariéře.

Úspěšné publikace mohou pomoci výzkumníkovi získat body, které mohou být přeměněny do výzkumného grantu a tím získu peněz na vybavení své a svých kolegů. Vzniká více dat, která mohou být převedena do argumentů, nových publikací atd. Důvěryhodnost takto pomáhá vědcům postoupit v kruhu:

Např. úspěšná investice může znamenat, že lidé mu telefonují, jeho abstrakty jsou akceptovány, další ukazují zájem o jeho práci, on je vnímán lépe a uváděn v různých přehledech s větší pozorností, je mu nabízena lepší pozice, jeho práce je více využívána, tok dat je důvěryhodnější a forma obrazu o něm se jeví okolí jako věrohodnější.

Akademický úspěch je viděn do velké míry jako posuzování a neustálý proces požadující uznání závislé na kapacitě vytvořených prací, které jsou hodnoceny kolegy.

Pozice a institucionální postavení jedince v oboru jsou jistě významnou motivační silou, ale reputace se těžce získává. V mnoha zemích a institucích je kladen důraz na publikace, které s sebou přinášejí určité zviditelnění. V akademické komunikaci se začíná prosazovat rychlého publikování „vědeckých dopisů“ v časopisech a je velký zájem o novinky.

Autocitace je výraz pro rostoucí soutěživost, ovlivňující viditelnost autorů. Profilují se v citačních indexech a může být na tom závislá i jejich kariéra. White například zjistil, že autocitace přesahují citace k dalším a reprezentují „jádro jádra“ citační identity každé z osmi vedoucích forem v informační vědě. Cronin a Shaw rozšířili studii Whiteho a ukázali také, že autocitování je důležitým komponentem vědeckých akademických identit a vytvářejí formou znovucitování spojení mezi předmětem a jejich případovými studii. Phelan zjistil, že jen dva z 56 nejvíce citovaných autorů v oblasti vzdělávání necitují sami sebe za období 20 let, zatímco 154 z 280 citací (55%) získaných jedním autorem byly výsledkem autocitací. Peritz a Bar-Ilan poukázali na to, že pět prací publikovaných v časopise *Scientometrics* v letech 1990 a 2000 obsahovalo přes 50% autocitací a poznamenali, že okolo 30 % prací za oba roky má poměr 20 % a více autocitací.

Studie shodně ukazují, že obvyklý počet autocitací je mezi 10% a 20% všech citací, v závislosti na oboru a stupni vývoje oboru. Mnoho pozorovatelů potvrzuje, že autocitace je významná část citační sítě a má jasné charakteristiky, tendenci být novější a projevit větší frekvenci opakování, představuje kontinuitu práce citované a jejího významu.

Jiný pohled představují ti, kteří doporučují, aby čtenářům byla poskytnuta neosobní informace, která tak vypadá „důvěryhodnější“ a objektivnější“.

Z tohoto důvodu mnoho manuálů a učebnic doporučuje vyvarování se osobního zájmena ve prospěch anonymnější osoby.

Tyto modely neosobnosti jsou ale dané způsoby propagační praxe, marketingem sama sebe a práce jedince analogicky k propagaci zboží. Zatímco požadavky mají potvrdit přiměřenou podporu a odkaz k existujícím znalostem, autoři musejí také zajistit, že čtenáři pochopí jejich *individuální příspěvek* a jejich prosazování *akademické priority*, jenž jsou základem vědecké reputace. Zvýšení jejich viditelnosti vůči čtenářům díky použití určitého výrazového aparátu, který zdůrazní jejich podíl na textu a odkaz na jejich dřívější práci, jsou důležitými způsoby, kterými toho mohou dosáhnout. Mnoho autorit nyní vidí tuto formu zmínění sama sebe jako pozitivní.

Strategické užití osobní citace a autocitace dovoluje autorům zdůraznit a najít potvrzení jejich vlastního příspěvku k oboru, vysílají jasný signál, že oni stojí za svým tvrzením. A tak může být neosobnost institucionálně posvěcená a autocitace chápána jako špatná, ale přes to jsou tyto konvence konstantně překračovány.

Infometrický výzkum poznamenává, že pouze několik článků neobsahuje žádné autocitace a že autoři citují jejich vlastní práci více než další.

Disciplína	Výzkumné články			Abstrakty článků		
	celkově	citace	Zmínění sama sebe	Celkově	citace	Zmínění sama sebe
Přírodní a technické vědy	17.6 (31.6)	5.6 (59.1)	12.1 (26.1)	0.3 (25.9)	0.0 (0.0)	0.3 (26.0)
Sociální a humanitní vědy	38.1 (68.4)	3.9 (40.9)	34.2 (73.9)	0.8 (74.1)	0.1 (100)	0.8 (74.0)
Celkově	27.9 (100)	4.7 (100)	4.7 (100)	0.6 (100)	0.0 (100)	0.6 (100)

Tabulka: frekvence autocitací na práci v procentech dle oboru⁶³

Faktory, které motivují autory citovat jejich vlastní práci jsou nepochybně různé a komplexní. Představují psychologické faktory ovlivněné individuální důvěrou autorů, zkušenostmi a vědomím vlastní ceny. Bonzi a Snyder informují hlavně o podobných důvodech pro vlastní a jiné citace, Tagliacozzo zdůrazňuje potřebu autorů propojit jejich současnou práci s předešlou. Žádný výzkum se nenachází v sociálním vakuu. Citace k dřívější literatuře indikuje upevnění v problému a je významným článkem v konstrukci nových znalostí v oboru.

Závislost argumentů v sítích citací neznamena jen kumulativní a lineární vývoj znalostí, ale skutečnost lokalizovat autory a jejich požadavky v pochopeném oborovém rámci. Odkaz autora na jeho vlastní dřívější práce zdůrazňuje jeho angažovanost ve společném profesionálním úsilí. Autocitace je důležitý prostředek demonstrující jedincovo oborové oprávnění a stálost. Autor se tak sám stává více viditelným a je

⁶³ HYLAND, Ken. Self-citation and self-reference.

v pozici důležitého hráče v oblasti, jehož práce by měla být brána vážně. Bonzi a Snyder zmínili potřebu upevnit jejich autoritu v oboru jako motivaci pro autocitace.

Existují podstatné rozdíly v používání autocitací v různých disciplínách. Odkaz

na předešlý výzkum hraje významnější roli v humanitních a sociálních vědách. Autoři více odkazují na literaturu, která má častěji větší historický a tematický rozptyl. To znamená, že citace jsou relativně rozptýlené a příležitosti k autocitování jsou porovnatelně menší než v přírodních vědách.

Kontrastem jsou odkazy v přírodních vědách a technických oborech, které mají tendenci být těsněji svázány s daným výzkumným tématem. Vědecké požadavky, pokud jsou akceptovány, jsou obecně považovány za objevy, jejich rozšíření je prověřené fakty. To, co je nové, znamená další kostku ve výzkumné skládance. Citace v těchto oborech úžeji definují specifický kontext a příspěvek ve smyslu lineárního vývoje. Důležitým rysem takovéto výzkumné praxe je, že vědci mají tendenci účastnit se velmi specializovaných oblastí výzkumu. Jsou donuceni provádět expertizu díky potřebě udržet si přehled v rychle se vyvíjejících oborech a vkládat hojně investice do procedurální schopnosti a technického vybavení. Výzkum konkrétního problému je často prováděn v omezeném počtu sítí a omezeném počtu výzkumníků, a to dovoluje autorům upozornit na jejich vlastní práci v daleko větším měřítku než v sociálních a humanitních oborech.

10. Modely autocitací v publikovaném výzkumu

Prominentní forma zmínění sama sebe, než odkazování na předešlý publikovaný výzkum daného jedince, je použití osobního zájmena. Existuje podstatný kontrast v tom, jak se chovají rozdílné obory.

Existují různé způsoby, jak vědci provádějí v rámci disciplíny výzkum a přesvědčují čtenáře, aby akceptoval jejich výsledky. Autoři v přírodních vědách hledají potvrzení v empirických výzkumných aktivitách, které představují přesné měření a systematické zkoumání omezeného počtu kontrolovaných proměnných. Znalosti těchto oborů mají tendenci být univerzální a jsou motivovány konceptuálními problémy. Výzkum se proto obvykle skládá z doprovodných experimentů a snahou je nabídnout řešení specifických problémů oboru.

Ve výsledku výzkumu může autor zdůraznit svou roli, aby zdůraznil fenomén studie, opakovatelnost výzkumných aktivit a univerzálnost zjištění. Nebo může zvolit méně osobní styl, aby posílil objektivitu interpretace. Takovou strategii může chytře sdělit základní ideologii, která navrhuje výzkumné výsledky .

Jinak je tomu v humanitních a sociálních vědách. Podpora efektivního stupně osobního podílu je extrémně hodnotná strategie. Navozuje spojení mezi entitami, které jsou obecně podrobnější, méně precizněji měřitelné a méně jasně oddělitelné než v přírodních a technických vědách. Proměnné jsou častěji více heterogenní a kauzální spojení je v sociálních vědách slabší. Úspěch autorů, jejichž snahou je přijetí jejich argumentů, závisí na větším rozšíření informací a na navázání na věrohodnou a angažovanou osobu. Rozhodnutí uplatnit první osobu zájmena v těchto oborech je spojeno se snahou prezentovat sama sebe jako informovaného důvěryhodného kolegu, silně identifikující sama sebe, dosáhnout povrzení pro individuální záměry a výzkumná rozhodnutí.

Zmínění sama sebe je důležité, protože hraje rozhodující roli ve zprostředkování vztahu mezi argumenty autorů v jejich disciplínách. Distribuce těchto charakteristik ukazuje, že zatímco existují tlaky k autocitování, ne všechny disciplíny používají stejná pravidla a existují různé sankce při rozdílných stupních prezentace autorů. Kromě vytváření přesvědčivého obrazu o tom, jak autor přispěl do mozaiky disciplíny, autor zdůrazněním svého přínosu demonstruje obeznamenost s řečnickými konvencemi a epistemologickými prvky dané disciplíny. Faktory, které ovlivňují způsob zdůraznění sama sebe jsou kromě jiného i pokročilejší věk, zkušenost a důvěra v sebe sama.⁶⁴

10.1. Výzkumná spolupráce: modely spoluautorství

Způsoby práce výzkumníku jsou založeny na mnoha způsobech a odrážejí spolupráci na rozdílných úrovních a rozdílných typech. Vzájemné působení modelů výzkumníků zrcadlí tok myšlenek a odráží přitažlivost modelů a tradic spolupráce. To je ovlivňováno podobnostmi a rozdílnostmi mnoha typů a také to odráží různé druhy bariér projevujících se při kontaktech mezi vědci. Model spolupráce obsahuje základní struktury jako je vzdálenost, jazyk, národnost, náboženství, kultura, ekonomické bloky, migraci (emigranti) a dokonce přátelství.

⁶⁴ HYLAND, Ken. Self-citation and self-reference

Tradiční spolupráce je velmi důležitá a tradice v ní se odlišuje od disciplíny k disciplíně, od univerzity k univerzitě, mezi univerzitním světem a dalšími producenty výzkumných výstupů, jako jsou soukromé společnosti, veřejné instituce a organizace.

10.1.1. Vůdci a následovníci: citační modely

Výzkumníci citují další výzkumníky pokud používají výsledky, diskusní práce, přemýšlejí o posuzování metod a o výsledcích zkoumání. Mnohé motivace je možné nalézt za způsoby, jak lidé citují. Model citování odráží úctu (vážnost) citované práce, ačkoliv existují další motivy pro citování.

Racionální modely citování jsou velmi komplikované. Analýza modelů citování umožňuje ukázat významný obraz vůdců a následovníků. Analýza citačních vazeb míří a identifikuje vedoucí citace, což dává další charakteristiku výzkumných modelů.⁶⁵

Citační analýzy nemusí pracovat dobře u rychle se rozvíjejících oblastí. To platí především pro patenty. Patenty mohou získat velký počet citací, ale nemusí to znamenat, že prezentují nejlepší technologie.

Nicméně to neznámá, že citační analýza nemůže být použita v těchto oblastech. Společnosti s velkou kapacitou výzkumu a vývoje pravděpodobně si zvyšují schopnost produkovat významné technologie v budoucnu.⁶⁶

Kvalita článků publikovaných v časopise a stupeň pravděpodobnosti jejich pozdějšího citování, může být také závislá na vydavateli časopisu. Hlavní těžkostí je, že jsou porovnávány časopisy, jejichž vydavatelé se ve stejných časových intervalech nemění.⁶⁷

Využití impakt faktoru se odráží v kvalitě citací. Problémem je interpretace počtu citací. Pracuje se jednoduše s počítáním citací, ale nikoliv s kontextem citace.

Např. De Jong a Schaper se zabývali počtem vědeckých prací, které nejsou nikdy citovány. Analyzovali 137 019 prací z oblasti klinické kardiologické medicíny v rámci zemí G7 a dalších sedmi menších zemí Evropy. 46% prací nikdy nebylo citováno. Nejlépe dopadlo Norsko, kde nebylo citováno 31% prací, nejhůře dopadlo Japonsko, kde nikdy nebylo citováno 69% prací.

Ahlgren, Jarneving a Rousseau rozebírali Pearsonův koeficient korelace (Pearson's Correlation Coefficient) jako podobnost měření v autorské kocitační analýze (ACA) s argumentem, že toto měření je citlivé na nuly. Přidání nul k dvěma proměnným by rozšířilo jejich podobnost, ale autoři ukazují pomocí empirických příkladů, že toto přidání může snížit koeficient korelace mezi proměnnými. Saltonův kosinus je navrhován jako možná alternativa, protože tato podobnost měření je necitlivá k přidání nul.

Doposud se Pearsonova korelace užívala v autorské kocitační analýze s pragmatickým argumentem, že rozdíly mezi užitím rozdílné podobnosti měření může být ve výzkumné

⁶⁵ MATTHIESSEN, Christian Wichmann; SCHWARZ, Annette Winkel ; FIND, Søren, The Top-level global research system, 1997-99 : centres, networks and nodality. An analysis based on bibliometric indicators.

MCMILLAN, G. Steven; HAMILTON, Robert D. Using bibliometrics to measure firm knowledge : an analysis of the pharmaceutical industry.

⁶⁶ BREITZMAN, Anthony ; THOMAS, Patric. Using patent citation analysis to target/value M & A candidates.

⁶⁷ ROMANO, Claudio; RATNATUNGE, Janek, Citation analysis of the impact of journals on contemporary small enterprise research.

praxi opomenuto. Pearsonovo r je vloženo ve více variantní statistice a protože normalizace zahrnuje v sobě toto měření, bere v úvahu negativní hodnoty. Problém s nulami může být objasněn aplikováním logaritmické transformace k datům. Dle jeho názoru je tato transformace v každém případě vhodná v případě dvou variantních normálních rozložení.

Jednoduchá vazba shlukování je dobře známá při spojování oblastí velké hustoty dohromady s jedním super shlukem. Kocitační shluky ve *World Atlas of Science* byly někdy spojeny tak zvaným efektem „řetězení“. Rozdílnost mezi použitím koeficientu Pearsonovy korelace a Saltonovým kosinem jsou v praxi chápány jako okrajové, protože korelace měření může také být vzata v úvahu jako kosinus mezi normalizovanými vektory. Normalizace je citlivá na nuly, ale to může být opraveno díky logaritmické transformaci.

Problém, jak vypočítat počet shluků, faktorů, skupin, dimenzí atd. je proniknutí do více variantní analýzy. Jestliže neexistují žádné *priori* teoretické důvody – jako je případ užití těchto technik ve výzkumu – může takový postup trvat libovolně dlouho. Ve faktorové analýze jsou metody vizuální kontroly dlouhého diagramu nebo přerušení jisté vlastní hodnoty společná praxe. V shlukové analýze je vícehlediskového míra rozhodnutí založená na vizuální kontrole výsledků společná.

$$H(n/m) = H(n,m) - H(m)$$

Odtud

$$H_0 = H(n) - H(n/m),$$

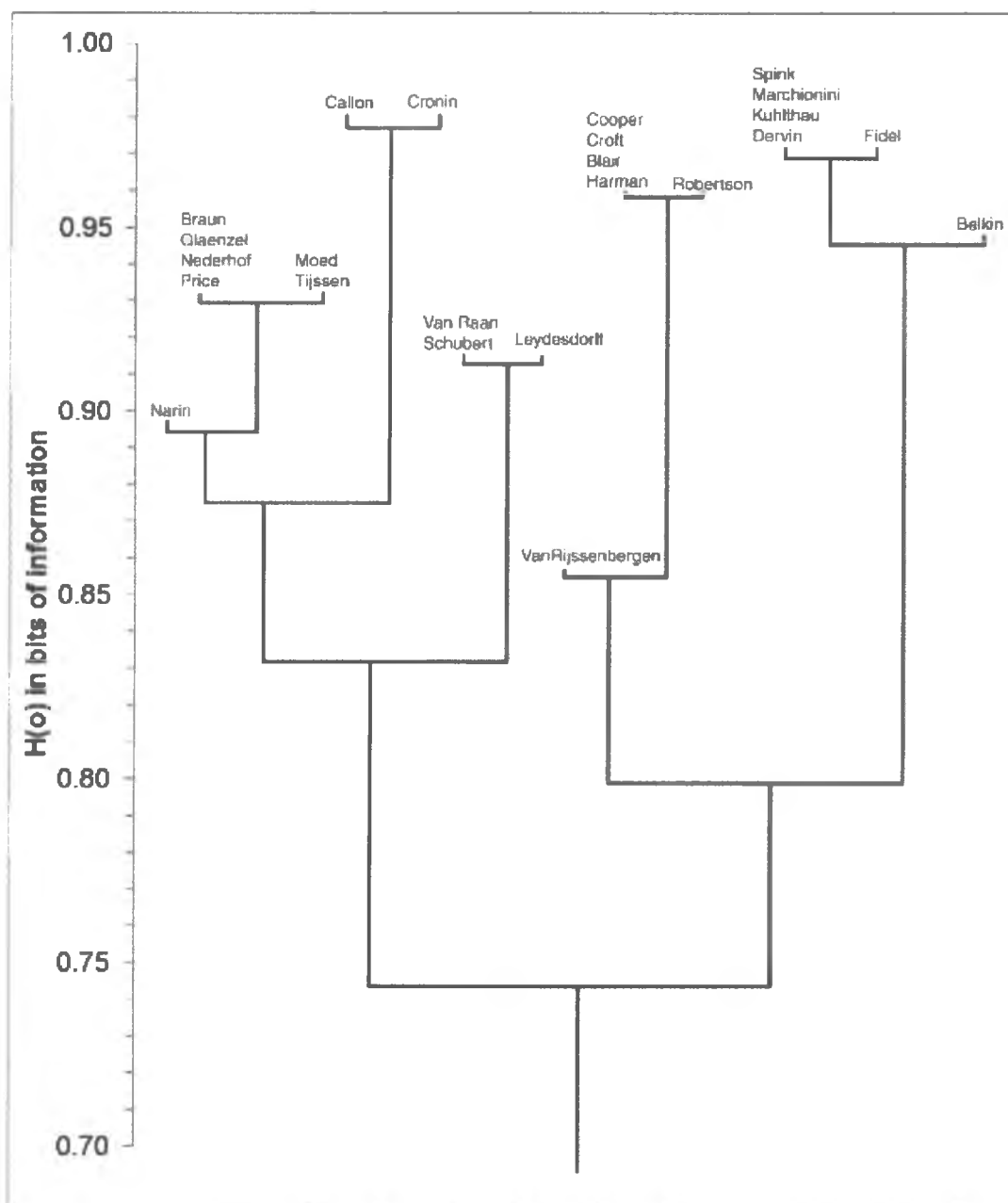
to implikuje:

$$H_0 = H(n) + H(m) - H(n,m)$$

Jinými slovy, zvýšení H_0 , pokud člověk rozlišuje přidanou skupinu (shluk, faktor atd.) je tvořeno částí, která je závislá jen na seskupení proměnné ($H(m)$) a částí, která je závislá na interakci mezi seskupení proměnné m a seskupení proměnné n . Interakce mezi dvěma proměnnými vytvoří $H(n,m)$ menší než součet $H(n)$ a $H(m)$. Dané číslo proměnných n má být seskupeno a pak se nabízí otázka: pro kterou hodnotu m bude funkce $\{H(m) - H(n,m)\}$ a tudíž H_0 jako indikátor rozdělení dosáhne maxima? Tento problém může být rozřešen numericky díky rekurzivnímu znovurozdělení případů do všech možných seskupení. (Normalizovaná) maximalizace H_0 takto poskytuje jednoznačné kritérium pro označení kvality.

Zkoumá se, zda uspořádání oddělené v jakémkoliv případě, vyústí do dvou podskupin. Byly zkoumány všechny možnosti a bylo rozhodnuto na základě nejlepšího dělení o zařazení do dvou podskupin. Analýza může být eventuelně opakována pro dvě podskupiny.

Po normalizaci H_0 v termínech hlavního součtu matice může být dendrogram vytvořen, což je obsaženo v termínech vertikálních vzdáleností mezi uzly a v termínech, jejichž další části vedou k podskupinám, které nejsou nižší v jejich entropii než jejich příslušné nahromadění. Tato úroveň v grafu odpovídá maximu pro H_0 .



Divizní shlukování matice autorské kocitace užívané v informační teorii

Informačně teoretický přístup nebere v potaz pouze nenulové hodnoty jako je kosinus, ale také zahrnuje informace o nadbytečnosti poskytované nulami v matici. Tato strukturální informace je požadována pro analýzu více proměnných, jak poznamenal Bensman. Ahlgren se spolupracovníky problém s nulami ovlivňující Pearsonovu korelaci řešil – ačkoliv člověk potřebuje toto měření pro parametrický přístup – nemusí také být zamítnut. Informační teorie nám umožňuje kombinovat neparametrický přístup s adresováním informací obsažených v nulách.

Hlavní výhodou použití informační teorie pro shlukování je uspořádání dvou parametrů mezer pro kritérium podobnosti a algoritmy shlukování. Člověk nepotřebuje další rozhodnutí o algoritmu shlukování (kromě rozhodnutí o kritériu podobnosti), protože informační teorie jich poskytuje dostatečný počet. Řešení je unikátní (poněvadž je exaktní) v užití jednoho kritéria, znamenitě se osvědčuje v maximalizaci v-mezi

skupinami nejistoty. Seskupování a rozdělování tak může být vyjádřeno přesně v termínech bitů informací obsažených v datech.⁶⁸

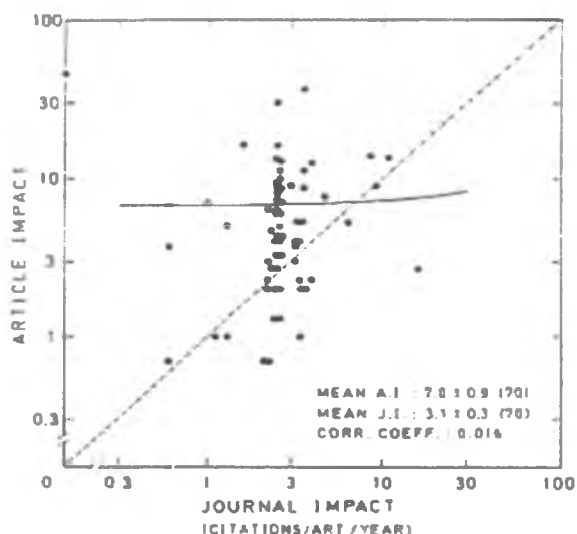
⁶⁸ LEYDESDORFF, Loet. Similarity measures, author cocitation analysis, and information theory.

11. Užití a zneužití impakt faktoru

Impakt faktor se používá jako nástroj pro hodnocení kvality nejen časopisů, ale také jednotlivých prací, vědců nebo skupin vědců. Později se právě toto stává hlavní příčinou, proč je citační analýza hodnocena jako méně atraktivní. Základním předpokladem je, že článek publikovaný v časopise adekvátně reprezentuje kvalitu daného časopisu jako celku. Pro individuální výstupy (např. hodnocení vědců nebo skupin vědců), může být hodnocení kvality velmi závažné. Není možné počítat jen průměr, ale je nutné pracovat se standardní odchylkou a standardní chybou významu (sémantickou). Pak je možné říci, že impakt faktor opravdu vyjadřuje hodnocení kvality časopisů.

11.1.1. Dovoluje impakt faktor hodnotit kvalitu jednotlivých autorů?

Analýza práce jednoho autora za dobu 17 let (viz obr. níže) znázorňuje impakt faktory časopisů, ve kterých práce tohoto autora byly publikovány. V průměru tento konkrétní autor publikoval práce v časopise s impakt faktorem 3,1, zatímco jeho články měly impakt faktor 7. To znamená, že autor byl více citován než další práce publikované v daných časopisech.



Porovnání impakt faktoru časopisů, ve kterých autor publikoval po dobu 17 let a impakt faktoru individuální práce.⁶⁹

Není možné používat impakt faktor pro hodnocení kvality práce jednotlivého autora. Neexistuje vztah mezi impakt faktorem časopisu a eventuálním hodnocením citací prací tohoto autora.

Impakt faktor je možné spolehlivě uplatnit pro ocenění kvality vědců jen v rámci velké skupiny (jako jsou univerzity).

Kvalita práce bude důležitější než impakt faktor časopisu. Na druhou stranu, viditelnost časopisů může zvyšovat citační podíl práce až o 80%.

Jestliže porovnáváme hodnocení národních komunit na základě výsledků citační analýzy, můžeme narazit na těžkosti. Aplikace citační analýzy na menší skupiny jako jsou výzkumné skupiny nebo jednotliví vědci je možná jen s velkou opatrností.

⁶⁹ OPTHOF, Tobias, Sense and nonsense about the impact factor.

Impakt faktor:

- Je platným nástrojem pro hodnocení kvality vědeckých časopisům,
- není platným nástrojem pro hodnocení kvality jednotlivých prací,
- není platným nástrojem pro hodnocení kvality jednotlivých vědců,
- není platným nástrojem pro hodnocení kvality skupin vědců jestliže vytvoří méně prací než 100 za dva roky.

Citační analýza:

- By měla preferovat prioritní předpoklad kvality prací pro hodnocení kvality individuálních prací, individuálních vědců a skupin vědců,
- může odrazit úspěšnost výzkumné politiky vlády či univerzity.⁷⁰

⁷⁰ OPTHOF, Tobias, Sense and nonsense about the impact factor.

12. Odměňování ve vědě

12.1. Systém odměňování a fenomén „okupantů 41. křesla“

Systém odměňování ve sféře vědy je podle K. Mertona založen na penězích a na shodě výzkumu. Obraz vědce v jeho očích a veřejný obraz vědců jsou zákonem významnosti.

Podle Craneho se užívá množství publikací (nehledě na kvalitu) jako měření vědecké produktivity. Vysoce produktivní vědci z velkých univerzit získávají uznání častěji než stejně produktivní vědci z menších univerzit. Zuckermanová a Coles zjistili, že vědci, kteří získali nějaké uznání za provedený výzkum na začátku jejich kariéry jsou později produktivnější než ti, kteří uznání nezískali.

Nobelova cena je považována za vrchol ocenění nejlepších vědců. Řada vědců, kteří cenu nezískali a pravděpodobně nikdy nezískají, přispěli stejně nebo možná více k rozvoji vědy než ti, kteří cenu získali. Toto lze popsat podle Mertona jako fenomén „41. místa“ (vychází z toho, že Francouzská akademie věd se rozhodla vybrat 40 vědců za členy jako nositele vědy a to nesmrtelně).

Ti, kteří jednou dosáhli určitého stupně znamenitosti, nespádnou nikdy dolů pod určitou úroveň. Ten, kdo získal Nobelovu cenu, vždy bude jejím nositelem. A vědci se snaží zachovat si znamenitost. Více a více je od nich očekáváno a to je motivem, ale také příčinou stresu.

12.2. Matthewův efekt

12.2.1. Matthewův efekt v systému odměn

Někteří laureáti Nobelovy ceny mohou být ohodnoceni za menší přínos světové vědě než ti, kteří Nobelovu cenu nezískali. Tento model odměňování přeje více zavedeným vědcům. Objevuje se hlavně v případech spolupráce a v případech více nezávislých objevů, které učinili vědci od sebe vzdálení.

Obvykle si více povšimneme jména, které je pro nás známé. A v podstatě všechna jména, která nám nejsou povědomá, jsou anonymní.

Uznání mladého autora práce, která byla napsána ve spolupráci s předním vědcem může mít opačné výhody. Autor, který je identifikován ve spojení s určitým vědcem a později pokračuje výborně, získává náležité ocenění. V mnoha případech mladý vědec skutečně je přijat a jeho práce obecně přijímána, pokud jednou byl spojován s někým známějším. Jestliže přibližně stejné myšlenky nebo zjištění jsou nezávisle sdělovány vědcem s velkou reputací a vědcem, který ještě není příliš známý, s velkou reputací získá hlavní ocenění.

Model špatného hodnocení vědecké práce je popsán K. Mertonem jako „**Matthewův efekt**“ (název byl přejat podle evangelia sv. Matouše - "Matthew" anglicky). „Platí, že pro každého, kdo měl by dát a měl by nadbytek a od toho, kdo nemá, by se mělo vzít stejně jako od toho, kdo má.“

Matthewův efekt spočívá v tom, že se většího uznání za jednotlivé vědecké příspěvky dostává vědcům s větší reputací a odpovídající podíl uznání je odepřen těm vědcům, kteří ještě nejsou tak známí. Nositelé různých ocenění a další význační vědci jsou si vědomi tohoto aspektu. V extrémních případech tito vědci odmítají informovat

spolupracovníky o výzkumu, na kterém spolupracovali, aby nedošlo ke zmenšení jejich ocenění díky jejich méně známým spolupracovníkům. Harriet Zuckermanová zjistila, že významní vědci mají tendenci být na prvním místě prací, na kterých spolupracovali s dalšími vědci.

Ve stále rostoucím objemu literatury vědci hledají určité záchytné body, snaží se, aby nepřehlédli podstatné práce. Jedním takovým záchytným bodem je profesionální reputace autorů. Problém nalezení pertinentní výzkumné literatury způsobuje, že tento efekt ještě narůstá.

12.2.2. Sociální a psychologické základy Matthewova efektu

Někdy je významná přítomnost výtečného jedince ve skupině, který svým charismatem rozšiřuje své nadšení mezi další, povzbuzuje je a umožňuje jim dosáhnout význačnosti.

Někdy mají rozhodující podíl na rozvoj vědy myšlenky více vědců (jedné výzkumné skupiny) nebo jednoho vědce. Přenášejí jejich poznatky normami a hodnotami, které určují významný výzkum. Později nebo po jejich smrti se toto charisma stává institucionalizované ve formě škol, myšlenek a uskutečněných výzkumů.

Podle Mertona je: „... stejně těžké udělat nevýznamný výzkum, často je to dokonce obtížnější, než ten významný...“

Držitelé Nobelovy ceny jsou lidé, kteří se sami řídí, pohybující se sebejistě v nových oblastech bádání. Jsou schopni se přenést i přes velké problémy spíše, než by se zabývali řešením problémů snadnějších a bezpečnějších.

Význační vědci svou sebejistotou diskriminují posouzení, mají tendenci zvýrazňovat významnost jejich přispívání, povyšující je mimo proud publikování více než vědci, kteří mají méně sociálně ověřenou vlastní cenu, kteří se častěji zaměstnávají rutinními vysvětleními.

Ačkoliv významní vědci mohou pravděpodobněji vytvořit významné příspěvky, nejsou zřejmě sami, kteří je vytvářejí. Vědci nezačínají jako význační. Historie vědy oplývá příklady základních prací, které byly napsány poměrně neznámými vědci, jen opomenutými po mnoho let.⁷¹

Nesprávné hodnocení vědecké práce platí i pokud publikovaná práce má jen jednoho autora, dosud neznámého vědce.

Celkový počet vědeckých prací publikovaných vědci se velmi liší. Rozmístění hodnocení je jednostrannější, když jsou práce hodnoceny na základě citací. Price rozšířil termín Roberta Boyla – *invisible college* (neformální výměna poznatků mezi vědci) na označení neformálního kolektivu vědců, kteří vzájemně působí v jejich výzkumu podobných problémů. Tyto skupiny omezené velikosti mohou být ovládnuty interpersonálními vztahy.

Mladí lidé, kteří učinili vstup do oblastí vědy, ale kteří se ještě neprojeví, jsou odsunuti do méně stimulujících pozic pro vědeckou práci, s omezenými zdroji.

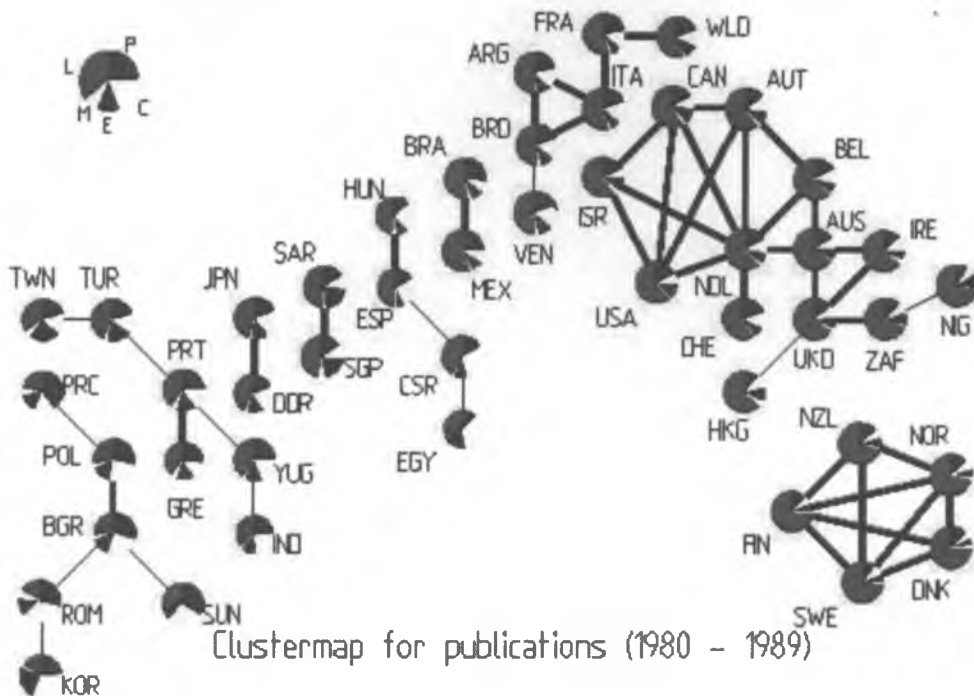
⁷¹ MERTON, Robert K., The Matthew effect in science.

12.3. Hromadění výhod a nevýhod v rámci vědeckých institucí

Centra, která historicky měla větší výkon ve vědě, přitahovala větší zdroje - jak lidské, tak materiální - než výzkumné organizace, které se ještě ve vědě příliš neprojeví. Prestižní univerzity také přitahují pravděpodobně nejnadanější studenty. Avšak i některé katedry na prestižních a bohatých univerzitách jsou chudé. Získávají podstatně méně zdrojů a ocenění. Mohou však své omezené zdroje zaměřit na určité oblasti a poskytovat konkurenčně přitažlivé mikroprostředí talentům první třídy v těchto oblastech.⁷²

12.4. Od výzkumu struktury světové vědy k objevení Matthewova efektu pro země

Na níže je zobrazena koncepce shlukování na základě spolupráce. Byly porovnány odborné dvojice profilů národního publikování, které se dělí na vědy o životě (L), fyziku (P), chemii (C), inženýrské obory (E) a matematiku (M). Obrázek ukazuje na pravé straně země s vysokým procentem vědy o životě v jejich vědecké struktuře, zatímco na levé straně jsou seskupeny země dominující v klasických vědách – chemii a fyzice.



„klasická“ mapa shluků⁷³

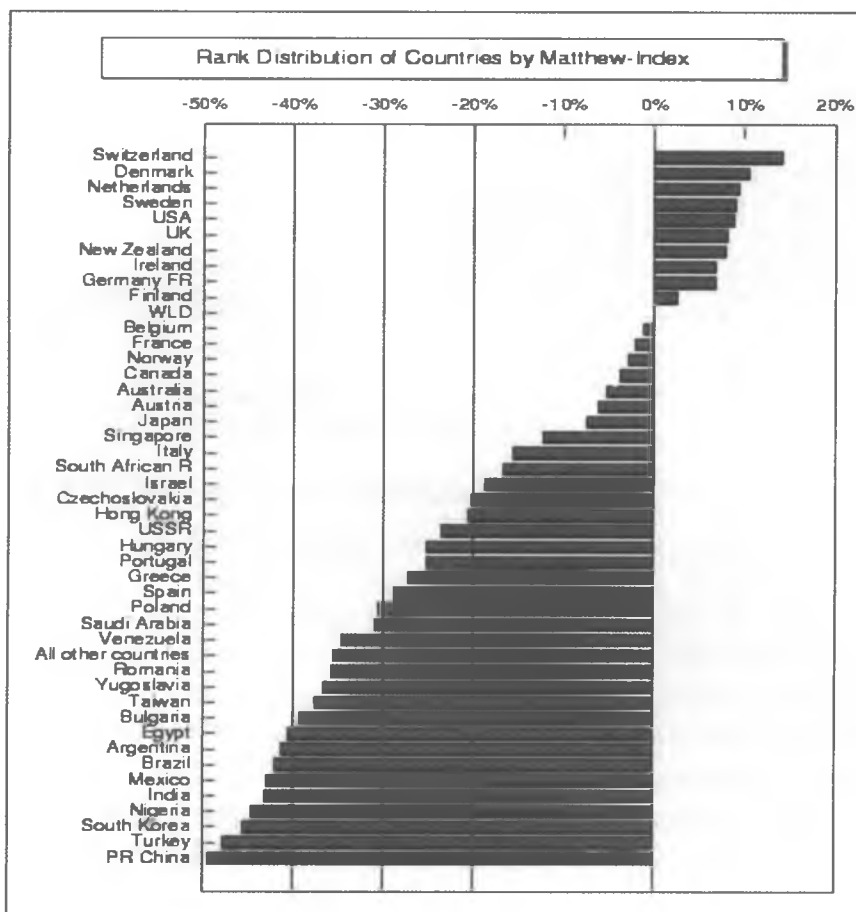
Země jsou řazeny podle Matthewova indexu, který je pozitivní pro země uvedené na pravé části obrázku 15, a negativní pro země uvedené v jeho levé části. Matthewův **index** je měřítkem Matthewova efektu pro země. Autoři napsali: „Malý počet zemí, očekávajících vysoký počet citací na vědeckou práci získala dokonce více citací, než se předpokládalo. Většina zemí, předpokládajících jen malý počet citací na vědeckou práci, získala dokonce ještě méně citací, než se předpokládalo.“

⁷² MERTON, Rober K., The Matthew effect in science II.

⁷³ BONITZ, Manfred; SCHARNHORST, Andrea, National Science Systems and the Matthew effect for countries.

12.4.1. Mikrostruktura Matthewova efektu pro země

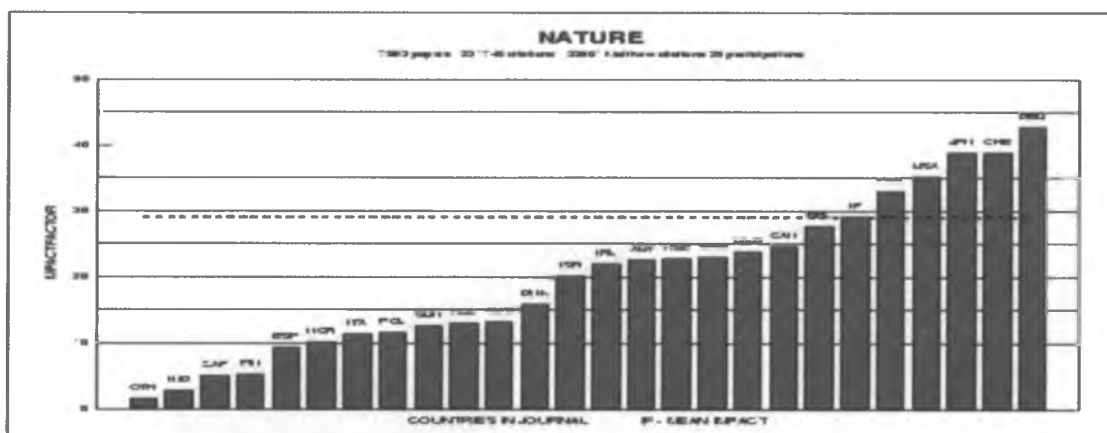
Matthewův efekt pro země (MEC) má malý účinek – je zde zahrnuto jen méně než 10% všech citací. MEC spočívá v systematickém odchýlení počtu aktuálních citací od počtu předpokládaných citací.



“klasické” řazení rozmístění ⁷⁴

Obrázek „Klasické“ řazení rozmístění ukazuje, jak efektivně země využívá svých “vědeckých talentů” v mezinárodní vědecké komunitě. Matthewův index je indikátor národní vědecké kvality. Nicméně interpretace tohoto makro-fenoménu není ještě objasněna. Je nezbytné studovat mikrostrukturu MEC.

⁷⁴ BONITZ, Manfred; SCHARNHORST, Andrea, National Science Systems and the Matthew effect for countries.



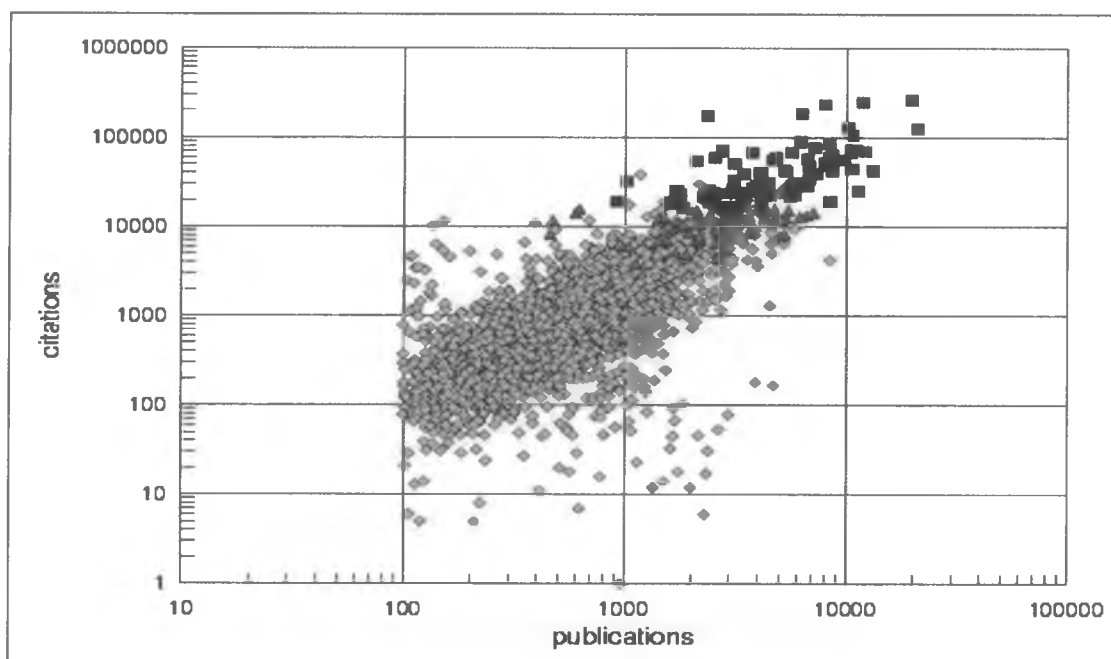
Nerovnoměrné impakty země v časopise Nature (větší obrázek v příloze)⁷⁵

12.4.2. Matthewovy citace a Matthewovy časopisy jádra

Každý časopis má svůj vlastní počet znovu rozmístěných citací. Například časopis *Nature* jich má 33 901, časopis *ANALYTICA CHIMICA ACTA* 1134. Tyto znovu rozmístěné citace se nazývají "Matthewovy citace", protože se ukazují být "atomy" Matthewova efektu pro země. Není potřebné ptát se, zda jistá citace se stane Matthewovou citací. Počet Matthewových citací v časopise může být vypočítán porovnáním počtu očekávaných citací se sledovaným počtem citací pro každou zemi.

Jen 144 časopisů z 2 712 na obrázku *Citace versus publikace v 2712 časopisech* odpovídá polovině všech Matthewových citací (proto polovině MEC). Těchto 144 časopisů se nazývá "**Matthewovy časopisy jádra**". Jedná se o typ vědeckých časopisů, které nikdy předtím nebyly popsány.

⁷⁵ BONITZ, Manfred; SCHARNHORST, Andrea, National Science Systems and the Matthew effect for countries.



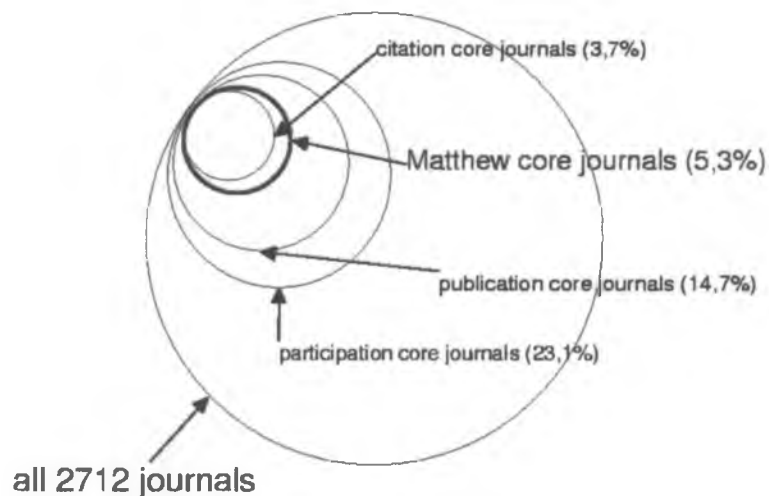
Citace versus publikace v 2712 časopisech ⁷⁶

E. Garfield a další se často zmiňovali o efektech koncentrace ve vědecké komunikaci. Identifikování Matthewových časopisů jádra je dalším potvrzení tohoto významného fenoménu. Shluky časopisů jsou seřazeny podle parametrů jako jsou publikace, citace, Matthewovy citace a účast (počet zúčastněných zemí). Matthewovy časopisy jádra jsou zvýrazněny v oblaku na obr. *Citace versus publikace v 2712 časopisech* a jsou vždy ve speciální pozici. Například Matthewovy časopisy jádra (MCJ) jsou obvykle umístěny v regionech s rozsáhlou publikační činností a s velkými počty citací. Nikoliv vždy platí, že časopis s velkým počtem prací a/nebo citací musí nezbytně být MCJ. Totéž platí pro časopisy se značnou účastí zemí a pro časopisy s vysokým impakt faktorem.

Počty publikací, citací nebo časopisů jádra odpovídají polovině všech počtů prací, citací nebo zúčastněných časopisů v celém příkladě. Časopisy jádra představují obvykle velmi malé procento (viz obr. *Relativní velikost rozdílných typů časopisů jádra*).

⁷⁶ BONITZ, Manfred; SCHARNHORST, Andrea, National Science Systems and the Matthew effect for countries.

Relative size of the different types of core journals



Relativní velikost rozdílných typů časopisů jádra ⁷⁷

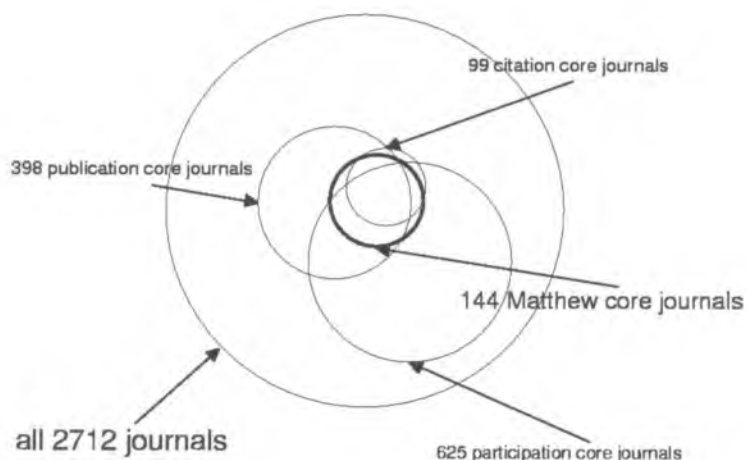
Legenda:

Nejmenší kroužek	- 3.7% citací časopisů jádra.
Zvýrazněný kroužek	- Matthewovy časopisy jádra – 5.3%.
Následující kroužek	- publikace časopisů jádra – 14.7%.
Druhý největší kroužek	- zúčastněné časopisy jádra – 23.1%.
Největší kroužek	- všech 2 712 časopisů.

Pokud čtyři typy časopisů jádra se částečně překrývají, zbývá v našem příkladu okolo 2 000 časopisů, které nejsou ani Matthewovy časopisy, ani publikace, ani citace ani zúčastněné časopisy jádra.

⁷⁷ BONITZ, Manfred; SCHARNHORST, Andrea, National Science Systems and the Matthew effect for countries.

Overlap of the different types of core journals



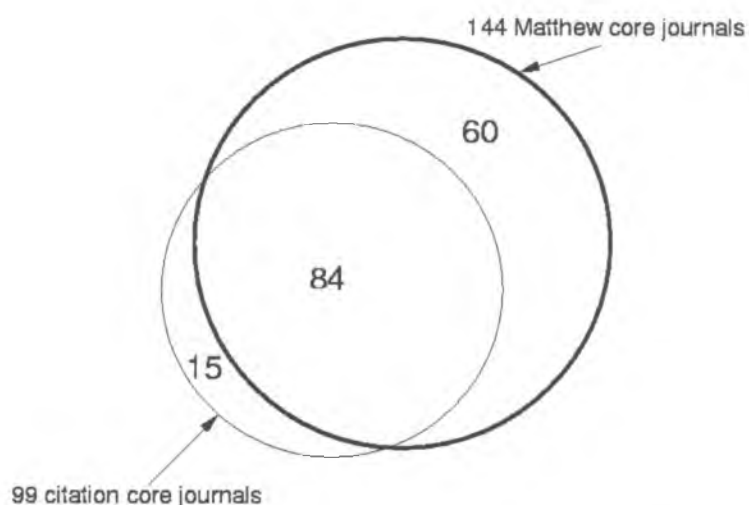
Přesah rozdílných typů časopisů jádra ⁷⁸

Legenda:

- | | | |
|-----------------------------|---|--|
| Nejmenší kroužek | - | zobrazuje 99 citací časopisů jádra. |
| Zvýrazněný kroužek | - | představuje 144 Matthewových časopisů jádra. |
| Další větší kroužek (vlevo) | - | zobrazuje 398 publikací časopisů jádra. |
| Větší kroužek (vpravo) | - | zobrazuje 625 zúčastněných časopisů jádra. |
| Největší kroužek | - | zobrazuje všech 2 712 časopisů. |

Překrytí Matthewových časopisů jádra a citací časopisů jádra je uvedeno na obr. *Překrytí mezi MCJ a citačním jádrem časopisů*. Mezi 144 časopisy MCJ jsou 84 také citace časopisů jádra, 60 nejsou citace časopisů jádra. 15 z 99 citací časopisů jádra nejsou MCJ.

Overlap between MCJ and citation core journals



Překrytí mezi MCJ a citačním jádrem časopisů ⁷⁹

⁷⁸ BONITZ, Manfred; SCHARNHORST, Andrea, National Science Systems and the Matthew effect for countries.

⁷⁹ BONITZ, Manfred; SCHARNHORST, Andrea, National Science Systems and the Matthew effect for countries.

Obr. *Překrytí mezi MCJ a citačním jádrem časopisů* představuje výsledek zkoumání alespoň 7 různých kategorií MCJ. Zobrazuje překrytí mezi časopisy MCJ a citacemi časopisů jádra. Větší kroužek představuje 144 Matthewových časopisů jádra, menší znázorňuje 99 citací časopisů jádra. Pro každý časopis se výzkumníci dotazovali, zda je časopis také členem dalších sad časopisů jádra. Pokud existují tři indikátory, každý s dvěmi možnými hodnotami (jsou a nejsou členy dalšího jádra), předpokládalo se 8 typických kombinací.

K typu A přísluší časopisy MCJ, které jsou také publikacemi časopisů jádra (PUB), citacemi časopisů jádra (CIT) a zúčastněnými časopisy jádra (PART). Jen sedm z možných 8 kombinací bylo nalezeno. Tyto kombinace jsou rozdílně zavedeny pomocí MCJ. Nejvíce zavedené typy A a D mají vztah s vědeckými oblastmi reprezentovanými odpovídajícími časopisy MCJ.

12.4.3. Typologie Matthewových časopisů jádra

TYP				POČET ČASOPISŮ Z MCJ TOHOTO TYPU
A	PUB	CIT	PART	73
B	NEPUB	NECIT	NEPART	6
C	PUB	CIT	NEPART	9
D	PUB	NECIT	PART	47
E	PUB	NECIT	NEPART	5
F	NEPUB	CIT	NEPART	2
G	NEPUB	NECIT	PART	2

Matthewovy časopisy jádra

Legenda:

- PUB/NEPUB – Matthewův časopis jádra je/není publikacemi časopisu jádra.
- CIT/NECIT – Matthewův časopis jádra je/není citací časopisu jádra.
- PART/NEPART – Matthewův časopis jádra je/není zúčastněný časopis jádra.

Exkluzivní role Matthewových časopisů jádra

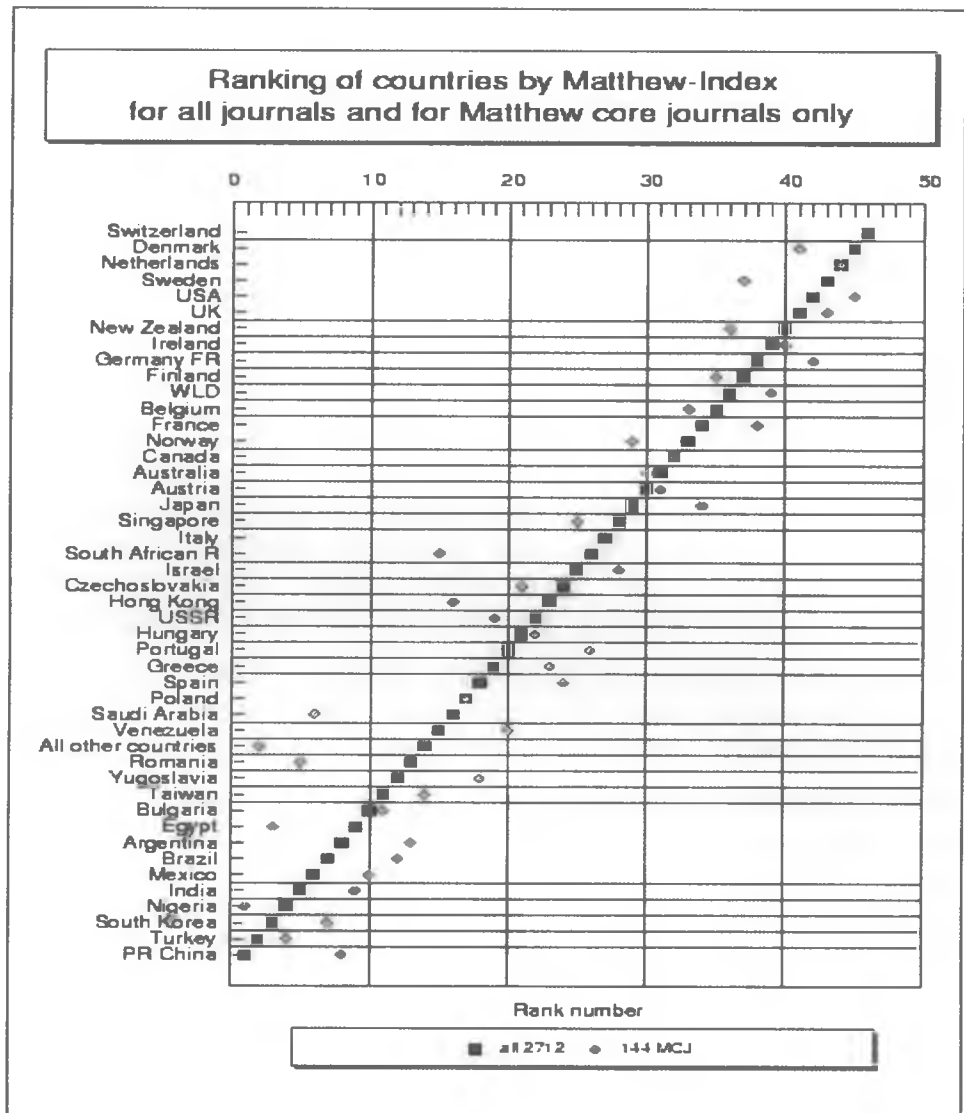
Výzkumné fronty ve vědě se „vaří“ v časopisech MCJ; mnoho zemí v nich soutěží; země dávají patrně své „nejlepší“ vědecké práce těmto časopisům a snaží se shromáždit nadbytek citací nad předpokládané citace, ale také berou na sebe riziko, že nedosáhnou předpokládaného impaktu. Časopisy MCJ jsou z toho hlediska nejvýznamnějšími časopisy ve vědecké komunikaci.

Na obr. 22 je zobrazeno řazení počtu zemí pro dva případy:

- Všech 2 712 časopisů (čtverce),
- jen 144 Matthewových časopisů jádra (kosočtverce).⁸⁰

Obr. *Řazení zemí pro Matthewovy časopisy jádra* znázorňuje časopisy MCJ podle „skutečného“ rozmístění. Země na levé a na pravé straně od značky čtverce zůstávají v jejich regionu. Nedochozí zde k dramatickým skokům z levé na pravou stranu a obráceně.

⁸⁰ BONITZ, Manfred; SCHARNHORST, Andrea, National Science Systems and the Matthew effect for countries.



Řazení zemí pro Matthewovy časopisy jádra⁸¹

⁸¹ BONITZ, Manfred; SCHARNHORST, Andrea, National Science Systems and the Matthew effect for countries.

13. Webometrie

Webometrie, kvantitativní studie fenoménu webu je oblast zahrnující poznatky z informační vědy, počítačové vědy a statistické fyziky. Její metodologie se rýsuje zvláště z bibliometrie.

Thomas C, Almind společně s Peterem Ingwersen razí termín webometrie (webometrics) a hrají klíčovou roli v této startující disciplíně. Vědci z této komunity se nazývají webometrici (webometricians (nebo téměř synonymní výraz cybermetrici – cybermetricians)).

Webometrie se stala částí hlavního proudu konferencí International Society for Scientometrics and Infometrics a kapitola webometrie se objevila v *Annual Review of Information Science and Technology* v roce 2005.

Webometrie obsahuje všechny kvantitativní studie fenoménu webu. Studium hyperlinků typicky čerpá ze zkušeností dospělejší vědy bibliometrie z myšlenek, poskytující spojení k informační vědě. Studium bibliometrie bylo v minulosti analyzování citací článků časopisů jako cíl k proniknout do procesu vědecké komunikace a dopadu vědy.

Jedním z úkolů webometrie je zjištění rozsahu, s nímž užitečné informace o komunikaci ve vědě mohou být získávány z odkazů. Např. je možné počítat odkazy na webovské stránky časopisu dávající nám takto podobnou informaci o impakt faktorech časopisů?

Kvantitativní studia odkazů mají širší potenciál aplikací než bibliometrické studie citací. Některé mají cíle počítačové vědy, jako je zlepšení výkonu webovských search engineů. Další mají více v zájmu cíle informační vědy, jako je používání odkazů jako pomoc při vytváření kolekcí online dokumentů. Dává nám mnoho rozdílných užití webu, webometrická studia často nepopisují vytváření specifických aplikací nebo produktů, ale popisují přirozenost webu nebo podsouborů webu, které budou užitečné pro vytváření webovských aplikací. Např. Menczerův článek se zabývá vlastnostmi webu, které jsou užitečné pro tvůrce tematických crawlerů. Jedna z předností webometrie je, že předmět studia, web, se těší takovému širokému užití, že webometrické výsledky mohou být ku prospěchu v mnoha rozdílných liniích výzkumu.

Cíl obsahového prolézání (content crawling) v prostředí webu je získat všechny jednoznačné stránky z prohledávané oblasti, zatímco jsou vyloučeny duplicity, zatímco odkazový crawler ignoruje všechny obsahové duplikáty, protože ten je jen zaměřen na topologickou strukturu. Obsahové crawlery prolézají nastavenou hloubku a ovlivňují počáteční vybranou sadu výsledků crawleru a schopnost výzkumu produkovat spolehlivé výsledky.

Vlastnosti webovských dokumentů mohou být použity k předpovídání, zda mají vědecký obsah. Webovský projekt Web-TAPIR, získává takovéto soubory. Z výsledků měření vyplývá, že odkazy nebyly zvláště efektivní při oddělování vědeckých a nevědeckých dokumentů. Nicméně soubory PDF často vypadaly, že se jedná o vědecký obsah. Zdá se, že jednoduché webometrické techniky nejsou zcela schopny identifikovat HTML vědecké dokumenty na webu, ale existuje naděje, že zlepšení může být dosaženo díky použitím heuristik založených na obsahu.

V případě čistě elektronických časopisů se ukazuje, že mají větší tendenci odkazovat na zdroje na obecném webu. Formát dokumentů byl shledán jako důležitý rozlišující faktor pro odkazy: HTML články obsahují mnohem více odkazů než PDF, málo PDF obsahuje jakýkoliv hypertextový odkaz. Formát dokumentu je důležitý faktor při rozhodování, zda zahrnout hypertextové odkazy v článku nebo naopak, časopisy v člancích, které potřebují tyto odkazy mohou preferovat formát HTML. Potřeba citovat

webovské materiály se velmi mění dle předmětu a podoboru. Úloha analyzování hypertextových odkazů je komplikovaná díky vydavatelům časopisů automaticky přidávající odkazy k některým článkům.

Filippo Menczer říká že: „Hypertextové odkazy jsou důležité pro řadící algoritmy search engineů a web mining“ (dolování webovských dokumentů). Odkazy mezi stránkami naznačují podobný obsah. Menczer testoval, zda stránka je podobná obsahově stránkám, které na ni odkazují a zda stránky jsou s větší pravděpodobností s podobným obsahem, jestliže jsou spojeny krátkým řetězcem až ke třetím odkazům. Např. akademické stránky v doméně .edu s větší pravděpodobností odkazují v rámci stejného tématu než stránky v rámci domény .com. Vztahy mezi odkazy a obsahem ukazují, že odkazové prolézání v prostředí sousedních relevantních stránkách je s největší pravděpodobností úspěšné.

Věk informací pro hypertextové odkazy může být užitečný při odhalování webovských trendů s možnou aplikací při informačním vyhledávání v prostředí webu. Dokonce tento jednoduchý přístup je schopný odhalit užitečné trendy. Několik příkladů ukazuje, jak rozdílný fenomén může být znatelný díky *time-stamped link profiles* (časové označení profilů odkazu; definované v článku AMITAY, Einat; CARMEL, David; HERSCOVICI, Michael, LEMPEL, Ronny a SOFFER, Aya. Trend detection through temporal link analysis).

Časové odkazové informace mohou být použity rozdílně starými stránkami – ty mají převážně staré odkazy na ně – z *aktuálních autorit* – stránky s novějšími odkazy odkazují na stránky s podobně založenými zájmy. Pravděpodobně search enginey budou preferovat návrat k pozdějšímu typu. Článek informuje, jak časové označení informací může být použito k dosažení výsledků pomocí navrhovaných modifikací dvou existujících algoritmů řazení výsledků vyhledávání.⁸²

Požadavek poloautomatických metod pro vytváření metadat využívá metody kocitací. Dřívější studie užívaly odkazy mezi webovskými stránkami k rozšiřování metadat z jedné stránky na další. Metoda užívá data odkazů ke kolinkovým datům (colink data). Koncepty a teorie kocitační analýzy jsou aplikovány na webovské colinky (colink) k vytváření shluků podobných webovských stránek. Techniky rozšiřují metadata z centra shluku k dalším stránkám ve shluku.⁸³

Knihovnická a informační věda (LIS = Library and Information Science) a souvztažné obory v sociologii vědy a ve studiu vědy a technologie rozvinuly řadu teorií a metodologií – nyní včetně webometrie – zaměřující se na obecné aspekty, jak jsou různé typy informací vytvářeny, organizovány, rozšiřovány a používány různými uživateli v rozdílných kontextech.

Dřívější studie odhalily důležité bibliometrické zákony jako Lotkův zákon rozmístění produktivity mezi vědci (Lotka, 1926), Bradfordův zákon rozptylu literatury o určitém tématu v různých časopisech (Bradford, 1934), Zipfův zákon frekvence slov v textech (Zipf, 1949). Podobně důležité zákony rozmístění byly identifikovány na webu, např.

⁸² Více k tomuto tématu :

Almind, T.C.; Ingwersen, P. Infometric analyse on the World Wide Web : methodological approaches to „Webometrics“. *Journal of Documentation*. 1997, vol. 53, no. 4, s. 404-426.

SMITH, A.G. A tale of two web spaces: comparing sites using web impact factors. *Journal of Documentation*. 1999, vol. 55, no. 5, s. 577-592.

⁸³ THELWALL, Mike; VAUGHAN, Liwen. Webometrics

distribuce TLD (top level domains) na dané téma nebo odkaz (inlink) v rámci webovské stránky. Průlom online citační analýzy se shoduje s pozdější lavinou webometrických studií, které využívají přístup k rozsáhlým webovským datům. Nejednoznačné podobnosti mezi citačními sítěmi a hypertextovými strukturami dokumentů webu způsobily velký zájem od poloviny 90. let 20. století.

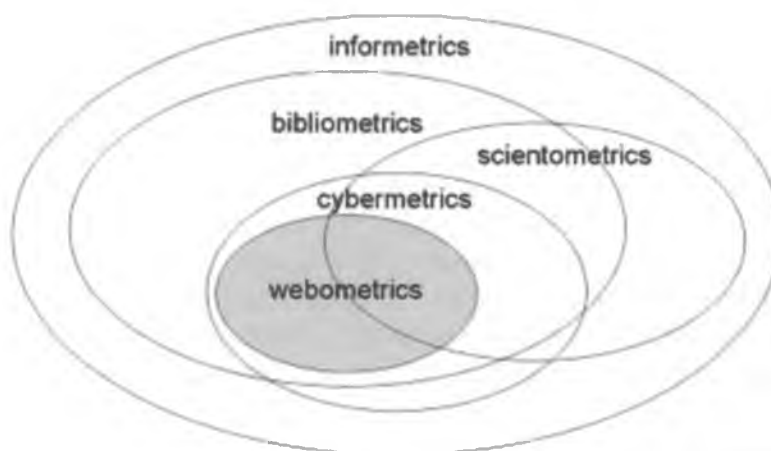
Web je široce používán ve formální i neformální vědecké komunikaci a spolupráci. Webometrie takto nabízí potenciál pro stopování aspektů tradičních vědeckých postů, které jsou více skryté pro bibliometrické nebo scientometrické studie. K nim patří např. využití výzkumných výsledků při vyučování a všeobecnou veřejností nebo aktuální užití vědeckých webovských stránek.

Nové termíny pro perspektivní výzkumnou oblast byly navrhovány od poloviny 90. let 20. st., např. netometrie (netometrics), webometry, internetometrics, webometrics, web bibliometry. Tyto a podobné snahy specifikují různorodost a vývoj v této oblasti a je dosti obtížné identifikovat, co je vlastně v příspěvcích analyzováno. Terminologické změny od internetometrie (internetometrics) k webometrii zaznamenáváme u stejných autorů jako jsou Almindem a Ingwersene, což vnáší do problematiky určitý zmatek.

Tomas C. Almid chtěl zpočátku zkoumat oba aspekty Internetu - sdělovací a síťové a analyzovat typologii, obsahy a charakteristiky národních webovských stránek stejně jako v tradičních bibliometrických analýzách publikací. Ale nebylo jasné, kde Internet končí a web začíná. Amid a Ingwersen se rozhodli pro studii publikační analýzy zveřejněnou v roce 1997, která byla zcela zaměřena na typy a vlastnosti webovských stránek – nepojednávala o komunikaci na Internetu. Proto se objevuje již v názvu tohoto klasického článku koncepce webometrie.

13.1. Webometrie, bibliometrie a infometrie

Web je výzkumným polem pro bibliometrii, scientometrii a infometrii. Webometrie a cybermetrie jsou v současné době dva nejvíce přijatelné termíny v knihovní a informační vědě pro vyvíjející se výzkumné pole. Všeobecně vzato, obrázek s termíny vztahů oborů knihovnické a informační vědy ukazuje, že jsou často používány jako synonyma.



vztah mezi obory knihovnické a informační vědy infometrie, bibliometrie, scientometrie, cybermetrie a webometrie. Velikosti přesahu elips jsou vytvořeny jen v zájmu srozumitelnosti.⁸⁴

⁸⁴ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

Webometrie je definována jako:

Studie kvantitativních aspektů vytváření a užívání informačních zdrojů, struktur a technologií na webu, přičemž se čerpá z postupů užívaných v bibliometrii a infometrii. (Björneborn, 2004)

Pokryté kvantitativní aspekty obou stran vytvoření a užití webu obsahují čtyři hlavní oblasti prezentující webometrický výzkum:

1. analýza obsahu webovských stránek
2. analýza struktury webovských odkazů
3. analýza užití webu (zahrnující přihlašovací soubory uživatelů, chování při vyhledávání a prohlížení)
4. analýza webovské technologie

Všechny čtyři hlavní výzkumné oblasti zahrnují průběžné studie změn dynamiky webu, např. obsahů stránek, struktury odkazů a používání modelů. Tak zvaná webovská archeologie by mohla v tomto webometrickém kontextu být důležitá pro poznání historického vývoje webu, např. významný Internet Archive (www.archive.org). Webometrie jako specifický termín inf. vědy v linii bibliometrie a scientometrie. Webometrie čerpá poznatky z bibliometrie a infometrie a zabývá se metodologickým vývojem webovsky specifických přístupů, včetně inkorporace postupů webovských studií v počítačové vědě, analýzou sociální sítě, hypertextovým výzkumem, studiem médií atd.

13.2. Cybermetrie

Cybermetrie je obecný termín pro:

Studium kvantitativních aspektů tvorby a užití informačních zdrojů, struktur a technologií na celém Internetu čerpajících z přístupů v bibliometrii a infometrii. (Björneborn, 2004)

Cybermetrie takto zahrnuje statistické studie diskusních skupin, seznamů e-mailů a další počítačově zprostředkované komunikace na Internetu. Cybermetrie také pokrývá kvantitativní měření internetové páteřové technologie, topologie a provozu. Šíře pokrytí cybermetrie a webometrie představuje velké překrytí s přístupy založenými na počítačové vědě v analýze webovských obsahů, struktur odkazů, webovského užití a webovských technologií.

Webometrie je uzavřené spojení s bibliometrií a infometrií a klade důraz na perspektivu knihovnické a informační vědy u webovských studií. Termín *Web bibliometrie (Web bibliometry)* byl použit v práci Chakrabartiho a spolupracovníků v roce 2002 a vyjadřuje speciální zájem, jelikož počítačový vědci takto označují pozůstatek z bibliometrického výzkumu, který je přenesen na oblast webovských studií. Další přístupy v počítačové vědě k analýze struktury odkazů také vzdávají hold inspiraci z citačních studií.

Oblast infometrie přesahuje do oblastí jako je bibliometrie a scientometrie jak to naznačuje definice např. Bookeseho, Eggheo a Rousseau, Tague-Sutcliffa. Dle Tague –Sutcliffa *infometrie* je „studie kvantitativních aspektů informací v jakékoliv formě, ne jen záznamů nebo bibliografií a v jakékoliv sociální skupině, ne jen vědců.“. Dále definuje *bibliometrii* jako „studium kvantitativních aspektů vytváření, rozšiřování a užívání zaznamenaných informací“ a *scientometrii* definuje jako „studium kvantitativních aspektů vědy jako disciplíny nebo ekonomické aktivity“.

Oblast webometrie je zcela zahrnuta v bibliometrii, protože webovské dokumenty, texty nebo multimédia, jsou *zaznamenané* informace. Toto zaznamenání může být pouze dočasné, jelikož ne všechny papírové dokumenty jsou řádně archivovány. Webometrie je částečně pokryta scientometrií a jako mnoho vědeckých aktivit je na webu, zatímco další takovéto aktivity, nezaznamenané, jsou dokonce mimo bibliometrii - konverzace mezi osobami. Kromě toho je webometrie zcela zahrnuta v oblasti cybermetrie.

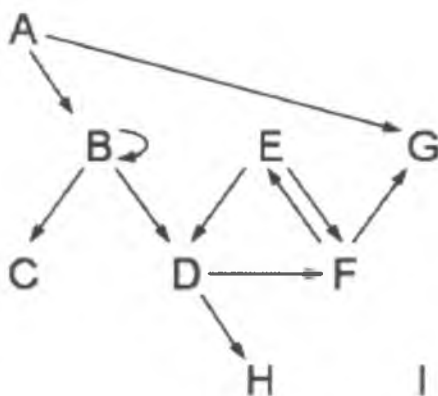
Oblast cybermetrie přesahuje hranice bibliometrie, protože některé aktivity v kyberprostoru normálně nejsou zaznamenané, ale spíše komunikovány souběžně, např. v chatovacích místnostech. Cybermetrické studie takovéto aktivity stále přiřazují do všeobecné oblasti infometrie jako studia kvantitativních aspektů informací „v jakékoliv formě“ a „v jakékoliv sociální skupině“.

Webometrie se rozšiřuje na oblast bibliometrie, stejně jako webometrie nevyhnutelně bude přispívat s dalším metodologickým vývojem specifických webovských přístupů. Jako idea zakořeněná v bibliometrii, scientometrii a infometrii přispívá k objevení webometrie, myšlenky ve webometrii mohou nyní přispívat k rozvoji těchto oblastí.

13.3. Terminologie a diagramy webovských uzlů

Odkaz získaný webovským uzlem (termín *uzel* zde označuje unikání analýzu jako webovskou stránku, adresář, někde se může jednat také o nadřazené domény země). Uzel se označuje jako *incoming link* (*příchozí odkaz*), *inbound link* (*vymezený odkaz*), *inward link* (*vnitřní odkaz*), *back link* (*zpětný odkaz*) a *sitation* (*stránka*).

Základní vztahy odkazů ilustruje pokus vytvořit konzistentní základní webometrickou terminologii pro odkaz (link) vztahující se k webovským uzlům. Web může být viděn jako „tak zvaný“ orientovaný graf, užívající grafický teoretický termín.



Základní vztahy odkazů

Základní vztahy odkazů. Písmenka mohou reprezentovat rozdílné úrovně webovských uzlů, např. webovské stránky, webovské adresáře nebo nadřazené domény zemí nebo všeobecné sektory.⁸⁵

Webovské uzly jsou spojeny pomocí řízených odkazů. Graf teoretických přístupů je užíván v bibliometrii a scientometrii od 60. let 20. století pro analyzování sítí citací a další informace o síti. Přehledový článek Parka a Thelwalla porovnává přístupy

⁸⁵ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

informační vědy ke studování webu s analýzou pro sociální síť. Zjistilo se, že informační věda má tendenci zdůrazňovat platnost dat a studium metodologických problémů, zatímco analýza sociální sítě navrhuje, jak existující teorie může být přenesena do oblasti webu.

Termín *outlink* (odkaz mimo danou stránku) označuje orientovaný odkaz a jeho dva sousedící uzly jsou viděny ze zdrojového uzlu poskytující odkaz, analogicky s užitím termínu citace v bibliometrii. Odpovídající analogie existuje mezi *oblinky* (*oblink*) a *citacemi* s cílovým uzlem jako divákova perspektiva na obrázku *rozdílná terminologie odkazů*. Odkaz křížící hranici webovské stránky, jako odkaz *e* na obrázku *Zjednodušený diagram webovských uzlů* se označuje jako *site outlink* (síťový odkaz mimo původní stránku) nebo *site inlink* (síťový odkaz v rámci stránky nebo webovského sídla) závisející na perspektivě diváka. Podobný význam obsahu terminologie předkládá bibliometrie, např. Price, který klade důraz na konceptuální rozdílnost mezi citacemi, které se hodí k rozdílu mezi outlinkem a inlinkem.

Termín *out-neighbor* (*soused mimo*) a *in-neighbor* (*soused v*) v navrhované terminologii jsou také používány v grafickém teoretickém výzkumu webu. Na webu jsou *self-linky* (to je odkaz na stejnou stránku, kde se odkaz objevuje) používány pro širší rámec než autocitace ve vědecké literatuře. To odráží specifický případ obecného rozdílu mezi outlinky/inlinky a citacemi a odkazy. Stránka *self-link* směřuje od jedné sekce k další v rámci stejné stránky. Stránka *self-linků* (také známá jako vnitřní odkazy) je typický navigační ukazatel z jedné stránky k druhé v rámci jednoho webovského sídla.

Web často představuje webovské stránky vzájemně odkazující mezi sebou, což není normálně možné v tradičním světě tištěných citací. Vzájemné odkazy, takové jako mezi uzly E a F na obrázku *Základní vztahy odkazů*, jsou rozšířený existující webovský termín pro vzájemné inlinky a outlinky mezi dvěma webovskými uzly. Tato vzájemnost není nezbytně zcela symetrická a jako zde může být více odkazů v jednom adresáři mezi dvěma webovskými uzly. Někdy vzájemné odkazy mohou být s rozmyslem vytvořeny tvůrci dvou webovských stránek jako pokus o dosažení lepšího zařazení search enginy využívající počítání inlinků v rámci řadícího algoritmu, jak tomu např. je u Google.

Triadicky odkazující uzly (*triadically linked nodes*; když si propojení webu představíme jako graf, platí pro tyto 2 uzly, že jsou spojeny třemi hranami) D, E a F odpovídají analytickému termínu sociální sítě a sepětí tří hran, používané pro označení pravděpodobnosti, že uzel D a F jsou přechodně spojeny, jestliže již existuje spojení mezi D a E a mezi E a F. V sociální síti jsou takovéto jednoduché struktury vytvořeny bloky větších sociálních struktur.

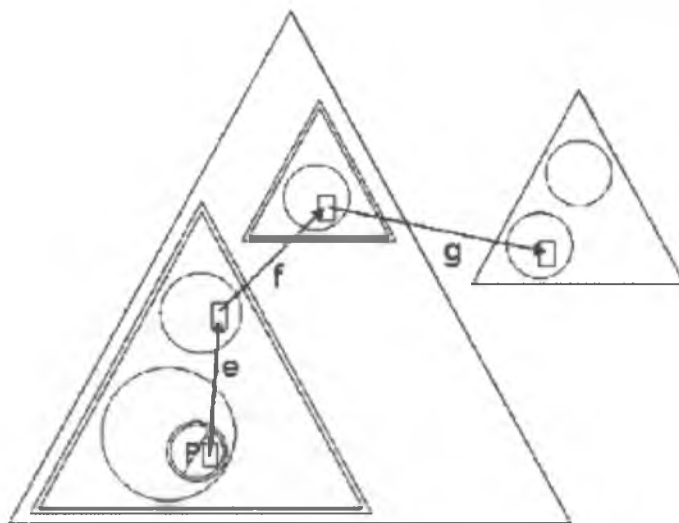
Slovní doprovod pro spojení vztahů na obrázku *Základní vztahy odkazů*

- B je inlink z A, B je odkazován; A je odkazující; A je soused ku B v rámci stejné webovské stránky
- B je odkaz na stránku C; B je odkazující stránka mimo danou stránku B; C je odkazovanou stránkou; C je soused stránky B (mimo danou stránku)
- B má *self-link* (odkaz v rámci dané stránky); B je *self-linking* (samoodkazující)
- A nemá inlinky (odkazy v rámci dané webovské stránky); A není inlinkováno (nemá odkazy v rámci dané stránky)
- C nemá outlinky (odkazy na další stránky); C není outlinkováno (nemá žádné odkazy na další webovské stránky)
- I nemá ani inlinky ani outlinky, I je izolováno

- E a F mají vzájemné odkazy; E a F jsou vzájemně odkazovány
- D, E a F všechny mají oblínky nebo outlinky spojující sebe navzájem mají spojeny tři hrany navzájem (triadically interlinked)
- A má průsečný outlink k G; fungující jako zkratka
- H je dosažitelné z A díky řízené cestě odkazu
- C a D jsou colinked stránkou B (jsou propojeny díky stránce B); C a D mají co-inlinky (mají navzájem propojující odkazy v rámci jedné stránky)
- B a E jsou colinking ku D (jsou vzájemně spojující); B a E mají co-outlinky (mají odkazy mimo danou stránku, které jsou pro obě stránky stejné (jedná se o odkazy, na stejné stránky))
- Co-inlinky a co-outlinky jsou oba dva případy colinků– tzn. totožné odkazy v rámci dané stránky nebo mimo danou stránku



Rozdílná terminologie odkazů pro stejný odkaz závisující na perspektivě diváka jako označované očima⁸⁶



Zjednodušený diagram webovských uzlů ilustrující základní úrovně webovských uzlů (Björneborn, 2004)⁸⁷

Motiv pro podobnou jednoduchou trojhrannou výstavbu bloků komplexních sítí obvykle např. v biochemii, neurologii, ekologii a technice.

Většina odkazů na webu spojuje webovské stránky obsahující příbuzná témata. Nicméně, některé odkazy v rámci sousedních webovských uzlů mohou rozbit takovéto typické modely spojení a spojení podobných tematických domén. Takové (volně definované) *transversální odkazy* (odkazy napříč) fungují jako zkratky napříč tématy a

⁸⁶ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

⁸⁷ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

mohou ovlivňovat tzv. fenomén malého světa na webu. Fenomén malého světa se týká krátkých vzdáleností mezi vzájemným propojením cest mezi uzly v grafu sítě. Např. krátké vzdálenosti mezi dvěmi libovolnými osobami skrze intermediální řetěz známý jako předmět studia v analýze sociální sítě a popularizovaný díky pojmu „šesti stupňů odloučení“. Watts a Strogatz zavedli síťový model malého světa charakterizovaný vysoce shluknutými uzly jako v řádných grafech, dosud s krátkou charakteristickou cestou vzdálenosti mezi párovými uzly jako v nahodilém grafu. Watts a Strogatz ukázali, že velmi malé procento dlouhodobých sítí je dostačující v síti malého světa, aby fungovaly jako zástupci spojení vzdálených uzlů sítě.

Dva co-linkované webovské uzly C a D na obrázku *Základní vztahy odkazů* s co-inlinky ze stejného zdrojového uzlu jsou analogické bibliometrickému konceptu kocitací. Odpovídající, dva co-linkované uzly B a E mající co-outlinky ke stejnému cílovému uzlu jsou analogické bibliografickým vazbám. Co-linky jsou zamýšleny jako obecný termín pokrývající koncept co-inlinků a co-outlinků. Základní předpoklad pro použití obou, bibliometrického a webometrického konceptu, je, že dva dokumenty (nebo dva autoři/odkaz tvůrců) jsou více podobné, tj. více sémanticky propojené, mají vyšší frekvenci sdílení outlinků (odkazů) nebo sdílení inlinků (citací).

13.4. Základní terminologie webovského uzlu a diagramy

Alternativní model dokumentu na obrázku *Zjednodušený diagram webovských uzlů* ukazuje diagram ilustrující některé základní stavební bloky ve shodném rámci webovského uzlu. V diagramu jsou vidět čtyři základní úrovně webovského uzlu, které jsou označeny jednoduchými geometrickými obrázky:

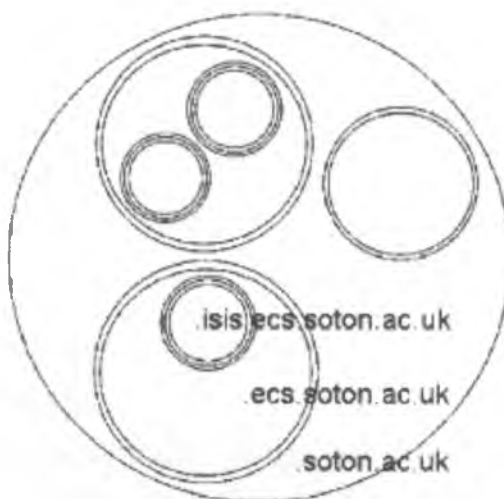
Čtyřúhelníky (webovské stránky), *diagonální čáry* (webovské adresáře), *kruhy* (webovská sídla) a *trojúhelníky* (doména země nebo obecná nadřazená úroveň domény, Tors). Podúroveň v rámci každého ze čtyř základních úrovní uzlu je označena dalšími hranicemi v odpovídajícím geometrickém obrázku. Např. trojúhelník s dvojitým ohraničením znamená obecnou druhou úroveň domény (SLD = second level domain), také známou jako sub-TLD, určenou mnoha vzdělávacími, komerčními, vládními a dalšími sektory vědy, např. .ac.uk, .co.uk, .ac.jp, .edu.au.

Webovský diagram uzlu na obrázku *Jednoduchého diagramu webovského uzlu* ukazuje stránku P lokalizovanou v adresáři podstránky v sub-TLD. Stránka má stránkový outlink e na stránku ve stejném sub-TLD. Outlinkovaná stránka střídavě odkazuje vně ke stránce v jiné sub-TLD ve stejné zemi. Cesta odkazu e-f-g končí na stránce v sídle v jiném TLD.

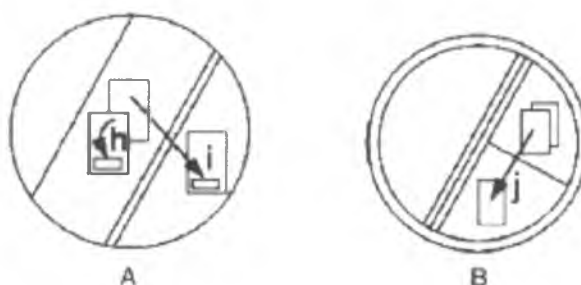
Jedno webovské sídlo může zahrnovat několik podjednotek ve formě podsídel, podpodsídel a tak dále, označované hierarchicky vytvářenými jmény domén. Např. na obrázku *Jednoduchý diagram webovského uzlu webovského sídla*, pod-podsídlo Image, Speech and Intelligent System Research Group (isis.ecs.soton.ac.uk) se nachází v rámci Department of Electronics and Computer Science (ecs.soton.ac.uk), další z mnoha podsídel na University of Southampton ve Velké Británii (soton.ac.uk). Podsídla a pod-podsídla jsou označována jako kruh s dvojitou až trojitou čarou. Podsouřadnice podúrovně by se logicky označovala dalším počtem čar.

Ačkoliv se některá webovská sídla dále rozdělují do odvozených jmen domén, další webovská sídla lokalizují stejný typ podjednotek ve složce adresářů v rámci hierarchie souborů jejich webovského sídla. Tyto různé praxe rozdělení a jmenování komplikují srovnatelnost v rámci webometrických studiích. Na obrázcích *Jednoduché diagramy webovských uzlů webovského sídla a podsídla*, jedna nebo více diagonálních čar

(podobající se šikmým čarám URL a odrážející počet úrovní adresářů níže kořenové úrovni URL) označují adresáře, podadresáře atd.



Jednoduchý diagram webovského uzlu webovského sídla obsahující podsídla a pod-podsídla ⁸⁸



Jednoduché diagramy webovských uzlů webovského sídla a podsídla s odkazy mezi úrovněmi rozdílných adresářů zahrnující subelementy stránky. ⁸⁹

Webové stránky mohou také obsahovat subelementy jako jsou sekce textu, rámec atd. Přidané pásmo ilustruje takové subelementy stránek v cíli samoodkazu stránky *h* a outlinku *i* ze dvou sourozeneckých webovských stránek ke stejnému adresáři na obrázku *Jednoduché diagramy webovských uzlů webovského sídla a podsídla*.

Jakákoliv schématická reprezentace velkorozměrových hypertextových struktur se stane příliš složitou pro praktické použití nebo pro interpretaci jakýmkoliv kvantitativním způsobem. Nicméně předkládané diagramy webovských uzlů s jejich jednoduchými a intuitivními geometrickými obrázky jsou zamýšleny ke zdůraznění a ilustrování kvalitativních rozdílů mezi zkoumanými úrovněmi webovských uzlů v rámci webometrické studie. Obrázek, který následuje v textu, ukazuje příklad takového diagramu webovského uzlu užívaného k ilustraci a zahrnuje a vylučuje webové uzly a odkazy v analýze propojitelnosti britského akademického webovského prostoru.

⁸⁸ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

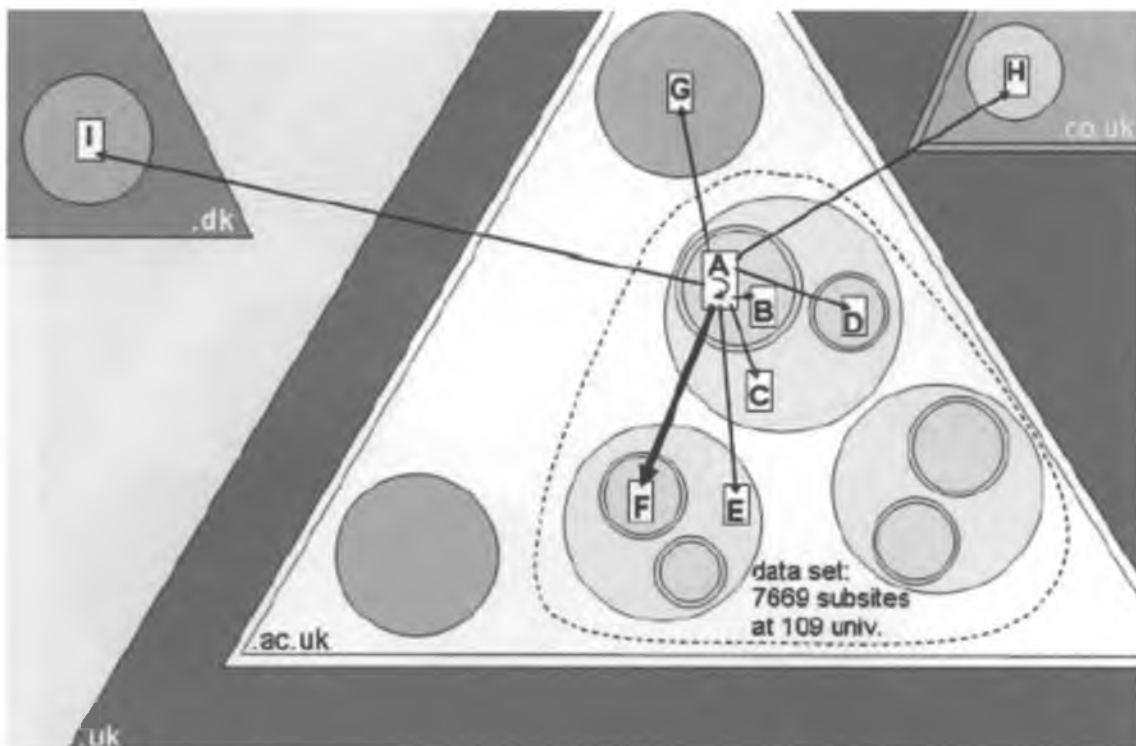
⁸⁹ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

Mikroúroveň webometrie se skládá z webovské stránky, webovského adresáře a malého pod-podsídla - např. které vytváří individuální webovská teritoria. Meziúroveň webometrie se zabývá kvantitativními aspekty velkých podsídel a sídel, mikroúroveň webometrie zahrnuje studie shluků mnoha sídel nebo se zaměřuje na pod-TLD nebo TLD.⁹⁰ Webometrické studie, zahrnující práce např. Larsena, Alminda a Ingwersena, užívají meziúrovně zaměřující se na vzájemnou propojitelnost typu síť-síť, stejně tak jako mikroúroveň analýzy TLD-TLD, primárně posuzují úroveň stránky počtem odkazů. Odkazy mohou být také shromažďovány na různých úrovních uzlu jako v dříve zmíněném Alternativní model dokumentu.

Odpovídající terminologie pro shromažďování souvisejících odkazů by měla zachytit obojí – úroveň odkazu v rámci zkoumání a dosáhnout každého odkazu. Taková terminologie by měla odrážet alespoň tři elementy:

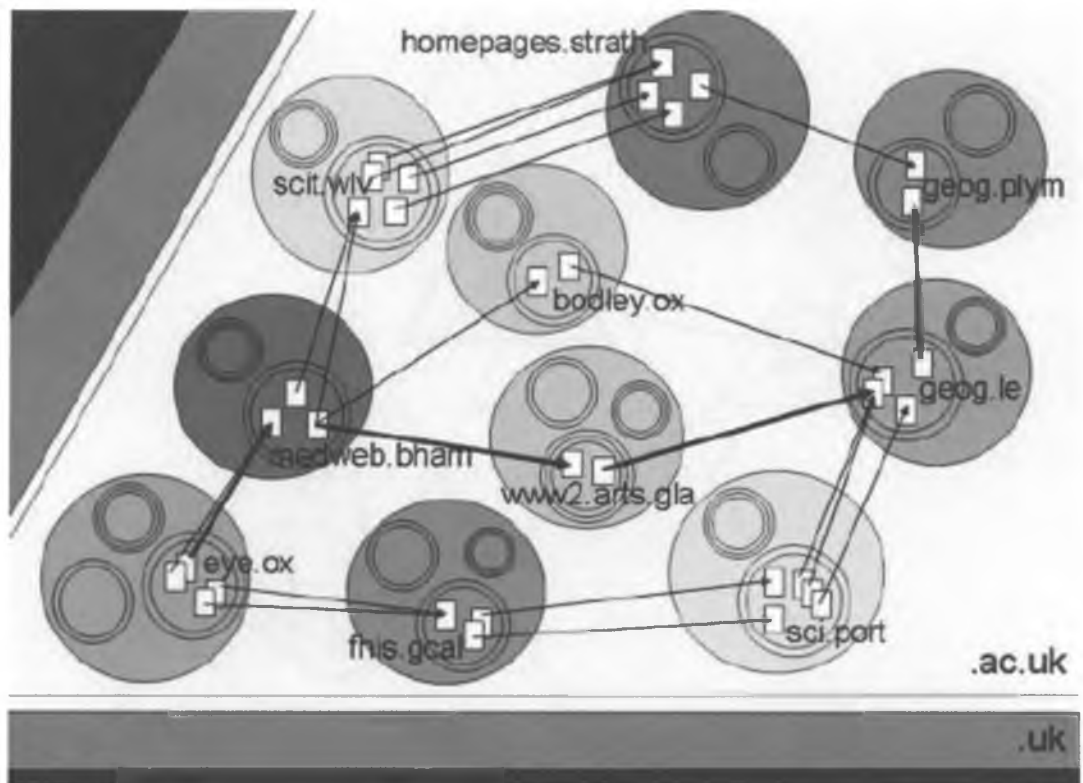
- 1) zkoumání úrovně odkazu
- 2) nejvyšší úroveň webovského uzlu jdoucí napříč odkazem
- 3) perspektivu diváka (obrázek *rozdílná terminologie odkazů*).

⁹⁰ TLD = Top level domény. TLD ukazuje součty odkazovačů dle domény, ve kterých se odkaz na vaši stránku nachází. TLD neboli Top Level Domény jsou přidělovány jednotlivým státům a nezávislým územím, proto lze podle TLD usuzovat, ve kterém státě se odkaz na vaši stránku nachází. V drtivé většině případů se na toto rozdělení dá spolehnout, ale 100% pravdivé to být nemusí, neboť doména není fyzicky nijak vázána k danému území a může být provozována na libovolném místě na Zemi.



Obr. *Příklad diagramu webovského uzlu* ilustrující kvantitativní rozdíly mezi odkazy a úrovněmi webovských uzlů ve webometrické studii. Obrázek ilustruje zahrnutí a vyloučení webovských uzlů a odkazů v analýze struktury odkazů tzv. malého světa napříč prostorem britského akademického webu. Zvýraznění odkazu AF symbolizuje všechny z 207 865 zahrnutých úrovní odkazů mezi 7 669 podsídly na 109 různých britských akademických webovských sídlech. Všechny další odkazy byly vyloučeny: AA (stránka samoodkazu); AB (podsídlo samoodkazů); AC a AD (sídlo samoodkazů); AE (sídlo outlinků k hlavním univerzitním stránkám); AG (sídlo outlinků na stránky ac.uk mimo datovou sadu); AH (pod-TLD outlinky, tj. linky na další britské sub-TLD) a AI (TLD outlinky, tj. linky na další TLD).⁹¹

⁹¹ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.



Příklad diagramu webovského uzlu ukazuje omezený podgraf. Obsahuje výňatek nejkratších cest odkazu (délka cesty 4) mezi podsídla o výzkumu oka (www.eye.ox.ac.uk) a podsídla o geografii (www.geog.plym.ac.uk) pro identifikování stránek a sídel, které poskytují odkazy (skrz různá témata) napříč různými tématickými doménami v britském akademickém webovském prostoru. Zvýrazněné odkazy ukazují jeden příklad nejkratší cesty odkazu mezi dvěma zmíněnými podsídly. Byly vzaty v úvahu jen odkazy spojující podsídla na různých britských univerzitách.⁹²

Jestliže webometrické studie porovnávají právě jednu úroveň odkazu, terminologie může být zjednodušena na pokrytí pouze dosaženého odkazu. V takovém případě l_p je outlink sídla, m_p outlink pod-TLD a n_p outlink TLD. Adresář a úroveň podsídla úroveň odkazů zde nejsou brána v úvahu.

⁹² BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

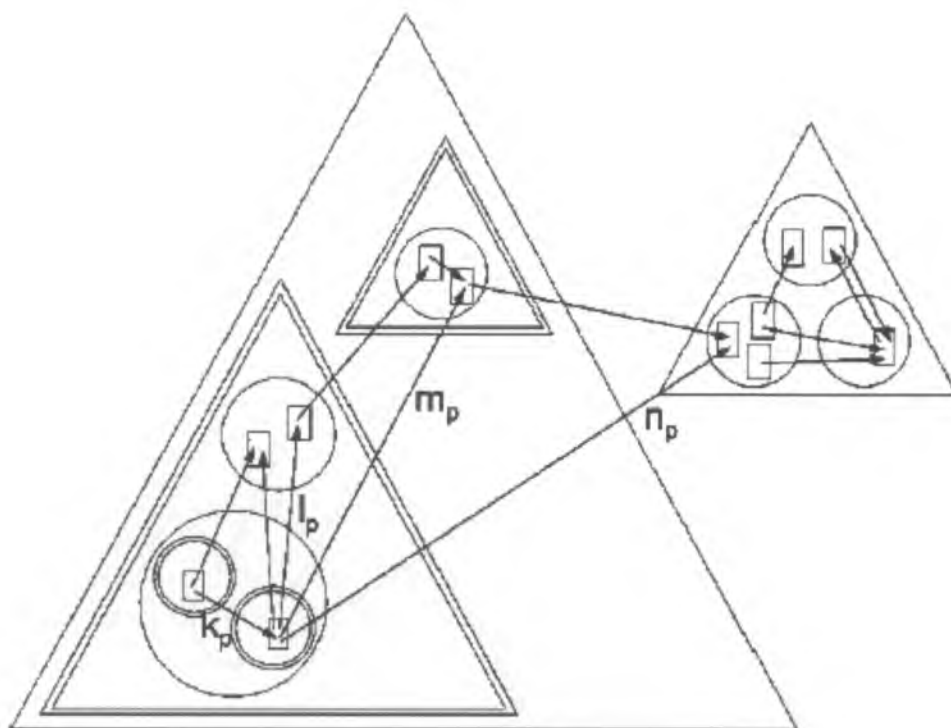


Diagram webového uzlu se stránkou úrovně odkazů ⁹³

Úroveň sídla odkazu vždy spojuje zdroj sídla s cílovým sídlem. Úroveň stránky sídla vždy spojuje zdrojovou stránku s cílovou stránkou. Cílové URL pro webovou stránku může klamně vypadat jako URL pro webovské sídlo. To je webová praxe pro kořen cílového URL hlavní stránky (nebo stránku nadřazenou) vložených stránek webového sídla. Např. místo psaní plného URL www.db.dk/default.htm v cílovém odkazu směřující na nadřazenou vloženou stránku **Royal School of Library and Information Science**, je vhodnější kořen URL k www.db.dk, protože webové servery automaticky hledají hlavní stránky pro kořenová URL. Nicméně toto kořenové URL stále označuje webovou stránku a ne webovské sídlo.

⁹³ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

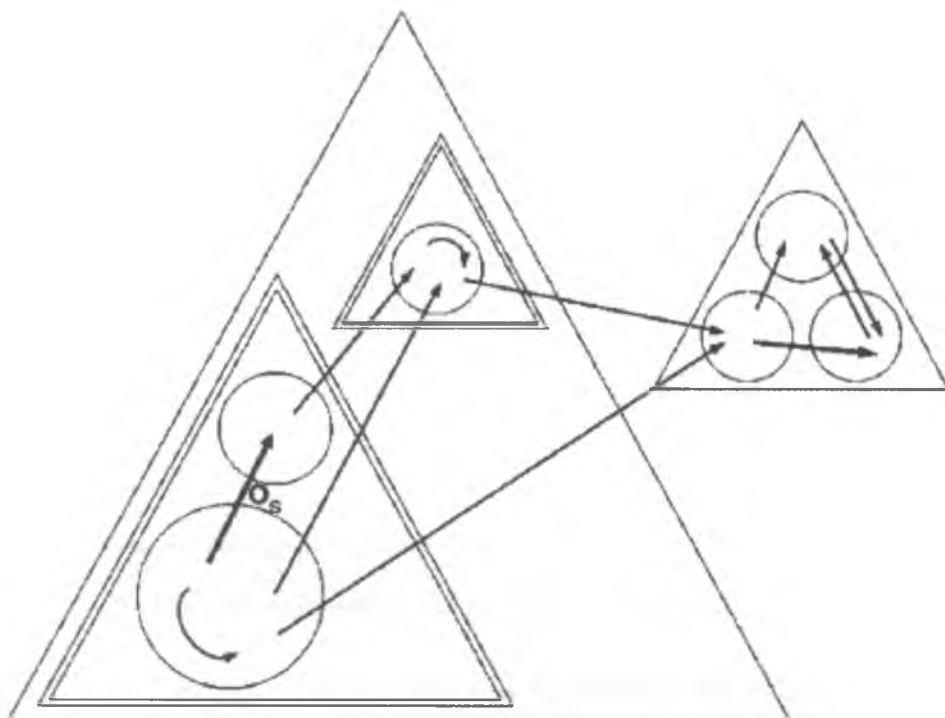


Diagram webového uzlu s odkazy úrovně sídla ⁹⁴

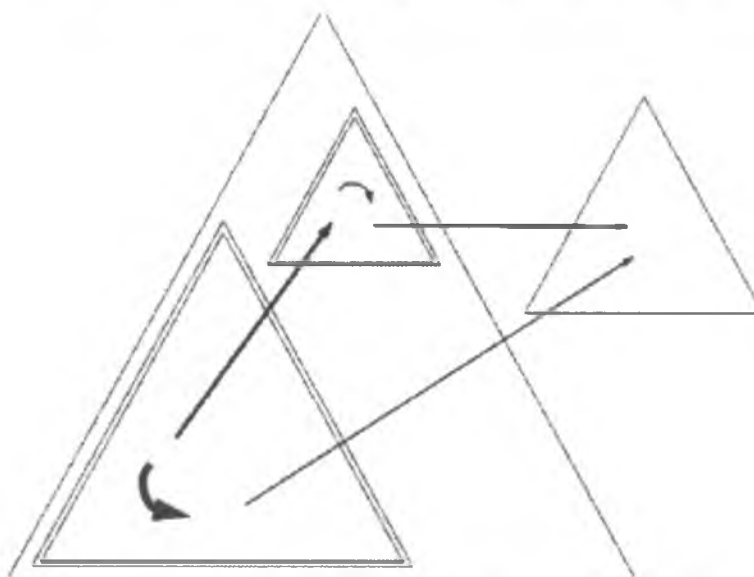


Diagram webového uzlu s odkazy úrovně sub-TLD ⁹⁵

⁹⁴ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

⁹⁵ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

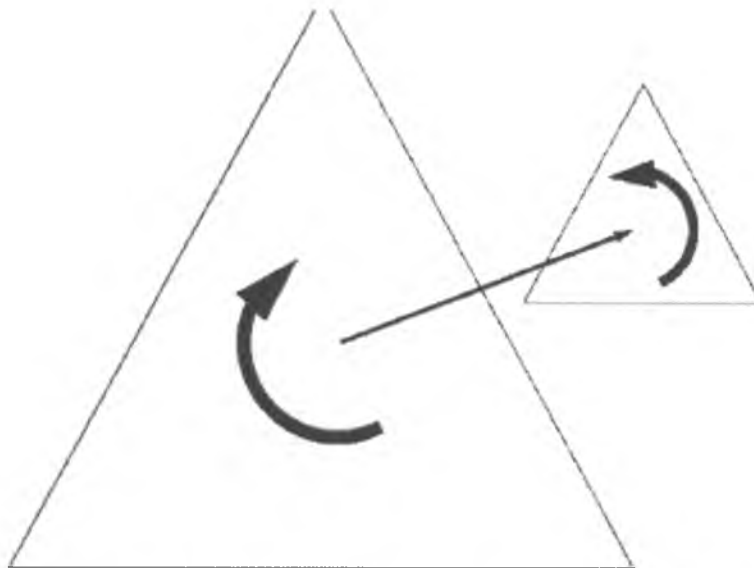


Diagram webovského uzlu s odkazy úrovně TLD ⁹⁶

Termín *station* představený McKiernanem a Rousseau se nezdá být konvenčním označením pro (in)linky. *Station* má stejné terminologické problémy jako termín *citace* a to, že může být interpretován jako synonymum s outlinkem, tj., výstupní odkazy na další práci či stránku.

Analogie existující mezi odkazy a citacemi a outlinky a oblinky existuje. Podobně jako tradiční kocitace a bibliografické vazby jsou technicky podobné pro co-linkované a co-linkující webové uzly. Důvody pro vědecké odkazy na další vědecké práce nejsou ještě zcela ujasněny a jsou rozdílné od poskytovaných outlinků v dynamickém prostředí webu. Nicméně můžeme usuzovat, analyzovat nebo mapovat projev takovýchto aktivit. Proto se analogicky k citační analýze musíme zajímat, jestli jsou vytvářeny interpretace analýzy odkazů v rozdílných webovských prostorech.

Je důležité si uvědomit, co je měřeno a co počítáno. Např. existují dost velké rozdíly mezi počítáním reálných čísel inlinků webovských stránek nebo sídel a počítáním počtů odkazů v sousedství ve formě webovských stránek nebo sídel spojující uvnitř dané stránky nebo sídla (alespoň jeden z některých webovských uzlů). Tento rozdíl je často přehlížen při výpočtech a při aplikování terminologie. Zde je opět analogie k citační analýze - zda počet citací, ne jen počet citujících článků, je počítán. Intelektuální a konceptuální nepřehlednost se zvyšuje hlavně v nově se objevujících oblastech infometrie. Zvláště jestliže bereme v úvahu, že to je přesně počet kocitujících článků, ne aktuální citace, které jsou společně aplikovány pro vypočítání síly kocitací.

Rozdíl mezi úrovněmi webovských uzlů, jejich terminologický dopad a návrh konzistentního diagramu značek jsou nezbytné pro porozumění topologie webu. Např., tento rozdíl je užitečný pro analyzující a ilustrující rozdíly úrovní webovského uzlu.⁹⁷

S rozvojem Internetu vyvstal problém využívání nevědeckých materiálů více než vědeckých např. v pracích studentů.

⁹⁶ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

⁹⁷ BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics.

Studenti dávají přednost elektronickým zdrojům, což je dáno jejich nedostatkem schopnosti rozlišovat důvěryhodné vědecké zdroje od populárních materiálů na Internetu, studenti mají také těžkosti s citováním toho, co našli.

Joann D'Esposito a Rachel Gardner říkají, že studenti jsou si vědomi důležitosti rozlišovat rozdíly mezi důvěryhodností a nedůvěryhodností informací z Internetu. Dle průzkumu studentů byly internetovské stránky nejvyšší kvality a důvěryhodnosti vytvářeny vládami, vzdělávacími institucemi a uznávanými společnostmi a korporacemi.

Při výběru webovských zdrojů před tradičními tištěnými může hrát velkou roli pohodlí. V současném smíšeném prostředí elektronické a tištěné literatury má elektronická literatura velkou výhodu ve snadné dostupnosti.

Změna ve tvoření citací může také odrážet změnu v dostupnosti zdrojů, které byly tradičně pro studentský výzkum méně dostupné jako např. novinky, informace společností a vládní informace. Tyto zdroje jsou nyní dostupné na základě nákupu knihoven v podobě předplatného databází nebo díky dostupnosti informací přímo z webovských stránek novin, obchodních společností a vládních stránek.⁹⁸

Webovské vyhledávací a dolovací nástroje (mining) kombinují textovou analýzu a analýzu (webovských) odkazů s cílem zlepšit algoritmy řazení a procházení webovskými linky. Existuje vztah mezi strukturou grafu webu a textem a významem stránek. Z obsahu odkazu můžeme usuzovat, že stránka je podobná stránkám, které odkazují na tuto stránku a se shluku odkazů můžeme usuzovat, že stránky o stejném tématu jsou shlukovány. Tento předpoklad se často bere jako platný a webovské vyhledávací nástroje jsou stavěné na takovýchto předpokladech.

Search engines užívají kombinaci technik vyhledávání informací a webovských algoritmů prohledávání za účelem indexování webovských stránek. To dovoluje uživatelům pátrat po informacích, tak že se přes webovské rozhraní výsledných databází na tyto informace dotazují. Ačkoliv každý search engine se odlišuje od zbytku tím, že nabízí některé specifické vlastnosti, všichni vykonávají dvě stejné funkce: procházení (což zahrnuje indexování) a řazení (v reakci na dotazy). Nejúspěšnější vyhledávače, kromě marketingových záležitostí, dosahují vysokého pokrytí webu, udržují aktualizovaný jejich index a řazení výsledků hledání předkládají takovým způsobem, který odpovídá představě uživatelské relevance.

Algoritmy řazení a procházení k určitému tématu užívají hlavně dva zdroje informací: slova a odkazy. Pokud uvažujeme o webu jako o fyzickém prostoru, člověk si může asociovat vjem slova s *lexikální topologií*, kde dvě stránky jsou blízké každé další, jestliže jsou podobné v termínech jejich obsahu. Podobnosti v tomto třídění jsou získávány z vektorového modelu, kde jednorozměrový vektor reprezentuje každý dokument či dotaz pro každý dokument či dotaz. Termín, váha a rozměr odhaduje podíl odpovídajícího termínu ve významu dokumentu. Lexikální topologie se proto pokouší posuzovat sémantiku stránek z jejich lexikální reprezentace. *Shluková hypotéza* v pozadí tohoto modelu znamená, že dokument v blízkosti vektoru k relevantnímu dokumentu je také s vysokou pravděpodobností relevantní. Lexikální metriky jsou tradičně používány search enginy pro řazení záznamů dle jejich podobnosti dotazu.

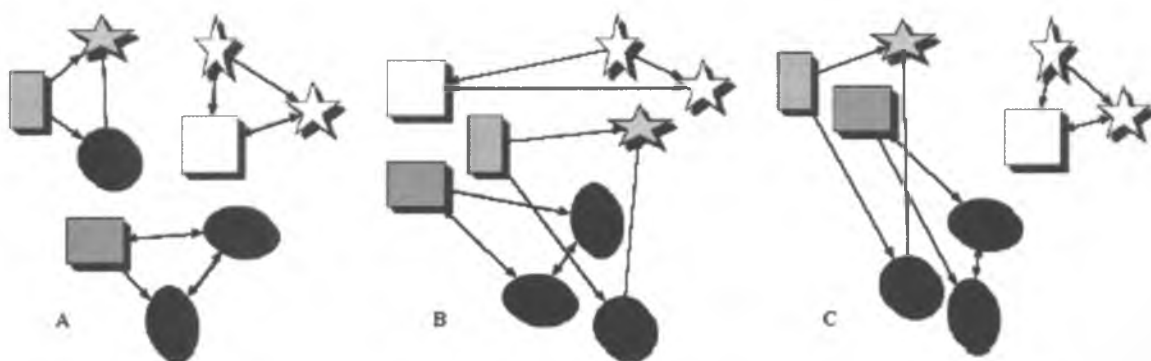
Ačkoliv lexikální topologie je založena na textovém obsahu stránek, *topologie odkazů* je založena na hypertextových komponentech webovských stránek – odkazů. Vjemy

⁹⁸ DAVIS, Philip M. Effect of the Web on undergraduate citation behavior: guiding student scholarship in a networked age

odkazu jsou tradičně užívané crawlers search enginů v úplných, centralizovaných algoritmech. Nicméně poslední generace webovských vyhledávacích nástrojů začíná integrovat lexikální a odkazové metriky ke zlepšení řazení a prozkoumávání výkonu pomocí lepších modelů relevance. Nejlepším případem je metrika *PageRank* užívaná Googlem: stránky obsahující lexikální vlastnosti dotazu jsou řazeny s použitím analýzy odkazu dle dotazu. Důležitost stránky pro další stránky využívá odkazování na ně. Odkazy jsou také používány ve spojení s textem pro identifikování ohniska a autority stránek pro jistý předmět, vymezení reputace daného sídla, průvodce vyhledávacími agenty prozkoumávající jméno uživatelů nebo tématických search enginů.

V pozadí algoritmů vyhledávání, řazení a prozkoumávání je, že užívají analýzu odkazů k vytvoření sémantického závěru, který má vztah mezi topologií webovských odkazů a významem stránek. Pokud přemýšlíme o webu jako orientovaném grafu, můžeme definovat metriku vzdálenosti založenou na nejkratší cestě mezi dvěma stránkami. *Hypotéza shluku odkazů* poukazuje na spojení z topologie odkazů na sémantickou hodnotu– můžeme usuzovat význam stránky, pokud se podíváme na stránky, které odkazují na tuto stránku.

Obrázek *Vztah mezi odkazem A, B lexikální a C sémantickou topologií* ilustruje vztah mezi lexikálním odkazem a sémantickou topologií, která je naznačena hypotézou shlukování a *hypotézou shluku odkazů*. Stránky s odkazy mezi sebou navzájem jsou k sobě, zatímco ve sféře lexikální jsou stránky s podobným textovým obsahem (reprezentované jako tvary) blíže k sobě navzájem. Pokud si představíme sémantickou sféru, v níž stránky s podobným významem (reprezentované jako šedé stíny) jsou shlukovány dohromady. V takovém prostoru by měla být metrika vzdálenosti pozitivně souvztažná s lexikální vzdáleností (hypotézou shlukování) a se vzdáleností odkazů (hypotézou shluku odkazů). Vztah mezi metrikami vzdálenosti naznačuje, že sémantický vztah se přibližuje a může být odvozován od vjemů lexikálních a odkazových.



Vztah mezi odkazem A, B lexikální a C sémantickou topologií⁹⁹

13.5. Hypotéza shluku odkazů

Prvním krokem k vytvoření spojení mezi lexikální topologií a topologií odkazů je poznámka, že jakákoliv daná dvojice webovských stránek (p_1, p_2) má dobře definované funkce vzdálenosti δ_1 a δ_l ve sféře odkazu a lexikální (textové) sféře. K vypočítání $\delta_1(p_1, p_2)$ byla použita struktura hypertextu k nalezení délky, nejkratší cesty v odkazech

⁹⁹ MENCZER, Filippo. Lexical and semantic clustering by web links

z p_1 do p_2 . Zde jsou některé výhrady: Neexistuje metrická vzdálenost, protože není symetrická v orientovaném grafu; metrická verze by byla $\min(\delta_1(p_1, p_2), \delta_1(p_2, p_1))$, ale z hlediska pohodlí δ_1 bude odkazováno na „vzdálenost“ ve zbytku práce.

Crawler má znalost inlinků (odkazů v rámci daného webovského sídla), která by naznačovala volný přístup k search enginu během prozkoumávání. Bylo vytvořeno minimální rozpětí stromu a prozkoumány stránky ve vyčerpávající šíři prvního pořadí. Rozvětvení webovských stránek představuje vážné omezení pro maximum δ_1 , které je nutné měřit. Při počítání $\delta_1(p_1, p_2)$ můžeme použít vektorovou reprezentaci dvou stránek, kde komponenty vektoru (váha) stránky p , w_p^k jsou počítány pro termíny k v textovém obsahu p , dané některým schématem vážení. Jednou z možností by bylo použít Euklidovskou vzdálenost v tomto případě u vektorové sféry nebo jakkoliv další normu L_z .

$$\delta_1^z(p_1, p_2) = \left(\sum_{k \in p_1 \cup p_2} |w_{p_1}^k - w_{p_2}^k|^z \right)^{\frac{1}{z}}$$

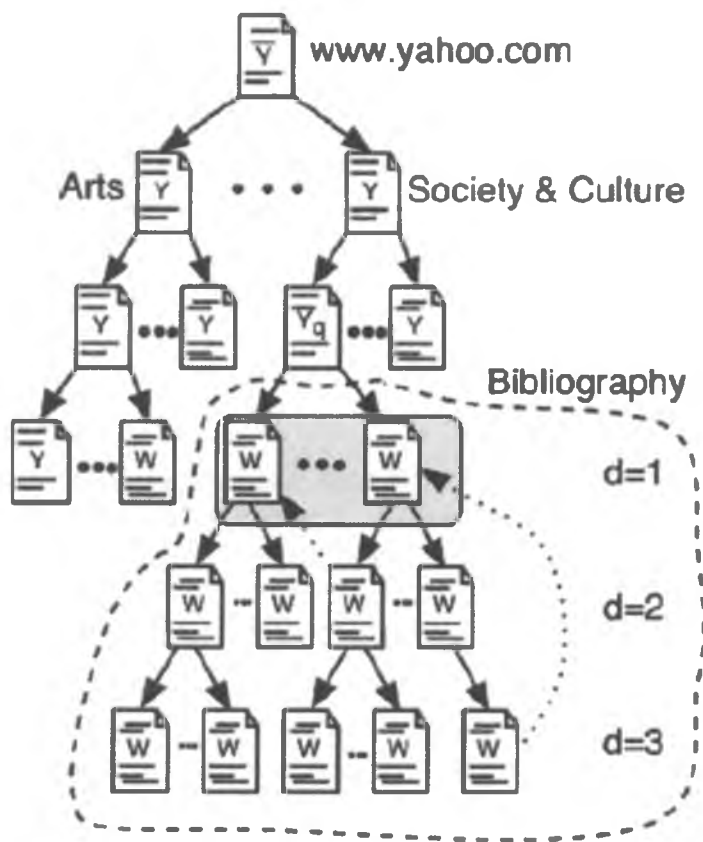
Metriky L_z jsou závislé na rozměrnosti stránek, tj. delší dokumenty mají tendenci mít větší vzdálenost jeden od druhého, než ty kratší. Zde hraje roli skutečnost, že dokumenty s méně slovy mají s větší pravděpodobností nulovou váhu (pro slova, která nejsou zahrnuta), což nepřispívá ke vzdálenosti. Z tohoto důvodu L_z metriky vzdálenosti nejsou použity při vyhledávání informací. Je použito více podobnosti měření, a proto se používá spíše lokalizace slov v dokumentech než jejich absence. Vzdálenost měření založená na podobnosti mezi stránkami byla proto definována:

$$\delta_1(p_1, p_2) = \frac{1}{\sigma(p_1, p_2)} - 1$$

Kde $\sigma(p_1, p_2) \in [0, 1]$ je podobnost mezi obsahem p_1 a p_2 . Použijeme funkci kosinus podobnosti, protože to je standardní měření užívané v komunitě informačního vyhledávání.

Hypotéza 1 σ je antisouvztažná s δ_1 (hypotéza obsahu odkazů).

Představuje měření vzdálenosti mezi dvěma vzdálenostmi měřenými u párů webovských stránek. Soubor užívaný pro tento záměr byl získán tak, že bylo započato od 100 tématických stránek v adresáři Yahoo a provedeno první širší zkoumání každé stránky. Yahoo byl vybrán jako startovní rozbočovač z důvodu velké popularity tohoto portálu.



Reprezentace datové kolekce. 100 tématických stránek bylo vybráno v adresáři Yahoo. Kategorie stránek Yahoo jsou označeny „Y“, externí stránky jsou označeny „W“. Tématické stránky byly vybrány při listování kategoriemi, tj. bez subkategorií. Tento způsob odkazování externích stránek tématickou stránkou (Y_q) reprezentuje relevantní sadu získanou z témat Yahoo (stínované). V tomto případě je to téma BIBLIOGRAFIE a SOCIÁLNÍ KULTURA. Šipka reprezentuje hyperlinky a tečkované šipky jsou příklady odkazů, které odkazují zpět na relevantní sadu. Zkoumaná sada pro téma q je reprezentována uvnitř přerušované čáry.¹⁰⁰

Společné TFIDF vážené schéma (term frequency – inverse document frequency = frekvence termínu – opačná frekvence dokumentu) ukazuje zobrazení každé stránky ve sféře vektoru slov. Tento model předpokládá globální měření vyjádřitelné maximem prozkoumané hloubky (parametr), opačná frekvence dokumentu byla vypočítána jako funkce vzdálenosti mezi počáteční stránkou a sadou dokumentů, s kterou byl vzdálen od zdroje. Formálně pro každé téma q , stránku p , termín k a hloubku d :

$$idf(k, d, q) = 1 + \ln \left(\frac{N_d^q}{N_d^q(k)} \right)$$

$$w_{p,d,q}^k = f(k, p) \cdot idf(k, d, q)$$

Kde N_d^q je velikost souhrnné sady stránky $P_d^q = \{p: \delta_1(q,p) \leq d\}$, $N_d^q(k)$ je velikost podsady stránek v P_d^q obsahující termín k a $f(k,p)$ je frekvence k na stránce p .

Souvztažnost mezi lexikální vzdáleností a vzdáleností odkazů

¹⁰⁰ MENCZER, Filippo. Lexical and semantic clustering by web links

Podobnost $\sigma(q,p)$ mezi každým tématem q a každou stránkou v sadě P_d^q . Vzdálenosti odkazů odpovídající podobnosti měření byly zprůměrnovány na základě souhrnné sady stránek pro každou hloubku:

$$\delta(q, d) = \langle \delta_i(q, p) \rangle_{P_d^q} = \frac{1}{N_d^q} \sum_{i=1}^d i \cdot (N_i^q - N_{i-1}^q)$$

$$\sigma(q, d) = \langle \sigma(q, p) \rangle_{P_d^q} = \frac{1}{N_d^q} \sum_{p \in P_d^q} \sigma(q, p).$$

13.6. Rozklad obsahu podobnosti

Pro analyzování poklesu spolehlivosti lexikálního obsahu odvozeného ze vzdáleností od tématické stránky ve sféře odkazů, může člověk provést nelineární zasazení nejmenšího čtverce v těchto datech pro rodinu modelu exponenciálního rozkladu:

$$\sigma(\delta) \sim \sigma_{\infty} + (1 - \sigma_{\infty})e^{-\alpha_1 \delta^{\alpha_2}}$$

Zde σ_{∞} je úroveň šumu v podobnosti, vypočítaná porovnáním každé tématické stránky s externími stránkami odkázanými z různých kategorií Yahoo:

$$\sigma_{\infty} = \left\langle \frac{1}{N_1^q} \sum_{p \in P_1^q} \sigma(q, p) \right\rangle_{\{q, q', q \neq q'\}} \approx 0.0318 \pm 0.0006.$$

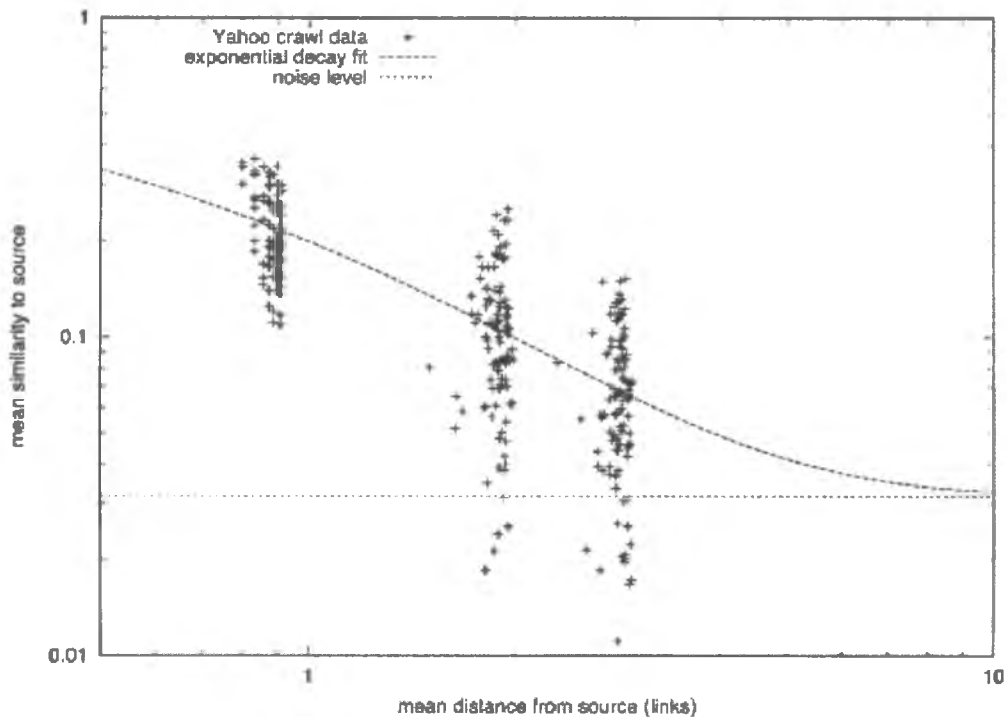
Podobnost rozkladu má podobu křivky s hrubým odhadem toho, jak daleko ve sféře odkazů může jedinec formulovat závěr o lexikálním obsahu.

13.7. Různorodost obsahového rozkladu

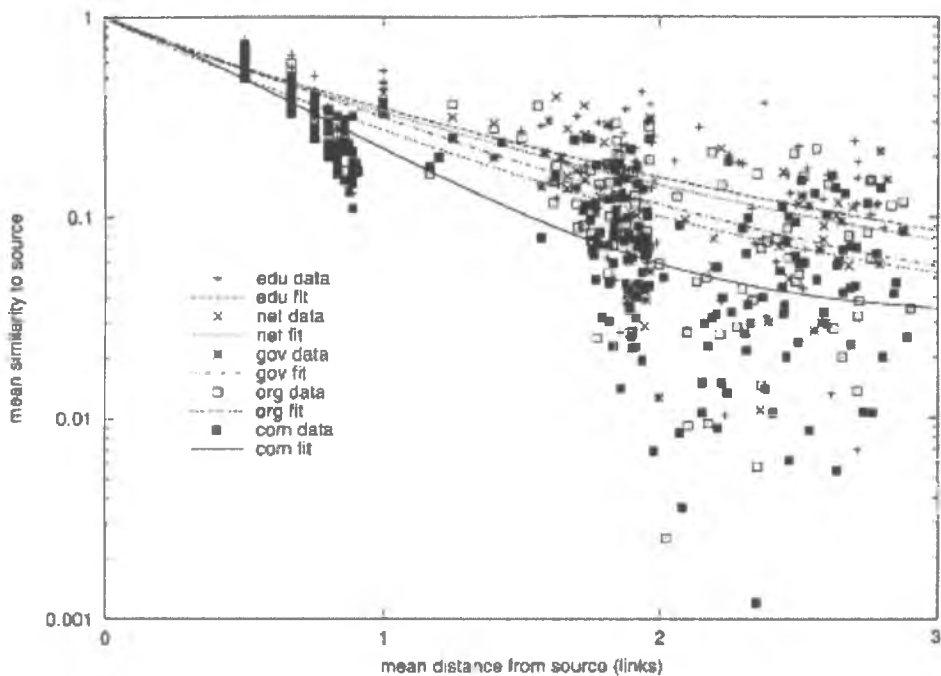
Výsledné sady jsou ekvivalentní k těm získaným crawlery první šíře, které jen sledují odkazy na servery uvnitř každé domény. Tento zakreslený rozptyl $\delta(q,d)$ a $\sigma(q,d)$ měření pro tyto doménově založené crawlery je na obrázku: Zakreslení rozptylu $\sigma(q,d)$ versus $y(q,d)$. Nákres ilustruje různorodost ve spolehlivosti lexikálních závěrů založených na odkazech napříč doménami. Parametry jsou získány ze zasazení dat každé domény k exponenciálnímu modelu rozkladu. Rovnice

$$\sigma(\delta) \sim \sigma_{\infty} + (1 - \sigma_{\infty})e^{-\alpha_1 \delta^{\alpha_2}}$$

udává, s jakou pravděpodobností odkazy vedou k obsahově podobným stránkám v každé doméně. Výsledky naznačují, že, např. akademické weby jsou lépe navzájem propojené než komerční stránky, je na nich lepší odkazování na další podobné stránky. Takové zjištění není překvapivé, když vezmeme v úvahu různé cíle těchto dvou komunit. Tyto výsledky mohou být užitečné při vytváření obecných crawlerů stejně tak jako u tématických prozkoumávajících algoritmů, které vytvářejí priority odkazů založené na textovém obsahu, kde bychom mohli vážít obsah odkazů na základě sídel jednotlivých domén.



Zakreslení rozptylu $\sigma(q,d)$ versus $\gamma(q,d)$ pro témata $q = 0, \dots, 99$ a hloubka $d=1,2,3$. Křivka znázorňuje exponenciálního rozpad dat a podobnost úrovně šumu.¹⁰¹

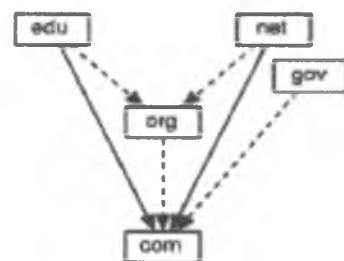


Zakreslení rozptylu $\sigma(q,d)$ versus $\gamma(q,d)$ pro témata $q = 0, \dots, 99$ a hloubka $d=1,2,3$ pro každou hlavní top úroveň domény. Křivka exponenciálního rozpadu je zobrazena pro každou doménu.¹⁰²

¹⁰¹ MENCZER, Filippo. Lexical and semantic clustering by web links

¹⁰² MENCZER, Filippo. Lexical and semantic clustering by web links

Domain	α_1	α_2
edu	1.11 ± 0.03	0.87 ± 0.05
net	1.16 ± 0.04	0.88 ± 0.05
gov	1.22 ± 0.07	1.00 ± 0.09
org	1.38 ± 0.03	0.93 ± 0.05
com	1.63 ± 0.04	1.13 ± 0.05



(vlevo) Parametry exponenciálního modelu rozkladu získané nelineárně křivkou nejmenšího čtverce pro data každé domény odpovídající křivce na předešlém obrázku, s asymptotickými standardními chybami.

(vpravo). Souhrn statisticky významných rozdílů (se spolehlivostí k 95%) mezi parametrickými výpočty, přerušované šipky reprezentují významné rozdíly v jen α_1 a nepřerušovaná čára významné rozdíly mezi α_1 a α_2 .¹⁰³

13.8. Hypotéza shluku odkazů

Hypotéza shluku odkazů je implikována nebo shledávána v mnoha různých formách. Jednou z nich může být nejjednodušší a obecné tvrzení v termínech podmíněné pravděpodobnosti, že stránka p je relevantní s ohledem na nějaký dotaz q , dané tak, že stránka r je relevantní a že p je v rámci d odkaz z r :

$$R_q(d) = \Pr[\text{rel}_q(p) | \text{rel}_q(r) \wedge \delta_1(r, p) \leq d]$$

Kde $\text{rel}_q()$ je binární relevance odhadu s ohledem na q . $R_q(d)$ je relevance pravděpodobnosti dané důkazem sousední relevantní stránky. $R_q(d)$ navozuje myšlenku, že stránka má větší šanci být odkazována, pokud je zahrnuta mezi několika odkazy dalších stránek na dané téma. Nejjednodušší forma hypotézy shluku odkazů je hledána porovnáním $R_q(1)$ a přednostní relevanci pravděpodobnosti G_q :

$$\lambda(q, d = 1) = \frac{R_q(1)}{G_q}$$

Jestliže odkaz sousedů bere v úvahu sémantické závěry, potom následující podmínka musí splňovat:

Conjecture 2 $\lambda(q, d = 1) > 1$ (slabá hypotéza shluku odkazů).

Pro ilustraci důležitosti hypotézy shluku odkazů, vezmeme v úvahu crawlera (nebo uživatele) vyhledávající stránky o tématu q . Vezněme $\eta_q(t)$ pravděpodobnost, že crawler narazí na relevantní stránku v čase t . Můžeme definovat $\eta_q(t)$ rekurzivně:

$$\eta_q(t + 1) = \eta_q(t) \cdot R_q(1) + (1 - \eta_q(t)) \cdot G_q$$

Stacionární záznam poměru dosahuje pro $\eta_q(t - 1) = \eta_q(t)$.

¹⁰³ MENCZER, Filippo. Lexical and semantic clustering by web links

Řešení rovnice předcházející je:

$$\eta_q^* = \frac{G_q}{1 + G_q - R_q(1)}$$

Hypotéza shluku odkazů je nezbytná a vhodné podmínky pro takovýto nahodilý crawler jsou lepší než šance poměrné četnosti, takto ohraničená efektivnost aktivity prozkoumávání (nebo prohlížení) :

$$\eta_q^* > G_q \Leftrightarrow \lambda(q, 1) > 1$$

Definice

$$\lambda(q, d = 1) \equiv \frac{R_q(1)}{G_q}$$

může být generalizována pro faktory pravděpodobnosti přes větší pravděpodobnosti

$$\lambda(q, d) \equiv \frac{R_q(d)}{G_q} \xrightarrow{d \rightarrow \infty} 1$$

a silnější verze hypotézy může být formulována jako následující:

hypotéza 3 $\exists \delta^* > 1$ s. t. $\lambda(q, d) \gg 1$ for $\delta(q, d) < \delta^*$ (zevšeobecněná hypotéza shluku odkazů), kde δ^* je vzdálenost kritického odkazu, kde sémantické závěry nejsou spolehlivé.

Byla zjišťována pravděpodobnost faktor $\lambda(q, 1)$ pro málo dotazů a bylo zjištěno, že $\lambda(q, 1) \gg 1$, ale tyto předběžné výpočty byly založeny na velmi rušeném odhadu relevance. Pro dosažení spolehlivé kvantitativní platnosti silnější hypotézy shluku odkazů bylo měření opakováno na větší a novější datové sadě popsané dříve. Pro předběžný výpočet $R_q(d)$ lze použít relevantní sadu získanou z Yahoo editorů pro každých 100 témat:

$$R_q(d) \approx \frac{|P_d^q \cap Q_q|}{N_d^q}$$

Kde Q_q je relevantní sada pro q . Dále byl počítán podíl odkazů mimo sadu, které směřují zpět ke stránkám v relevantní sadě. Pro G_q lze použít:

$$G_q \approx \frac{|Q'_q|}{|\cup_{q' \in Y} Q'_q|}$$

Všechny relevantní odkazy pro každé téma q jsou zahrnuty v Q'_q , a to dokonce i témata, kde jen prvních 10 odkazů bylo použito při procházení ($Q'_q \supseteq Q_q$) a sada Y ve jmenovateli zahrnuje všechny listované Yahoo kategorie. Do konečného měření pro rovnici

$$R_q(d) \approx \frac{|P_d^q \cap Q_q|}{N_d^q} \quad \text{a} \quad G_q \approx \frac{|Q'_q|}{|\cup_{q' \in Y} Q'_q|}$$

byly zapojeny do rovnice

$$\lambda(q, d) \equiv \frac{R_q(d)}{G_q} \xrightarrow{d \rightarrow \infty} 1$$

pro získání $\lambda(q, d)$ vypočítáno předběžně pro $1 \leq d \leq 3$.

300 měření $\lambda(q, d)$ takto dosažené jsou plánovány versus $\delta(q, d)$ z rovnice

$$\delta(q, d) \equiv \langle \delta_l(q, p) \rangle_{p_d^q} = \frac{1}{N_d^q} \sum_{i=1}^d i \cdot (N_i^q - N_{i-1}^q)$$

Závěr relevantní stránky ve sféře odkazů je vysoce predikovaná relevance, zvýšení pravděpodobnosti relevance pomocí pravděpodobnosti faktoru $\lambda(q, d) \gg 1$ přes rozmezí sledované vzdálenosti a dotazy.

13.8.1. Rozklad předpokládané relevance

Bylo provedeno nelineární zasazení těchto dat do rodiny exponenciálního rozkladu funkcí využívajících 300 bodů jako nezávislé případy:

$$\lambda(\delta) \sim 1 + \alpha_3 e^{-\alpha_4 \delta^{\alpha_5}}$$

Poznámka, že tento tří-parametrový model je komplexnější než v rovnici

$$\sigma(\delta) \sim \sigma_\infty + (1 - \sigma_\infty) e^{-\alpha_1 \delta^{\alpha_2}}$$

protože $\lambda(\delta = 0)$ musí také být vypočítáno z dat ($\lambda(q, 0) = 1/G_q$). Souvztažnost mezi vzdáleností odkazu a faktorem sémantické pravděpodobnosti ($\rho = -0.1, p = 0.09$) je menší než mezi vzdáleností odkazu a lexikální podobností. Regrese dává parametrický propočít $\alpha_3 \approx 1000, \alpha_4 \approx 0.002$ and $\alpha_5 \approx 5.5$. Tato data, která zapadají do modelu exponenciálního rozkladu, nám poskytují informaci, jak daleko ve sféře odkazu můžeme učinit závěr o sémantice (relevanci) stránek, tj. do kritické vzdálenosti δ^* mezi čtvrtým a pátým odkazem.

13.8.2. Implikace pro předmětné webovské crawlery

Pokud vezmeme v úvahu lokalizované crawlery, zaměříme se na stránky v hloubce $d = 1$ odkaz. Z rovnice

$$\eta_q^* = \frac{G_q}{1 + G_q - R_q(1)}$$

můžeme kvantifikovat relativní zvýšení v poměrné četnosti libovolného crawleru s šancí četnosti:

$$\frac{\eta_q^*}{G_q} - 1 = \frac{R_q(1) - G_q}{1 + G_q - R_q(1)}$$

Při použití 100 bodů z $d=1$ sad jako nezávislých případů zjistíme, že pro témata v naší datové sadě, jednoduše startující z dobré sady stránek, poskytuje nahodilý crawler výhodu odpovídající poměrné četnosti zvýšení mezi 50% a 1000%. Toto zvýšení je nepravidelně lineární v $\lambda(q, 1)$. Indukuje, že stupeň, se kterým je hypotéza shluku odkazů platná pro určité téma, má významný vliv na funkční charakteristiku nahodilého

crawleru vyhledávajícího stránky o daném tématu. Takový efekt je pravděpodobně šířen zpracovanějšími tematickými crawlery.

Hypotézy shluku odkazů a obsahových odkazů mají být formalizovány všeobecným způsobem, který charakterizuje vztahy mezi lexikální a odkazovou topologií a mezi sémantickou a odkazovou topologií

Hypotéza obsahu odkazů je empiricky platná pro kvantifikování souvztažnosti mezi lexikální a odkazovou vzdáleností.

Lexikální podobnost zobrazuje hladký exponencionální rozpad v rámci rozsahu několika odkazů.

Značná heterogenita byla zjištěna ve spolehlivosti lexikálních závěrů založených na vjemech odkazů napříč webovskými komunitami široce asociovanými s doménami serverů.

Hypotéza shluku odkazů je empiricky hodnocena, dvě stránky jsou významně pravděpodobněji příbuzné, jestliže jsou v rámci nemnoha odkazů jedna na druhou odkazována.

Pravděpodobnost relevance je v rámci dosahu okolí tří odkazů, poté rapidně klesá.

Být v blízkosti relevantních stránek významně ovlivňuje výkon tematického crawleru.

Existuje mnoho omezení, která musejí být vzata v úvahu. Bylo by vhodné rozprostřít současnou analýzu do hloubky $d > 3$. Naneštěstí, přesné měření vzdálenosti odkazů požaduje znalost nejkratší cesty, a proto je potřeba užít kompletní procházení šire první. Jestliže testujeme odkazy v našem procházení, mohli bychom dosáhnout větší vzdálenosti, ale měření vzdálenosti odkazů by nadhodnotilo skutečnou vzdálenost, protože nejkratší cesty by nebyly garantované. Proto pro daný exponenciální růst procházené sady s d je maximální hloubka limitována našimi současnými výpočtovými zdroji a širokým pásmem zdrojů.

Je velmi obtížné spolehlivě identifikovat relevantní sady pro velký počet témat na webu bez přetřídění manuálně udržované adresáře jako Yahoo nebo Open Directory.

Search enginy obvykle omezují přístup, dokonce i pokud je přístup zprostředkován nástroji jako GOOGLE API. V alternativních přístupech (Menczer) má omezení být krokem k přiblížení se vzdálenosti odkazu přes sousední funkci, která začleňuje kcitaci a bibliografické vazby. Takové přiblížení dovoluje mapovat vztahy mezi nezávislými obsahy, odkazy a sémantickou podobností rozdělenou mezi větším počtem párových stránek.

Výsledky měření prezentované v práci ověřují existenci silné vazby mezi topologií webovských odkazů a sémantickým obsahem. Navzdory decentralizaci webovské struktury, rozmanitosti informací, svobodě obsahu a stylu, hypertextové odkazy vytvořené autory vytváří signál, že lze zjistit, bez ohledu na šum, dané vzdálenosti alespoň tři odkazů.

Interpretace úspěchu algoritmů takových jako PageRank a HITS nám poskytuje novou cestu. V každé z těchto lexikálních topologií, fungujících jako filtr, je užíváno shromažďování sady potencionálních stránek a poté topologie odkazů k identifikování nejlepších zdrojů (např. nejrelevantnější nebo nejautoritativnější stránky).

To je shodné s krátkým řazením odkazů v sousedství, v nichž mohou být detekovány významné lexikální a sémantické signály. Jestliže webovské stránky nejsou shlukovány ve sféře odkazů způsobem, který odpovídá jejich významu, analýza odkazů by nepomohla identifikovat relevantní zdroje.

Souvztažnost mezi webovskými odkazy a obsahem má dodatečně význam ve světle studií odkazové analýzy, která nám říká, že web je „malý svět“ síť se silou zákona uspořádání stupňů odkazů.

Sítě malého světa mohou míchat shluklou lokální strukturu a náhodné odkazy, které vytvářejí krátkou cestu mezi stránkami. Prezentované výsledky ukazují na to, že webová lokální struktura může být asociována se sémantickými shluky vyplývajícími z autorského odkazování stránek na příbuzné zdroje.

Jestliže prolézání neproniká díky uživatelům, kompromis může být velmi neefektivní, např. aktualizování stránek velkým nebo potencionálním zájmem. Uzavření cyklu z uživatelského dosahu zpět k prozkoumávání povede k dynamičtějším a měřitelnějším search enginům, které mohou být lépe přizpůsobovány informačním potřebám uživatelů.¹⁰⁴

Webové odkazy jsou studovány informačními vědci už několik let, ale až z vývoje v posledních letech je jasné, že počty odkazů k vědeckému webu (univerzit a institucí) mohou mít souvislost s hodnocením kvality a důvěryhodnosti webovských stránek. Jsou zkoumány specifické otázky jako třeba, zda obsah stránky a věk stránky hrají významnou roli nebo zda má souvislost počet odkazů na webové stránky časopisu a jeho impakt faktor v databázi Journal Citation Reports. Ve studii *Scholarly use of the web* se používá nová metodologie - užívá se Internet Archiv k získání dat o webovských stránkách. Výsledky ukazují, že věk stránky a obsah stránky jsou významné faktory u disciplín jako je knihovnická a informační věda. Porovnání mezi těmito dvěma obory také ukazuje disciplinární rozdíly v charakteristikách webovských stránek. Vědci a vydavatelé by si měli být vědomi toho, že bohatší obsah webové stránky časopisu vede k vytváření odkazů na tyto stránky, a tím může zajistit vyšší návštěvnost této stránky.

Web jako informační zdroj je předmětem speciálního zájmu informačních specialistů a věnuje se mu mnoho nových výzkumů. Konkrétním zájmem je vědecké využití webu a hypertextových odkazů jako kolektivního zdroje nových informací. Studium webovských odkazů ukazuje, že je možné tímto způsobem získat informace, které byly dříve nedostupné nebo těžko kvantifikovatelné. Na druhé straně pokud může kdokoliv vytvářet webovou stránku a dát ji na Internet a odkazovat na jakoukoliv další stránku, musíme se ptát, jak zde funguje kontrola kvality a zda je srovnatelná s recenzním řízením např. vědeckých časopisů. Další problém je s interpretací počtu odkazů. Webové stránky je obtížné spočítat a mohou se měnit nebo zmizet v čase. Navíc není možné studovat celý web, ať jsou používány jakékoliv nástroje. Navzdory tomu existují konkrétní důvody věřit, jak říká Davenport a Cronin, že je zajištěna metodologická ochrana. Důvody k optimismu jsou dvojnásobné i proto, že byla ověřena série významných korelací mezi metrikami webovských odkazů a provedena další měření v rozdílných kontextech. Úspěch algoritmu odkazu potvrzuje i popularita technologií jako Google a podobných přístupů založených na odkazech, které adoptovaly další hlavní search enginy jako je Alta Vista a NorthernLight.

Webové stránky časopisů by mohly hrát důležitou roli ve vědecké komunikaci ze tří důvodů: zvýšení užití webu jako informačního zdroje uvnitř a mimo akademické prostředí; centralizace časopisů při rozšiřování vědeckého výzkumu; a úžasné zvýšení v počtu časopisů dostupných přes web v posledních letech, včetně nových elektronických časopisů a tradičních tištěných časopisů, které mají svou online verzi. To vede hlavní vydavatelské domy k vytvoření sofistikovaných webovských stránek a

¹⁰⁴ MENCZER, Filippo. Lexical and semantic clustering by web links

k množství diskuzí o přístupu k vědeckému výzkumu. V tomto kontextu existuje potřeba rozvinout model, který pomůže porozumět roli webovských stránek časopisů. Například viditelnost webovských stránek se vztahuje k faktorům jako je impakt asociovaného časopisu, akademické disciplíny, věk webové stránky nebo informační obsah poskytnutý na stránce. Viditelnosti webové stránky (její měření) se projeví v počtu odkazů, které k ní vedou. A čím více odkazů na stránku, tím se zvyšuje pravděpodobnost, že stránka bude navštívena a tudíž bude mít větší vliv v rámci vědecké komunity. Výzkumníci si mohou také přát vědět, jaký vliv jejich výzkum má, což je podmíněno veřejně dostupnou informací, která bude umístěna na webu.

13.9. Studie týkající se daného problému

Výzkum webovských odkazů v informační vědě začíná Larsonem, Rousseauem a Almindem a Ingwersenem. Všichni používají analogii mezi citacemi a hypertextovými odkazy jako pojistku výzkumu tohoto nového fenoménu. Dále pak následují Leydesdorff, Curran a další vědci. Nově rozvinuté techniky zkoumají funkcionalitu search engine Alta Vista k provedení Booleovských příkazů s cílem nalézt stránky ve webovském prostoru, které odkazují na další. Tento postup umožňuje získat přístupy v řádu sekund a shrnuje statistiky zaměřené na propojení odkazů stovek milionů webovských stránek. Ingwersen navrhl nové měření, webovský impakt faktor, jehož modelem je impakt faktor časopisů. Počáteční experimenty vytvářely klamavé výsledky, nicméně problém byl nalezen ve spolehlivosti search engineů a v pokusu utvrdit se, že počty odkazů se budou plnovýznamově vztahovat k dalším známým proměnným, které ztroskotaly. Ale poté se objevila řada pozitivních výsledků: AltaVista, hlavní search engine pro sbírku dat, se stala stabilnější; počty odkazů k univerzitám by mohly být souvztažné s akceptovatelnými ohodnoceními, když data jsou přiměřeně analyzována; odkazy k fakultám by mohly být souvztažné k řazení; pokročilé modely pro počítání webovských odkazů by mohly být dramaticky zredukovány tak, že by byla vyloučeny odkazy, které se konkrétních dat netýkají.

Webové stránky časopisů jsou významné, jestliže vědecké využití webu je modelováno nebo porozuměno, primárně díky jejich asociaci s klíčovým komunikačním médiem časopiseckých článků. Velmi přibližná analogie může být také vytvořena mezi citační analýzou a analýzou odkazů v jejich kontextu a jedna logická otázka je, zda měření založené na odkazech může vytvářet výsledky, které korelují s existujícími měřeními impaktu asociovaných časopisů. Smith studoval inlinky webovských stránek časopisů (to je stránky mimo danou stránku, které obsahují alespoň jeden odkaz na jakoukoliv stránku v rámci daného webového sídla) a zjistil počet těchto stránek, je nespolehlivý indikátor vlivu. Harter a Ford nezjistili žádnou korelaci mezi impakt faktory časopisů a počty stránek inlinků elektronických časopisů a došli k závěru, že měřili něco odlišného. Např. Vaughan a Hysen objevili souvztažnost mezi počtem inlinků webovských stránek časopisu a impakt faktory časopisů pro obor knihovnická a informační věda (LIS) a stejná asociace byla později odhalena pro právo (Vaughan a Thelwall).

Ve stejné době bylo překvapivě málo výzkumů zabývajících se přímo obecnými důvody vytváření webovských odkazů, ačkoliv některá zjištění, která se vztahují k tomuto problému se objevila. Odkazy byly studovány v člancích elektronických časopisů na základě rozhovorů s autory a byly objeveny nové důvody pro vytváření odkazů, které nebyly použitelné v tradičních citacích. Goodrum a další zjistili, že citační modely pro online vědecké články se liší od citačních modelů nalezených v databázích Institute of Scientific Information, protože existuje více online konferenčních článků a ty mají tendenci citovat další konferenční články více než se tomu děje v člancích časopisů.

Kling a McKim rozsáhle zkoumali problém elektronických publikací a došli k závěru, že heterogenita bude trvajícím vlastností vědeckého užívání elektronických komunikačních médií. Nicméně, existuje několik empirických studií, které testují hypotézy jako je vznik sociální informatiky.

13.10. Výzkumné otázky

Asociace nalezená mezi tradičními měřeními vlivu a počty inlinků webovských stránek časopisů je indikátorem toho, že studium hypertextových odkazů k tomuto vědeckému prostoru webu je pravděpodobně informace objevující se v oblasti vědecké komunikace. Studie informuje o pokusu prohloubit porozumění faktorům zahrnutým v odkazování na webovské stránky časopisů ve víře, že webovské stránky jsou stále více významným komponentem vědecké komunikace. Má se za to, že čím více obsahu na webovské stránce časopisu, tím více odkazů, a tím je také vyšší návštěvnost dané stránky. Výchozí model naznačuje, že věk stránky, obsah stránky a disciplína časopisu ovlivní celkově počet inlinků stránky. Dalšími faktory, které mohou mít významný vliv, jsou např. kvalita stránky, ale to by bylo velmi obtížné měřit na základě empirického testování.

Primární hypotézy testované ve studii jsou následující:

- odkazy k webovským stránkám časopisu (ve formě počtu inlinků stránek) bude souvztažné pozitivně s věkem stránek
- odkazy k webovským stránkám časopisu (ve formě počtu inlinků) bude korelovat s typem informací v obsahu stránky
- budou zde rozdílnosti v rámci oborů v měření webovských odkazů, což kontrastuje s oborovými rozdílnostmi v impakt faktorech časopisů

Například právo a informační a knihovnická věda jsou podobné disciplíny v tom, že oba jsou to profesionální obory a mají velkou tradici v aplikovaném výzkumu. Nicméně tyto dva obory převážně nepřesahují svým významem, vyjma jednoho časopisu (Law Library Journal). Kromě toho je užití webu nedílnou součástí praxe informačních profesionálů, zatímco to stejné nelze říci o právu, což poskytuje zajímavý kontrast. Bylo by možné namítat, že pro studii mohly být vybrány i jiné dvě další disciplíny s velkými rozdíly (nebo protiklady, dvě s užší vazbou). Konečně rozdílné vědecké obory jsou ve spojení rozdílnými způsoby a neexistuje žádná optimální dvojice oborů, které by byly ideální pro porovnání. Porovnání rozdílných dvojic může odhalit rozdílné informace a výběr specifických oblastí není zvláště důležitý z hlediska poznání.

13.11. Online nástroje pro sběr dat

Je obtížné zjistit věk webovské stránky časopisu. Z jednotlivých stránek „last modified by“ mohou být sebrána data, ale to nepomůže, pokud nevíme, zda stránka nebyla změněna od data vytvoření původní stránky. Alternativní strategie je napsat e-mail editorovi nebo webmasterovi, ale v době spamů je nepravděpodobné získat dostatečné množství odpovědí a výsledky mohou být ovlivněny krátce trvajícím faktory, např. novější architektura stránek může být volnější v komunikaci o jejím vytvoření. Jednou nověji dostupnou alternativou je nástroj WayBack Internet Archive (www.Archive.org). Je to velmi ambiciózní projekt, který uchovává historický archiv dat vytvářených crawlery webu komerčními search enginey a dalšími zdroji. WayBack nástroj je pojmenován po webovském rozhraní, které kdykoliv poskytuje URL odkaz

k archivované kopii vyvolaných stránek. Tato informace je velmi spolehlivá ve smyslu, že stránka prohledávaná v archivu může být viděna a neexistuje žádný důvod věřit, že data crawlerem poskytnutá jsou jiná než přesná. To může být použito k získání dřívějších informací o času vytvoření URL.

1. Search engine nepokrývají celý web, a tak absence webovské stránky v archivu neznamena, že nebyla již vytvořena, jen že URL nebylo objeveno nebo již indexováno crawlery plnicími archiv.
2. URL je sensitivní na přesná jména souboru a užívanou cestu. Mnoho stránek má více ekvivalentních URL, např. <http://www.blackwellpublishers.co.uk/JOLS/descript.htm> a <http://www.blackwellpublishers.co.uk/journals/JOLS>, ale archiv zaznamená tyto URL odděleně, dokonce i když stránky jsou identické. Abychom získali spolehlivější dřívěji známe datum vytvoření musejí být všechny možné verze URL kontrolovány odděleně.
3. Webovské stránky a sídla mohou být vynechány roboty z indexování tak, že úplné stránky mohou být vynechány. To je případ pro vydavatele Elsevier (<http://www.elsevier.nl/robots.txt> zpřístupněné 30. května 2002).

13.12. Kategorie pro klasifikování obsahu webovské stránky

Základní	Webovská stránka obsahuje základní informace o časopise jako jsou informace o předplatném, členech redakční rady a instrukce pro autory.
Titul	Stránka obsahuje volný přístup k aktuálnímu přehledu obsahu. Musí obsahovat alespoň jeden rok.
Abstrakt	Stránka obsahuje volný přístup k současnemu seznamu abstraktů. Seznam musí pokrývat alespoň jeden rok.
Full text	Stránka poskytuje volný přístup k současným plným textům. Musí pokrývat alespoň jeden rok.

13.12.1. Klíčový faktor 1: impakt časopisu v rámci vědecké komunity

Široce používaný impakt faktor časopisu publikovaný ISI byl využit jako měření vlivu časopisu v rámci vědecké komunity. Potvrdilo se, že impakt faktor časopisu není vhodný pro porovnání mezi jednotlivými časopisy. Ale tvrdí se, že to je vhodný pro měření v rámci disciplíny. Výpočet impakt faktor časopisu podle vzorce je znám, ale pro počítání inlinků stránek nebyla žádná ekvivalentní normalizace představena. To vyplývá z hypotézy, že inlinky se převážně zaměřují na webovské stránky časopisu jako entita, typicky přes jeho webovskou stránku, spíše než individuální stránky (nebo články, když jsou dostupné). Např. Harter a Ford zjistily median 26 inlinků odkazující na články, ale medián 135 inlinků pro celou stránku (či webovské sídlo) užívající výběr webovských stránek elektronických časopisů obsahující rozhodující články. Porovnali celkové citace s celkovými inlinky, ale usuzují, že šlo o odlišná měření.

Spermanovy testy koeficientu souvztažnosti byly použity k vymezení vztahu mezi impakt faktorem časopisu a počty inlinků. Byl použit spíše Spearman než test souvztažnosti Pearsonova koeficientu, protože frekvence distribuce impakt faktoru časopisu a počty inlinků byly velmi zešíkmené. Výsledky testu ukazují významnou korelaci ($p < 0.0$) mezi dvěma proměnnými pro oblast práva a knihovnické a informační vědy (koeficienty korelace 0.33 a 0.48) Stejné výsledky byly získány v dřívější studii

užívající podobnou sadu dat (Vaughan a Thelwall). To znamená, že časopisy mající větší vliv (impakt faktor časopisu) ve vědecké komunitě přitahují více inlinků.

13.12.2. Klíčový faktor 2: úroveň obsahu webovské stránky časopisu

Pro získání kontroly vlivu počtu inlinků byl zkoumán vliv úrovně obsahu počtu inlinků. To je dosaženo měřením poměru odkaz-impakt faktor, jestliže je počet inlinků dělen výsledkem impakt faktoru. Poměr odkaz-impakt faktor byl vypočítán pro každý časopis. Jestliže by impakt faktor byl jen faktor, který by odrážel počet inlinků, poté by poměr odkaz-impakt faktor byl stejný pro všechny časopisy. Rozdílné výsledky poměru odkaz-impakt faktor mezi časopisy dovolují zkoumat efekt dalších faktorů, zatímco je kontrolován efekt impakt faktoru. Ukázalo se, že webovské stránky, které více lákají svým obsahem, přitahují tak více inlinků relativně k jejich výslednému impakt faktoru.

13.12.3. Klíčový faktor 3: věk webovské stránky časopisu

Bylo zkoumáno, zda věk webovské stránky má vztah k úrovni obsahu stránky. Pokud by měl vztah, pak by vztah mezi věkem a poměrem odkaz-impakt faktor časopisu identifikoval, což by se přisuzovalo vztahu mezi úrovní obsahu a poměrem odkaz-impakt faktor. Výsledky testu ukazují nedostatečný důkaz, že úroveň obsahu ovlivňuje věk. Asociace mezi věkem a obsahem není statisticky významná.

Studie představila novou metodologii, kterou užívá Internet Archive pro kolekci datum a vztahující se informace. Potvrdilo se, že vyšší impakt časopisu, daný vyšším impakt faktorem časopisů přitahuje více inlinků. Ačkoliv pozitivní zjištění zaměřující se na počet odkazů aplikuje jen část webu pokryté Alta Vistou, tento search engine se nezdá mít dobré pokrytí univerzitních webovských stránek a zdá se rozumné předpokládat, že jeho pokrytí nebude systematické ve smyslu stránek, které by odkazovali na konkrétní klasické časopisy. Tento výsledek slouží jako dílčí ujištění, že nezbytnost omezení sbírky dat k oblasti pokryté search enginem nezruší platnost dosažených závěrů.

Zkoumáním obsahu webovských stránek vedlo k nalezení různé úrovně obsahu informací. Některé stránky poskytující jen základní popis časopisu, zatímco další obsahují názvy, abstrakty nebo volný přístup k plným textům článků. Porovnání ukazuje, že webovské stránky, které poskytují více informací, přitahují více inlinků. Spojení mezi obsahem a inlinky poskytuje důkaz o významu webu pro vědeckou komunikaci: časopisy, které poskytují více informací online, přitahují více zájmu. To potvrzuje názor, že informativní webovské stránky jsou hodnotné pro časopisy nejen v oblasti technologických LIS, ale také v oblastech jako je právo, pro které technologie není tak nedílnou součástí disciplíny. Existuje logická možnost, že časopisy, které existují online déle a s lepším obsahem, zvyšují svou viditelnost natolik, že se tento faktor odráží v jejich impakt faktoru. To si ale zasluhuje ještě podrobnější zkoumání.

Věk webovské stránky také ovlivňuje počet inlinků, starší jich získávají více. Obecně nově vytvořené odkazy (to je odkazy na nejnovějších stránkách na webu nebo nové odkazy, které byly nedávno přidány na starší stránky) jsou indexovány AltaVistou s menší pravděpodobností. Jedno vysvětlení pro souvztažnost je, že to znamená pro autory více času webovské stránky identifikovat a poté, po jejich objevení, vytvořit odkazy na webovské stránky časopisu. Alternativní vysvětlení je, že webovští autoři nepravidelně nebo nikdy neaktualizují některé stránky, které jednou vytvořili.

Byl nalezen důkaz pro určení, že webovské stránky časopisů s lepším obsahem jsou viditelnější a přitahují více odkazů a tudíž potencionálně jsou více navštěvovány. To je povzbuzením pro vydavatele, aby na svých stránkách umísťovali co nejvíce informací. Podobně, vědci mohou vzít v úvahu obsah webovských stránek jako faktor, když se rozhodují, které časopisy si předplatí, a to dokonce i v případě tištěných časopisů. Věk

webovských stránek je ovlivněn viditelností stránky - starší webovské stránky jsou viditelnější. Proto změny URL adresy nejsou žádoucí, protože mohou mít negativní vliv na viditelnost webovské stránky, a tak redukovat návštěvnost stránky. Věk webovské stránky je komplikující faktor, který by měl být vzat v úvahu, když užíváme počty odkazů k porovnání webovských stránek. Ačkoliv počet stránek inlinků měří vliv způsobem, že se asociuje s průměrem impaktu citací. Ukazuje se, že aktuální měření jsou někdy rozdílná, a proto by měla být použita volným způsobem. Citace mohou měřit vliv v rámci bezprostřední vědecké komunity, zatímco webovské odkazy mohou měřit vliv širšího okruhu včetně lidí z praxe a studentů.¹⁰⁵

Webovské prohledávání pomocí search enginů je dnes již běžnou praxí. Procházení jako proces identifikování a zobrazování webovských stránek, obvykle se chápe jako indexování pomocí procházení grafu webovských odkazů ze souboru startovacích bodů. Webovští crawleři se stěhují z uzlu (nebo dokumentu) do uzlu pomocí hyperlinků, které každý uzel obsahuje a který definuje ostří webovského grafu. Tato technika webovského prohledávání (crawling), spojená s indexováním dokumentů a procedurami vyhledávání, demonstruje zvláště úspěšnost v podporování systémů webovských search enginů. To pomáhá vyhledat relevantní obsah přístupných webovských dokumentů uživatelům. Webovské prohledávání je také technika sbírání dat, která podporuje strukturální a infometrickou analýzu webovského grafu. Jedna z forem analýzy webovského grafu bere v úvahu inlinky a outlinky z uzlů, které reprezentují dokumentové citace a odkazy. Obsahové prohledávání se vztahuje k předmětu nebo k cíli procedury webovského prohledávání (převzaté k podpoře webovských search enginů) užívané ve spojení s objevením a indexováním obsahu dokumentů, které tvoří web. Obsahové prohledávání např. ignoruje duplikáty dokumentů a takto se liší od prohledávání odkazů, které to nedělají.

Technika může přinést konzistentní výsledky, které ovšem nemusejí být platné. Vědci potřebují přístup k hodnotným a platným technikám, jestliže jsou jejich zkoumání založena na soudržném a konzistentním základu znalostí. Spolehlivost společně s problémem platnosti může být důležitým parametrem při výběru jednotlivých postupů webovského prohledávání.

Borgman a Furner identifikovali "explicitní aplikaci bibliometrických metod do studia world wide webu". Tyto metody a jejich aplikování do infometrické analýzy webu jsou přezkoumávány Bar-Ilanem a v práci Bar-Ilana a Peritze. Existující search enginey poskytují nespolehlivý základ pro infometrickou analýzu webu a autoři poznamenali, že platnost těchto infometrických zjištění by měla být podrobena kritice. Mettrop a Nieuwenhuysen jsou důraznější a varují, že kvantitativní analýza webu je narušena rozporuplností v reakcích search enginů, což oni zjistili. Identifikované těžkosti by mohly mít původ v proměnlivosti, jak operátoři search enginů manipulují s výsledky prohledávání, což ale nesevřídčí o spolehlivosti prolézání.

Hawking a spolupracovníci si všimli, s ohledem na jejich experiment informačního vyhledávání, "že zdroje problému (špatné efektivnosti search enginu) mohou spočívat v pavučině webu", ale Amitay a spolupracovníci tvrdili, že data search enginu "jsou velmi závislá na konfiguraci (nebo politice) crawlera. Thelwall se podílel na vyhledávání webovských dokumentů v průzkumech, které se týkali crawlerů a říká, že "rozdílné metodologie výběru stránek mohou přinést rozdílné výsledky na stejné otázky a, že design parametrů crawleru může ovlivnit průzkum výsledků. Varoval, že informační

¹⁰⁵ VAUGHAN, Liwen; THELWALL, Mike. Scholarly use of the web

vědci by si měli být vědomi rozdílů při zkoušení prohledávání a že malé příklady (vzorky) mohou být nereprezentativní.

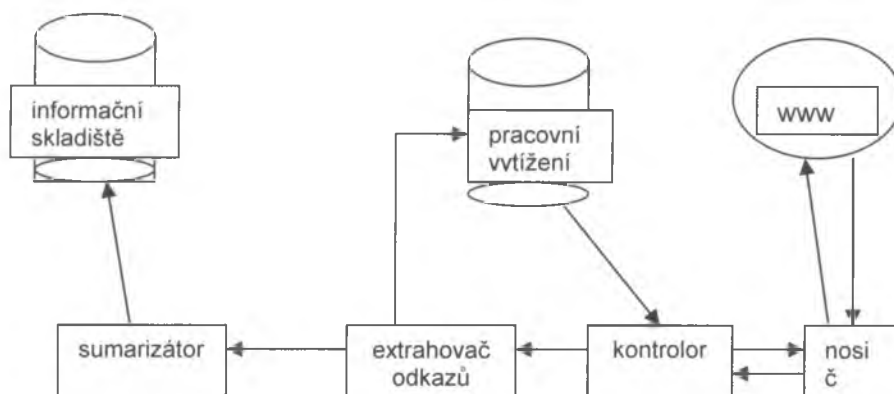
Je vhodné identifikovat duplicitní webovské dokumenty tak, že crawler je ignoruje. To je užitečné pro zamezení redundance při prozkoumávání a zvýšení indexování a naopak to umožní snížení počtu síťového provozu a pracovního zatížení serveru způsobené webovskými crawlery. Primární požadavek designu webovských crawlerů je přizpůsobit se etice webovského prohledávání. To znamená respektování práva požadavků a minimalizování provozu a přetížení serverů.

13.13. Terminologie webovského prohledávání

Obrázek *architektura web crawleru* : nosič (fetcher) požadavků webovských dokumentů (nebo zdrojových uzlů) pod dozorem kontrolora (controller). Hyperlinky jsou získávány ze získaných dokumentů pomocí extrahovače odkazů (link extractor) a jsou přidány k „nákladu“ webovského crawlera. Proto dokumenty nebo citované uzly, ke kterým tyto outlinky odkazují, jsou samy přinášeny.

Sumarizátor komponentů webovského crawlera postupuje obsahem dokumentu a ukládá informace, které ohodnotí jako relevantní pro záměry webovského crawlera. Tato součást sumarizátora nese odpovědnost ze indexování dokumentů společně s uchováváním informací a odpovídá zájmům komunity vyhledávání informací.

Dnes již nestačí extrahovat informace (link extractor) pomocí webovského crawlera jen jako hyperlinky obsažené v dokumentech vytvořené hypertext markup language (HTML). Pokročilý design webovského crawlera by měl umět extrahovat potenciaální uzly i dalších dokumentových formátů jako plaint text (.txt), Microsoft Word (.doc) a Adobe přenosný dokumentový formát (.pdf) , stejně tak jako by měl umět projít zkomprimované dokumenty. Pokročilé designy crawlerů se umí také přenést přes webovskou sféru (nebo jednotný lokátor zdroje (URLs) popsanou pomocí internetového http a vlastní schémata http do např. ftp nebo news).



Architektura web crawleru ¹⁰⁶

Vysoká kvalita ukazuje na shromažďování obsahu webovských dokumentů a zvláště na problém aktuálnosti. Design web crawlerů je založený na dávkách (nebo periodicitě) prolézání nebo pokračování (nebo přibývání) prolézání. Dávka webovského prolézání (nebo snímkování webu) může být zamýšlena jako prolézání napříč webovským grafem, kde každý uzel je vidět nejvýše jednou. Pokračující webovské prolézání má za

¹⁰⁶ COTHNEY, Viv. Web-crawling reliability

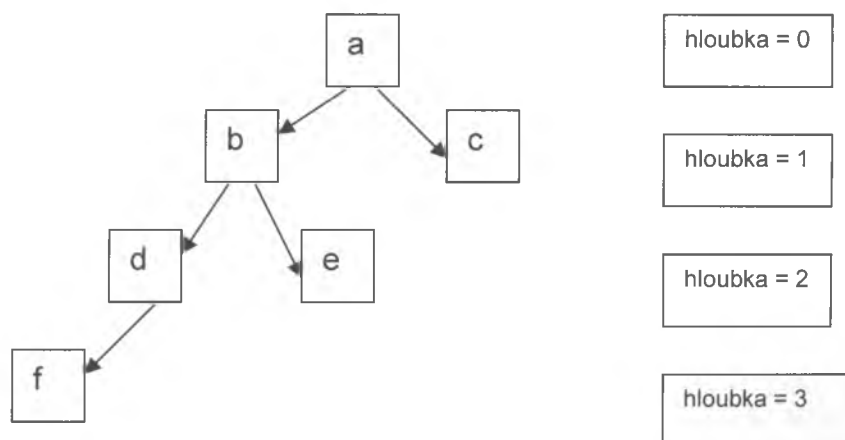
cíl znovu navštívit uzly z důvodu znovu shromáždit a reindexovat webovské dokumenty, které jsou již zastaralé.

Jednoduché prolézání znamená následující kroky:

1. crawler vezme zdrojový uzel z pracovního vytížení
2. jestliže uzel dosud nebyl navštíven, crawler přenesse webovský dokument odpovídající uzlu
3. jestliže webovský dokument je html, crawler extrahuje všechny citované uzly z dokumentu a přidá je do pracovního vytížení jako zdrojové uzly
4. crawler opakuje proceduru z bodu 1, předmět v jakémkoliv designu zúží nebo rozšíří, dokud zde nejsou nenavštívené zdrojové uzly zanechané v pracovním vytížení.

Existují tři způsoby výběru zdrojových uzlů vzaté z pracovního vytížení. Jsou nazývány *breadth-first*, *depth-first* a *best-first* (šíře první, hloubka první a nejlepší první). Hloubka citovaného uzlu je daná počet odkazů. To znamená, že citovaný uzel je odstraněn z některého relevantního zdrojového uzlu. Tento zdrojový uzel může být např. počáteční uzel nebo webovská stránka popsána pomocí URL ve formě: <http://some-host.org/>. Společný design prolézání zúžením je ve smyslu omezení hloubky crawlerem.

Další obrázek ilustruje malou hierarchii spojených dokumentů. Zde crawler začíná s dokumentem a, potom podle parametru „první šíře“ dává crawler přednost získání šesti uzlů v sekvenci a,b,c,d,e a f. V porovnání s parametrem „první hloubka“ by crawler usiloval o získání šesti uzlů v sekvenci a,b,d,f,e a poté c. Pod nátlakem parametru hloubky by se omezil na dvě úrovně a pak by dokument f nebyl zahrnut ani v postupu „první šíře“ ani „první hloubky“. Postup „první šíře“ je atraktivní prolézací politika, protože je jednoduchá na implementaci a v porovnání s „první hloubkou“ se pravděpodobněji vyvaruje přetížení individuálních serverů.¹⁰⁷



Nejlepší je postup, který umožní užití "PageRank" (pořadí stránek) sloužící k výběru dalších dokumentů, které budou vyvolány s úmyslem zredukovat strojovou a síťovou strojovou zátěž, jestliže se obsah prolézání bude snažit předpovědět právě z obsahu těchto webovských stránek, které jsou pravděpodobně vysoce kvalitní. Najork a Wiener nicméně zjistili, že s ohledem na kvalitu obsahu, sofistikovanější vykonávací politika „nejlepší první“ takové jako PageRank nenabízejí výhodu oproti pravidlu „širší první“.

¹⁰⁷ COTHNEY, Viv. Web-crawling reliability

Spekulovali, že to je proto, že "nejdůležitější stránky na ně mají odkazy z velkého množství „hostů“ a tyto odkazy budou brzy nalezeny bez ohledu na to, jakým hostem či stránkou crawler začal.

Existuje omezení celkového počtu uzlů platící během prolézání nebo omezení při prolézání. Potom pravidlo „širší první“, „hloubka první“ a „nejlepší první“ upřednostňující uzel budou poskytovat rozdílné výsledky. Jestliže omezení na počet pěti je uplatněno, pak díky pravidlu „širší první“ by crawler vyloučil dokument e v porovnání s vyloučením dokumentu c při prolézání s uplatněním pravidla „hloubka první“. Duplicitní webovské stránky se mohou objevit, jestliže soubor s obsahem dokumentu je explicitně zkopírován a uložen na více než jednom místě (a na více jak jednom hostitelském serveru) a poté jsou používány rozdílné URL jako citace odkazu se stejným obsahovým souborem.

Často se najde, že dvě citace ve formě <http://foo.com> a <http://foo.com/index.htm> odkazují na stejný soubor a mají identický obsah. Je to speciální třída duplikátů nebo podobných duplikátům webovských stránek, kdy mluvíme o tzv. zrcadlení hostitelských počítačů.

Politika prohledávání do šířky je oddělením nebo rozdělením grafu sféry crawleru do podgrafů a prozkoumání každého podgrafu nezávisle. Vhodná dělicí metoda je např. zahrnutí nebo vyloučení uzlů závislých na jméně hostitelské domény v URL. Proto všechny URL mající přesné jméno hostitelské domény rozdělíme. Prolézání postupuje rozdělením rozdělení, ačkoliv v rámci každého rozdělení webovského prolézání může být stále pravidlo „širší první“. Existují URL mimo rozdělení. V případě archivu, tyto vzájemné rozdělené odkazy jsou uloženy a používány k identifikování nových sídel nebo hostu pro prolézání.

Politiku webovského prolézání může také zvětšit spíše než zúžit nosič. Např. na obrázku *hierarchie webu* URL citovaného uzlu b odkazovaného ze zdrojového uzlu a může být ve formě <http://different-host.org/one/two/page.html>. Jednoduchý webovský crawler extrahuje toto URL a přidá ho do pracovního vytížení. Rozmnožení jednoduchého webovského crawlera může také přidat jedno nebo více ze tří URL <http://different-host.org/one/two>, <http://different-host.org/one> a <http://different-host.org>, což může být derivováno z citace. Politika webovského prolézání, která užívá všechny možné derivované parametry z uzlu je popsána zde jako zvětšení pomocí vzestupu cesty. Takto má cestou vzestupující webovský crawler příležitost objevit jinak nespojené uzly.

13.14. Obsahové prolézání versus prolézání odkazů

Web crawlery vytvořené pro vyhledávání webovských obsahů, což jsou *obsahové crawlery*, zahrnují nástroje pro vynechání duplikátů webovských dokumentů. Na rozdíl od *odkazových crawlerů*, které nevylučují stejné webovské dokumenty právě proto, že jejich obsah je duplikován (ačkoliv se jim mohou vyhnout z dalších důvodů). Pro search enginey z hlediska providerů není shromažďování duplicitních webovských stránek výhodou, protože se redukuje nadbytečné dokumenty při shromažďování a ukládání. Také verze grafu webovských odkazů, který je používán pro zlepšení procedur vyhledávání obsahů, pracuje lépe, když odkazy z duplicitních dokumentu jsou eliminovány.

Webovské crawlery vytvořené pro sbírání obsahů také mají za cíl uchovávat *aktuální* obsah. Některé žánry webovských dokumentů, zvláště ty, které zahrnují zprávy nebo programy vysílání, rychle zastarávají. Webovské stránky, které rapidně mění obsah

budou více náchylné k zastarávání. Aktualizace pak může probíhat tak rychle, jak search engine rychle zaznamenává změny jednotlivých webovských dokumentů.

Obsahové prolézání se objevuje jako pokračující činnost, zatímco periodické (ale rozšířené a obsažné dávky prolézání provedené organizací Internet Archive slouží za vzor). Neustálé prolézání webovských odkazů vede k potencionálnímu problému s údržbou zastaralých dokumentů. Cílem tvůrců search engineů je poskytnout relevantní čerstvé webovské stránky jako odpověď na uživatelův požadavek. Není ale jasné, co se stane se starými verzemi dokumentů, které mohou obsahovat zastaralé nebo duplicitní citace, když dokumenty v databázích search engineů nejsou obnoveny. Na jednu stranu tyto verze mohou být odstraněny, ale na druhou stranu mohou být pouze méně preferovány při určování relevance. V důsledku toho mohou celkové počty dokumentů a odkazů obsahovat násobky počtu obnovených dokumentů také kvůli zahrnutím odkazů na staré dokumenty.

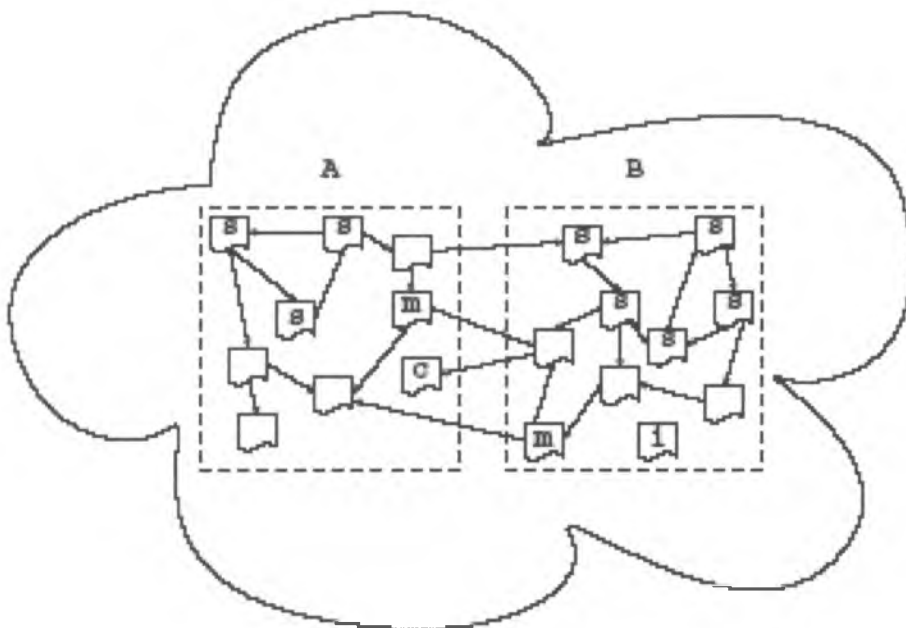
Search enginey pokrývají web a existuje malé překrytí mezi nimi. Všechny [search] enginey jsou směřovány k dobře propojeným, obsahově bohatým stránkám. Henzinger z toho odvodil, že různé obsahové crawlery generují rozličně směřované ukázky webu. Henzinger nahlíží, že obsahové crawlery různých search engineů dělají různá rozhodnutí, a tak získávají rozdílné množiny webovských dokumentů. Např. crawler může zkusit prolézat s vyšší preferencí "kvalitnější" webovské stránky. Algoritmy nebo politika prolézání webovských stránek, které jsou použity k definování vysoké kvality a k řízení ukázek výběru obsahových crawlerů, nejsou publikovány. Uživatelé provádějící druhotný výzkum nejsou tudíž informováni o základu výběru a směřování vzorků.

V principu tato výběrovost ovlivňuje hlavně infometrickou analýzu odkazů webu, protože verze grafu webu vyvinutá jako vedlejší produkt prolézání obsahu nebude pravděpodobně kopírovat skutečný web, nebo bude jeho podmnožinou. Namísto toho je specifickou interpretací grafu skutečného webu. Zda-li tento rozdíl mezi webovskými grafy je podstatný, bude záviset na přesnosti prováděného infometrického výzkumu.

13.15. Spolehlivost a platnost webovského prolézání

Některé specifické problémy spolehlivosti a platnosti prolézání webu jsou diskutovány z ohledem na graf zobrazený na obrázku *Příklad prostoru prolézání*. Ten ukazuje procházecký prostor složený ze dvou částí A a B, který může reprezentovat rozdílné hostitele nebo domény. Nesrovnalosti webového prolézání budou závislé na přijatých politikách prolézání.

V tomto příkladě předpokládáme, že 19 dokumentů zobrazených na obrázku *Příklad prostoru prolézání*, jsou všechny rozdílné kromě páru označeného n . Dokumenty označené m se navzájem zrcadlí nebo duplikují. Jednoduché neomezené obsahové prolézání dokumentů skrz celý prostor by, při nejlepším, prošel 17 těchto dokumentů. To proto, že izolovaný (nebo ostrovní) dokument i by nebyl objeven a pouze jedna (první objevená) duplicitní stránka by byla získána. Který z těchto dvou dokumentů by to byl, záleží na okolních vlivech jako je doba odezvy serveru, takže opakované identické obsahové prolézání by mohlo ukázat buď dva odkazy v oblasti A do B a jeden z B do A nebo jeden odkaz z A do B a dva z B do A. Tudíž analýza odkazů založená na tomto obsahovém procházení není spolehlivá. Výzkumníci používající identickou techniku dostanou různé výsledky. V tomto případě je také jasné, že data získaná na základě obsahového prolézání neplatně reprezentují interpretaci odkazové struktury procházeného prostoru, ačkoliv mohou platně reprezentovat jeho obsah.



Příklad prostoru prolézání ¹⁰⁸

Dokument *i* nemůže být objeven, protože není inlinkem. Nicméně to je závislé na struktuře cesty dokumentů serveru a zda nebo ne adresář stránek serveru je odhalen, cestou stoupající crawler může najít *i*, protože by byl zahrnut jako outlink v jeho původním adresáři. Objevení izolovaných dokumentů by mohlo být zvláště důležité pro analýzu odkazů, jestliže obsahují outlinky, dokonce jestliže byly jen k dalším dokumentům v rámci stejného oddílu.

Velké rozdíly v nálezů potencionálních uzlů naznačují, že průchody na skutečném webovském grafu mohou být senzitivní ke zvolené prolézací politice.

Obsahové procházení produkuje dvě rozdílné odpovědi, zatímco odkazové prolézání produkuje třetí odpověď. Procházení založené na oddílech by mohlo znamenat další ztrátu spolehlivosti v závislosti na interpretaci odkazové informace (informace, kterou nese odkaz v sobě). Dokument *c*, je podmíněně izolován (je citován pouze dokumentem v jiném oddílu B a nebude nalezen jednoduchým vyhledáváním oddílu A; dokument *c* bude nalezen pouze jednoduchým vyhledáváním, pokud plně využijeme interpretační outlink na něj z oddílu B). Tudíž procházení prostoru po oddílech může produkovat rozdílnou analýzu pro oblast A ve srovnání s procházením prostoru v celku (nikoliv po oddílech). To první může vynechat podmíněně izolovaný dokument *c*, zatímco to druhé nikoliv. Rozdíly v principu mezi dvěma třídami webového prolézače, obsahového procházeče a odkazového procházeče, mají praktický důsledek a výsledky s rozdílnými odpověďmi, pokud jsou aplikovány na problém analyzování odkazové struktury zmíněného prolézacího prostoru. Tedy široká technika webového prolézání není spolehlivá vzhledem ke zjištění odkazové struktury. Rozdíl, který je nalezen mezi dvěma na první pohled ekvivalentními zkoumánými odkazové struktury, může odrážet nedostatek spolehlivosti technik datového výběru a nikoliv skutečný rozdíl v datech.

¹⁰⁸ COTHNEY, Viv. Web-crawling reliability

13.15.1. Experiment

Cílem experimentu, který provedl Cothney, není porovnat obsahový crawler s odkazovým, ale demonstrovat rozdíl prolézacích politik nebo operačních charakteristik. Cílem experimentu je zjistit, zda jsou mezi politikami nějaké rozdíly nikoliv doporučit jednu z nich.

Experiment bere v úvahu sedm následujících politik procházení webu:

1. zvětšování stoupající pěšinou (augment by path ascent)
2. žádná omezení
3. omezení na standardní serverový port pro zdrojové uzly
4. omezení vynechat dotazovací URL pro zdrojové uzly
5. omezení na maximální hloubku 9
6. omezení na maximální hloubku 6
7. omezení na maximální hloubku 3

Politika "bez omezení" je v zásadě standardní crawler. Tento je srovnáván se třemi identickými crawlery, ale pro jednu změněnou charakteristiku. Tyto crawleři mají vliv na serverový port (který byl důležitý protože experimentální podgraf leží uvnitř ochranného firewallu), zda-li URL má dotazovací část a na omezení hloubky. Tyto dvě charakteristiky jsou zvoleny, protože se vyskytují jako ukázky omezení použitých v praxi. Hodnoty vybrané pro hloubku omezení jsou zvoleny, aby poskytovaly dobrý rozsah hodnot. Vzrůstající crawler, který používá stoupající pěšinu (patch ascent), aby podpořil objevování uzlů, je nový přístup k procházení ve srovnání s běžnými popisy, jak webové prolézače pracují. Pokouší se načíst adresářovou strukturu serveru, aby rychleji objevil uzly, které jinak nemají žádné inlinky.

Obecně se má za to, že čím větší množství počátečních dokumentů, tím lépe. Na druhou stranu, dodatečné počáteční dokumenty, které jsou mezi sebou propojeny, nemají žádný další význam.

13.15.2. Metoda

Numerický rozdíl mezi frekvencí zdrojových uzlů a citovaných uzlů v subdoméně odráží proceduru jednoduchého procházení, ve které zdrojové uzly musí být v HTML formátu. Ačkoliv obrázkové soubory vytvářejí většinu rozdílů, jsou také jiné než html webové formáty, např. Microsoft word, které mohou být zdrojem outlinků.

Zdrojové uzly	Frekvence zdrojových uzlů	V rámci wlv.ac.uk	Kdekoliv jinde	Celkově
		Frekvence citovaného uzlu		
Všechny	123 375	302 680	36 277	338 957
Nemající dotaz URL	122 479	301 353	36 253	337 606
Na daném portu	121 466	299 946	35 894	335 840

Charakteristika odkazů subgrafu

Výsledky

Výsledky ukazují, že použité politiky jsou si blízké. Tabulka *charakteristika odkazů subgrafu* udává, median a standardní odchylku zdrojových a citovaných objevených uzlů a poměr citovaných uzlů ku zdrojovým uzlům při tisíci zdrojových uzlech (seed set). Výzkum zjišťuje, že dokumenty topologicky blízké domácí stránce, jsou bohatší na outlinky. Toto je součástí vyhledávání do šířky (pravidlo šíře první) při procházení odkazů.

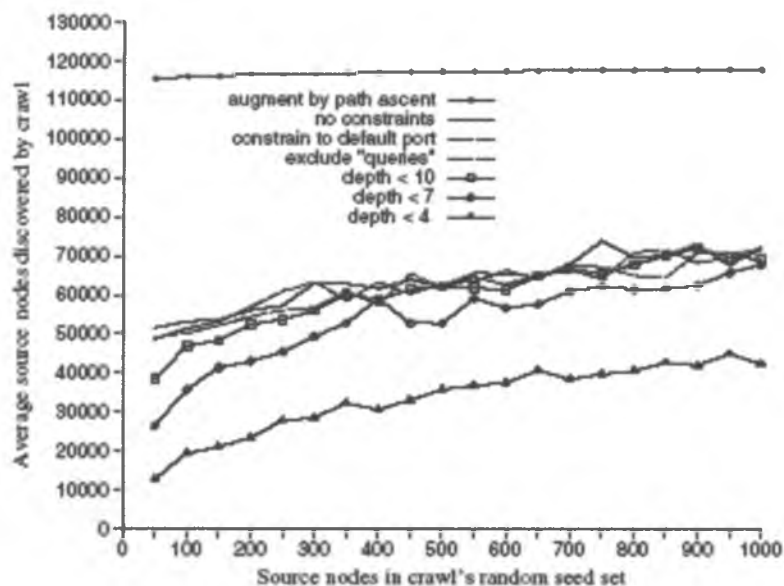
V případě absence informací o prolézací politice, která není odhalena, dva výzkumníci užívají data nashromážděná z různých politik, která by chybějící data doplnila, např. rozdíl v citování uzlů odkazující na zdrojový uzel mohou být objeveny.

Výsledek jednoduchého webovského prolézání je obecně sensitivní vzhledem ke zdrojovým uzlům. Není to případ pro rozšířené prolézání.

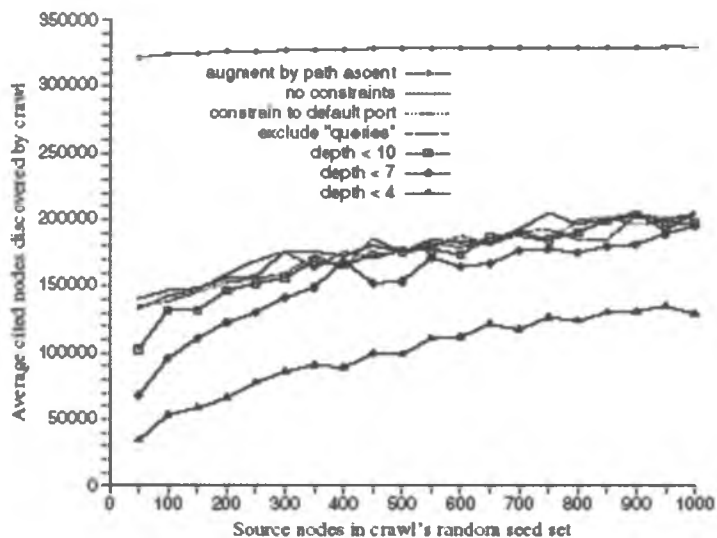
Konečná podobnost ve výsledku pěti ze sedmi zkoumaných prolézacích politik ukazuje, že velký počet zdrojových uzlů nebo zdrojový uzel, který je dobrou reprezentací zdrojů ve sféře crawlerů, eliminuje pro některé z crawlerů omezující podmínky, které je nutné vzít v úvahu. Ukazuje se, že hodnota kritické hloubky v rámci níž výsledky jednoduchého prolézání jsou poměrně limitovány. Pro wlv.ac.uk subdoménu tato hodnota je méně než 6.

Zdrojové uzly zdrojových uzlů při webovském prolézání mohou být pečlivě vybírány. Možný mechanismus k vysvětlení zjištění Najorka a Wienera a nedostatek překrytí mezi rozdílnými search enginey je, že webovské prolézání bylo v rámci kritické hloubky, kde výsledek prolézání je sensitivní k výběru zdrojových uzlů, zvláště když je malý. Najork a Wiener samy navrhují, že většina vysoce kvalitních webovských dokumentů (to je cíl obsahového prolézání) jsou v malé hloubce.

Výsledek funkcí rozšířené webovské prolézací politiky se liší s ohledem na citlivost k velikosti zdrojových uzlů a hodnot. Důležitý efekt rozšíření nedostatku cesty stoupání je, že zdrojové uzly jsou implicitně rozšiřovány vzhledem k zahrnutí domovské stránky každého reprezentovaného webovského serveru. To je maximalizováno příležitostí webovského crawleru využít jakoukoliv přirozenou hierarchii v URL struktuře webovských dokumentů. Každá větší sada zdrojových uzlů efektivně vytváří stejnou webovskou stránku směrem k prolézání celé sféry prolézání. Toto explozivní prolézání zahrnuje objevování izolovaných dokumentů z informací adresáře serveru, kdykoliv to je možné. Počet takovýchto dokumentů je nafouknutý, protože mnoho stránek adresáře nebudou mít žádné odkazy na ně, ale budou znamenat nárůst procedury vyhledávání (bude nutné je prozkoumat).



Zdrojový uzel objevený prolézací politikou ¹⁰⁹



Citovaný uzel objevený prolézací politikou ¹¹⁰

Politika prolézání	Zdrojové uzly (n= 1 000)	Citované uzly (n = 1000)	Citované uzly na zdrojový uzel
Rozšířené na cestu stoupání	117 770 (se = 51)	329 561 (se = 110)	2.80 (se = 0.0006)
Žádné omezení	72 390 (se = 1 879)	205 974 (se = 4 260)	2.85 (se = 0.02)
Jen daný port zdroje	71 564 (se = 2 320)	202 325 (se = 5 143)	2.84 (se = 0.03)
Vyjma „dotazu“	71 481 (se = 2 320)	203 837 (se = 5 143)	2.86 (se = 0.02)

¹⁰⁹ COTHNEY, Viv. Web-crawling reliability

¹¹⁰ COTHNEY, Viv. Web-crawling reliability

zdroje	050)	4 837)	
Hloubka < 10	69 346 (se = 2 139)	196 770 (se = 5 163)	2.85 (se = 0.02)
Hloubka < 7	67 786 (se = 2 651)	195 377 (se = 6 209)	2.90 (se = 0.03)
Hloubka < 4	42 391 (se = 663)	129 196 (se = 3 024)	3.04 (se = 0.04)

Tabulka: efektivita politiky prolézání ¹¹¹

Druhý příspěvek k rozdílnosti je, že dokumenty v subdoméně citované jen jako <http://host.wlv.ac.uk/path/index.html> jsou zahrnuty dodatečně v subgrafu odkazů jako <http://host.wlv.ac.uk/path>, což je izolované a proto tedy neobjevitelné jednoduchým crawlerem.

Rozdílné třídy webovských crawlerů, obsahové crawlery a odkazové crawlery jsou rozporuplné a obsahové crawlery nespolehlivě poskytují odkazovou strukturu dat. Webovské prozkoumávání crawlery jsou nekonzistentní v jejich výsledcích. Jsou proto nespolehlivé. Zejména, crawlery, které jsou identické kromě rozdílnosti v jejich politice hloubky, produkují rozdílné výsledky. Proto, s ohledem na záměr experimentu, to ukazuje, že politika crawlerů má význam, změnění politiky může změnit výsledky. Tudíž interpretace nebo analýzy porovnávání dat z prolézání jsou založeny na tom, že člověk musí znát politiku prolézání s níž byla data vybrána. To znamená, že webovské prolézání jako popis technik výběru dat je neadekvátní v tom, že zahrnují příliš mnoho skrytých variabilit (třídy, politika, sady, atd.) a při absenci dodatečné kvalifikace to vede k nekonzistenci výsledků. Takto webovské prolézání, samo o sobě, není spolehlivé.

Co se týká subdomén a dané vhodné zdrojové uzly, jakákoliv politika prolézání z těch zkoumaných je totožná a produkuje konzistentní výsledky. Spolehlivé webovské prolézání je možné a je dosažitelné odhalením nebo informováním skrytých možností.

Jakýkoliv graf, který obsahuje podgraf bude také produkovat nekonzistentní výsledky, dokonce jestliže to je omezeno na podgraf, potom nedostatek spolehlivosti přetrvává.

Politiky prolézání, včetně těch hlavních search enginů nejsou sdělovány veřejnosti. Uživatelé dat crawlerů, takový jako infometričtí vědci studující web, jsou zvláště s tímto nespokojeni, protože nejsou informováni o tom, jak data v jejich sadách dat mají být vybírány z jednoho zdroje pro porovnání s dalšími zdroji.

Proto je také nejasné, zda nebo ne sklon a selektivita u vzorků hlavních search enginů skutečně modifikuje infometrická zjištění. Použití dat z prolézání při absenci většího informování představuje také možnost, že separátně prováděné studie neúmyslně budou založeny na stejných datech jak navrhuje Bailey a další.

Etické webovské prolézání požaduje také, že soubory, které nejsou upravovány by neměly být vyvolány. To zredukuje nadbytečné sítě a přetížení serverů. ¹¹²

¹¹¹ COTHNEY, Viv. Web-crawling reliability

¹¹² COTHNEY, Viv. Web-crawling reliability

14. Další citační rejstříky

14.1. Taiwan Humanities Citation Index (THCI)

Knihovna Chinese Academy of Sciences vytváří Chinese Science Citation Database (CSCD) od roku 1989. Nanjing University v Číně a Hong Kong University of Science and Technology vytvářejí Chinese Social Science Citation Index (CSSCI). Tyto databáze citačních indexů mají velký vliv na výzkum. Někteří výzkumníci aplikují citační data do hodnocení příspěvku vědců, časopisů nebo institucí. Některá aplikovaná citační data slouží k analyzování struktur specifických oborů výzkumu, jiná jsou využívána retrospektivně k hodnocení výzkumných trendů a vytvářejí náznak budoucích trendů. Většina časopisů v citačních indexech ISI jsou publikovány v angličtině. V Tajvanu je potřeba lokální databáze citací pro lokální požadavky. CSCD a CSSCI poskytují citační data pro výzkumníky v Číně. Jestliže ale chceme mít souhrnný obrázek o tajvanském výzkumu v umění a humanitních vědách, je potřeba vytvořit Taiwan Humanities Citation Index (THCI). Cílem projektu THCI je vytvoření citačního indexu pro humanitní časopisy publikované na Tajvanu. A vědci tak budou moci využít THCI pro výzkum v oblasti humanitních věd, k analýze změn modelů humanitních studií a předpovídání budoucích trendů v humanitních oborech.

14.1.1. Cíle projektu THCI:

- Vytvoření citačního indexu pro humanitní obory
- Poskytnutí nezbytných vyhledávacích nástrojů (vyhledávání podle názvu časopisu, autora, dle citovaných autorů, zdrojových článků a titulu citovaných článků)
- Umožnit citační analýzy (THCI může identifikovat frekvenci citací pro časopisy a autory: výzkumníci by mohli užít tato data k vypočítání impakt faktoru a immediacy indexu- faktor bezprostředního vlivu?)

Chinese Academy of Science v Číně vytvořila Chinese Science Citation Database (CSCD) v roce 1989. Indexuje 315 vědeckých a technologických časopisů publikovaných v Číně jako suplement SCI. První práce o CSCD byla publikována v roce 1995 a první CSCI CD ROM vyšel v roce 1996. Nanjing University se zaměřila na sociální vědu a umění a humanitní časopisy a vytváří Chinese Social Science Citation Database od roku 1997. V roce 1999 Nanjing University v Číně podepsala kontrakt o spolupráci na CSSCI s Hong Kong Science and Technology University. První CSSCI CD ROM byl publikován v roce 2000.

Několik projektů citačního indexu probíhá také na Tajvanu. V roce 1997 National Science Council (NSC) republiky Číny podpořil pilotní projekt na vytvoření Taiwan Science Citation Index, který indexoval několik časopisů z vědy a techniky publikovaných na Tajvanu v letech 1996 a 1997. V roce 1999 NSC podpořil dva projekty výzkumných center: Social Science Research Centre (SSRC) a Centre for Humanities Research (HRC). Hlavními úkoly obou center je vytvářet Taiwan Social Science Citation Index (TSSCI) a Taiwan Humanities Citation Index (THCI).

THCI je samostatný citační rejstřík, který je vytvářen lokálně na Tajvanu a neobsahuje žádná data z dalších citačních databází.

14.1.2. Vytvoření THCI

Hlavním cílem THCI je vytvářet citační index časopisů z oblasti umění a humanitních věd na Tajvanu. Úkolem je i tvorba vyhledávacího nástroje a nástroje pro hodnocení. Citační databáze může být zdrojem informací o charakteristikách, vývoji a výzkumných trendech na Tajvanu v oblasti umění a humanitních oborů. Užití THCI jako výzkumného nástroje je jádrem filozofie projektu.

14.1.3. Kritéria pro indexaci v THCI:

- předmět indexovaných časopisů musí být umění a humanitní obory
- indexované časopisy by měly být publikovány na Tajvanu

Byl vytvořen nejprve seznam časopisů z oblasti umění a humanitních věd. Poté byl seznam posouzen komisí složenou z odborníků z různých disciplín. Konečný seznam obsahoval 245 titulů s třemi rozdílnými prioritami. Snahou je indexovat co nejvíce časopisů, ale přes to byly některé tituly vyřazeny.

Ačkoliv THCI indexuje mnoho časopisů, ne každý článek z časopisu je indexován. Vybírány jsou výzkumné práce a akademické články. Data v databázi obsahují základní bibliografická data pro zdrojové práce, citované práce a vztahy mezi citacemi. Bylo definováno množství atributů pro všechna možná užití THCI.

Discipline	Number of titles
Arts	16
Literature	82
History	48
Language	46
Library Science	23
Philosophy	17
Religion	13
Total	245

Statistika indexovaných časopisů dle disciplín ¹¹³

Rozlišuje se 6 typu citací užívaných v čínských člancích z umění a humanitních oborů. Některé články mohou užívat více než jeden typ citací simultánně, definuje se priorita pro každý citační typ. THCI uchovává záznamy jen pro citované práce s vyšší prioritou než citační typ užívaný v každé zdrojové práci.

Citation type	Priority
Citation	1
Notes	2
Footnotes	3
References (bibliography)	4
Notes in articles	5
Quotation	6

Typy citací ¹¹⁴

¹¹³ CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index

THCI identifikuje rozdílné typy citovaných prací. THCI indexuje jen články z časopisů, konferenční zprávy, technické zprávy, teze a disertace. V umění a humanitních vědách je zaznamenáno užití jazyků jiných než angličtina, proto i jazyky citovaných prací jsou také zaznamenány. Jedná se o angličtinu, čínštinu, japonštinu, francouzštinu, němčinu, italštinu, španělštinu, ruštinu, korejštinu a další.

V rámci THCI je identifikováno 5 hlavních disciplín a 8 poddisciplín z oblasti humanitních věd.

Main discipline	Sub-discipline
Humanities	Arts
	Chinese Literature
	Foreign Literature
	History
	Language
	Library Science
	Philosophy
	Religion
Social Sciences	
Natural Sciences	
Biological Sciences	
Engineering & Technology	

Disciplíny v THCI ¹¹⁵

Pokud autor (autoři) změní afiliaci, rejstřík se snaží je vystopovat. Proto jsou vytvořeny dva vztahy k osobám.

První vztah je pro autora (autory) zdroj nebo citovaná práce, druhý pro výzkumníky samotné. Výzkumníci mohou publikovat různé články pod hlavičkou různých institucí nebo organizací. Tabulka „researcher“ uchovává aktuální informace. Stopují se historická data publikací pro vědce, spíše než jejich přemístění s instituce do instituce.

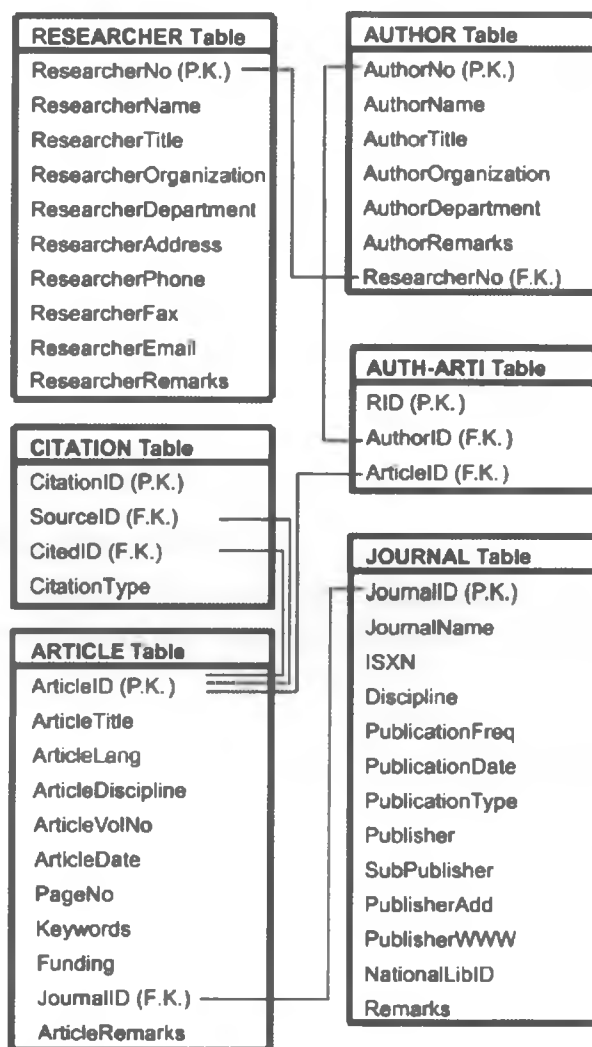
RESEARCHER Table	AUTHOR Table
ResearcherNo (P.K.)	AuthorNo (P.K.)
ResearcherName	AuthorName
ResearcherTitle	AuthorTitle
ResearcherOrganization	AuthorOrganization
ResearcherDepartment	AuthorDepartment
ResearcherAddress	AuthorRemarks
ResearcherPhone	ResearcherNo (F.K.)
ResearcherFax	
ResearcherEmail	
ResearcherRemarks	

Výzkumníci a autoři ¹¹⁶

¹¹⁴ CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index

¹¹⁵ CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index

¹¹⁶ CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index



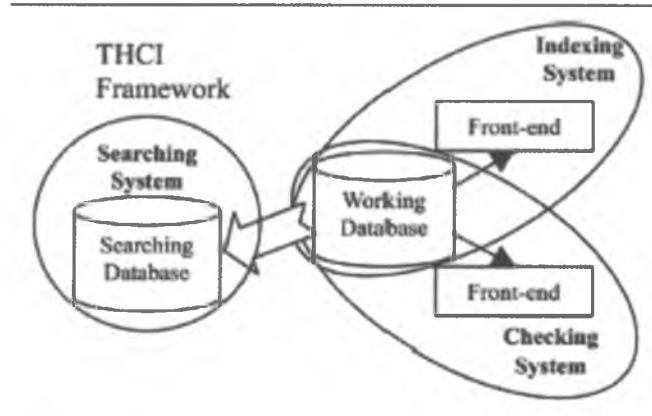
Tabulka vztahů v THCI ¹¹⁷

14.1.4. THCI struktura

THCI je založeno na třech komponentech. První je indexační systém, druhý kontrolní systém a třetí je vyhledávací systém. Indexační systém se skládá z databáze a klientského systému. Kontrolní systém také obsahuje databázi a klienta. Vyhledávací systém má strukturu složenou z webovského browseru, webovského serveru a systému vyhledávací databáze. Indexační systém a kontrolní systém užívají stejnou databázi. Vyhledávací databáze je periodicky importována z pracovní databáze.

¹¹⁷ CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index

14.1.5. Struktura THCI ¹¹⁸



14.1.6. Politika a pravidla THCI

Protože formát pro citace v humanitních časopisech na Tajvanu je velmi nekonzistentní, kompletní pravidla nebyla dopředu vytvořena. Nicméně některé pravidla byla definována:

Politika

- indexované články jsou výzkumně orientované a z oblasti humanitních věd
- je přijat princip „širšího indexování“
- THCI je více výzkumně založená databáze než hodnotící databáze
- Zdrojové časopisy jsou publikované na Tajvanu
- Citované práce jsou omezeny na články z časopisů, knih, sbírek, výzkumné zprávy, teze a disertace a další ¹¹⁹

14.1.7. Aplikace THCI

- do dubna 2004 bylo indexováno 3 550 časopisů, uloženo 35 097 zdrojových článků a 499 932 citovaných titulů. Kromě toho, že databáze je vyhledávacím nástrojem pro bibliografii z humanitních věd, THCI je zamýšlen jako výzkumný nástroj, který slouží k zachycení výzkumných trendů, zobrazení citačních map, pro nalezení důležitých objektů (institucí, časopisů, výzkumníků a článků) a pro identifikování výzkumných podoborů. Hlavním cílem projektu THCI je vytvořit výzkumný nástroj, ačkoliv je možné na rejstřík pohlížet také jako na hodnotící nástroj.

Aktuální verze je dostupná na:

www.hrc.ntu.edu.tw/thci/thci.htm (bohužel hlavní stránka se pro nás zobrazuje špatně, pokud nemáme nainstalovanou čínštinu).

<http://www.hrc.ntu.edu.tw/en/index2.htm> (anglická verze)

¹¹⁸ CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index

¹¹⁹ Další pravidla viz článek CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index

Na hlavní stránce je možné bibliografické vyhledávání, pokud uživatel chce užít další vlastnosti, může kliknout na „Citation Analysis Retrieval System“ a „Related Articles on Citation Analysis“. Je možné prolístovat časopisy dle disciplíny.

Pokud uživatel chce vyhledávat konkrétního autora, vloží jméno do vyhledávacího pole a zmáčkne tlačítko pro vyhledávání.

Kliknutím na jeden článek se objeví citované práce pro tento článek. Kliknutím na jednu z citovaných prací se umožní zobrazení seznamu článků, které citují citovanou práci.

14.1.8. Vyhledávací nástroje

THCI online služby neposkytují výzkumné vlastnosti přímo. Výzkumníci musí identifikovat rozsah jejich výzkumu. Výzkumníci poté potřebují vyhledat požadovaná data, stáhnout a upravit tato data pro další analýzu. Např. kocitační matice může být vytvořena v širší, kterou si lze domluvit.

國科會人文學研究中心
Center For Humanities ResearchHome.....How to contact usSearch.....Chinese version

News Introduction THCI Forum Resources

THCI

The Taiwan Humanities Citation Index (THCI) lists the citations of articles in Taiwan's humanities journals, which could further give us an understanding of the number of citations of papers published in these journals. The analysis could shed light on the development of a certain research topic and further infer about the future trend of research in that area of study. Moreover, the analysis of the literature cited in journals and the mutual citation between articles could help in the assessment of the degree of influence of a particular journal in the field of humanities. This information could then be provided to researchers for their reference.

Basic Information Retrieval System	Citation Analysis Retrieval System	Related Articles on Citation Analysis
-provides the basic retrieval function of the citations database, including the retrieval of key words of a title, retrieval of the name of an author, retrieval of a journal article, retrieval of an article using another article, and the retrieval of the cited article	-allows the search of the number of times that an article is cited, the number of times that an author is cited, the coefficient of influence, immediate citation index, self-citation index, and the rate of self-citation	-provides related article or citation analysis published in Taiwan and abroad, including Mainland China

Hlavní stránka THCI ¹²⁰

¹²⁰ CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index



Basic Information Retrieval		Listed Journals in database			
Choices	<input type="radio"/> Name of Journal	Category of journals	number	Category of journals	number
	<input checked="" type="radio"/> Name of Author				
	<input type="radio"/> Name of Source Article	<input type="checkbox"/> all journals	245	<input type="checkbox"/> philosophy	17
	<input type="radio"/> Name of Cited Article	<input type="checkbox"/> General subjects	82	<input type="checkbox"/> history	48
Key words	<input type="text" value="陳光"/>	<input type="checkbox"/> library information	23	<input type="checkbox"/> arts	16
		<input type="checkbox"/> religion	13	<input type="checkbox"/> languages	46
<input type="button" value="Search"/> <input type="button" value="Reset"/>		<input type="button" value="Search"/> <input type="button" value="Reset"/>			

© 1999-2004 Center For Humanities Research 國科會人文學研究中心

Vyhledání konkrétních článků autora ¹²¹

Basic Information Retrieval	
Choices	<input checked="" type="radio"/> Name of Journal
	<input type="radio"/> Name of Author
	<input type="radio"/> Name of Source Article
	<input type="radio"/> Name of Cited Article
Key words	<input type="text" value=""/>
	You must enter data to search by...
<input type="button" value="Search"/> <input type="button" value="Reset"/>	

Listed Journals in database			
Category of journals	number	Category of journals	number
<input type="checkbox"/> all journals	251	<input type="checkbox"/> philosophy	
<input type="checkbox"/> General subjects	84	<input type="checkbox"/> history	
<input type="checkbox"/> library information	23	<input type="checkbox"/> arts	
<input type="checkbox"/> religion	13	<input type="checkbox"/> languages	
<input type="button" value="Search"/> <input type="button" value="Reset"/>			

© 1999-2004 Center For Humanities Research

(vzhled z července 2005)

¹²¹ CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index

Nástroj hodnocení

Před užitím citačního indexu jako hodnotícího nástroje musíme porozumět motivaci tvorby citací. Užití citačního indexu jako hodnotícího nástroje je stále sporné, a to i tehdy, pokud všechny citace jsou správné. Stále existuje problém např. užití nepatřičného měření pro hodnocení. Tvůrci chtějí vytvořit metrický „impakt faktor s daty“. ISI užívá pro výpočet tohoto parametru dva roky. Pro disciplíny orientované na inženýrské obory jsou dva roky vhodné, ale příliš krátké pro další disciplíny. Výzkumníci z Leiden University v Nizozemí zjistili, že dokonce 6 let je krátká doba pro disciplíny jako biologie a medicína. Pro obory jako umění a humanitní věda bude potřeba asi ještě delší období. Aplikace impakt faktoru s jasným popisem časového rámce ujasní základ pro vypočítání a umožní nám porozumět potencionálním problémům.



國科會人文學研究中心
Center For Humanities Research

.....Home.....How to contact us.....Search.....Chinese version

News Introduction THCI Forum Resources

Match Author: 24 Records		Sort by: Author Name	
No	Author Name	Institute Name	Department Name
11	陳光化		
12	陳光國		
13	陳光慶		
14	陳光色		
15	陳光華=Chen, Kuang-Hua	國立臺灣大學	圖書資訊學系
16	陳光勳		
17	陳光榮	私立元智大學	企管系
18	陳光學	私立義興大學	企業管理學研究所
19	陳光堯	私立清雲技術學院	國貿系
20	陳光磊		

Priority | Next | First | Last | Requery | Correct | Print

Vyhledávání autora ¹²²

¹²² CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index



作者資料						
作者姓名	陳光華=Chen, Kuang-Hua					
服務機構	國立臺灣大學		服務單位	國語資訊學系		
<input type="checkbox"/> 本資料庫收錄之作者文獻 (11筆) <input type="checkbox"/> 國語人文學系收錄之文獻 (11筆) <input type="checkbox"/> 國語人文學系收錄之文獻 (11筆)						
本資料庫收錄之作者文獻						
請選擇排序的欄位： <input checked="" type="radio"/> 文獻篇名 <input type="radio"/> 題名 <input type="radio"/> 編號						
編號	文獻篇名	書刊	文獻卷期	出版日期		
1	「電子圖書館」資訊課程之設計	圖書館學刊=National Taiwan University Journal of Library Science	12	1997-12		
2	1999年全島圖書館週報——多學科發展	圖書館學與資訊科學=Journal of Library and Information Science	28(1)	2000-04		
3	TREC環境及其對資訊科學研究之影響	圖書館學季刊=Bulletin of Library and Information Science	29	1999-06		
4	XML Metadata 分類與單一Metadata之設計=The Design of XML Metadata Management System	中國圖書館學會會刊=Bulletin of the Library Association of China	65	2008-12		
5	中文資訊檢索系統之設計與製作	資訊傳播與圖書館學=Journal of Information, Communication and Library Science	8(3)	2000-03		
6	中文學術研究發展現況	書刊	18118	1998-06		
7	文件內容之分析——語料庫為本的模型=Content Analysis: A Corpus Based Model	圖書館學刊=National Taiwan University Journal of Library Science	11	1998-12		
8	日本圖書館學教育的發展與現況=The Development and Current Status of Library	國立臺灣大學圖書館學季刊=National Taiwan University Journal of	16	2001-06		

Články autora indexované v THCI ¹²³



來源文獻						
文獻篇名	適應向量檢索打斷語言障礙					
文獻作者	陳光華=Chen, Kuang-Hua					
期刊書刊	大學圖書館=University Library Quarterly					
文獻卷期	30(1)	起始頁碼	87-99			
出版日期	1998-01	文獻語言	中文			
學門類別	國語資訊學	學位類別				
研究主題						
附註						
本資料庫下列文獻：16 筆						
請選擇排序的欄位： <input checked="" type="radio"/> 文獻篇名 <input type="radio"/> 題名 <input type="radio"/> 編號						
編號	文獻篇名	書刊	文獻卷期	出版日期	作者姓名	
1		圖書館學季刊=National Taiwan University Journal of Library Science		1998	陳光華=Chen, Kuang-Hua	
2		Adaptive Vector Space Text Filtering for Monolingual and Cross-Language Applications		1996	Oard, Douglas	
3	Across Language, Across Cultures	D-Lib Magazine		1997-05	Peters, Carol	
4	An Approach to Conceptual Text Retrieval Using the Euro-WordNet Multilingual Semantic Database	Proceedings of AAAI-97 Spring Symposium Series on Cross-Language Text and Speech Retrieval		1997	Silvertanz, Julio, Dorigo, Julio, Yardeni, Felipe	

Citované práce daného článku ¹²⁴

¹²³ CHEN, Kuang-hua. The construction of the Tajwan Humanities Citation Index

¹²⁴ CHEN, Kuang-hua. The construction of the Tajwan Humanities Citation Index



被引文獻					
文獻題名	An Approach to Conceptual Text Retrieval Using the Euro-WorldNet Multilingual Semantic Database				
文獻作者	Gilarranz, Julio, Gonzalez, Julio, Verdejo, Felisa				
來源書刊	Proceedings of AAAI-97 Spring Symposium Series on Cross-Language Text and Speech Retrieval				
文獻卷期	第04卷第01期				
出版日期	1997				
學科類別	圖書資訊學				
授權來源	學位論文				
附註					
本頁被下列文獻引用：1筆					
將選擇排序的欄位： <input checked="" type="radio"/> 文獻題名 <input type="radio"/> 摘要 <input type="radio"/> 摘要					
編號	文獻題名	書刊	文獻卷期	出版日期	作者姓名
1	語法資訊學與語言學	大學圖書館=University Library Quarterly	2(1)	1998-01	蔡光華=Chen, Kuang-hua
上一頁 下一頁 第一頁 最後頁 重新查詢 檢視完整 列印本頁					

Články citující konkrétní citovanou práci ¹²⁵

14.1.9. Těžkosti a problémy

Při vytváření THCI se narazilo na některé těžkosti a problémy. Prvním je měnící se formáty citací. Rozdílné disciplíny mají rozdílné formáty. Různé časopisy užívají rozdílné formáty a dokonce různé články užívají ve stejném časopise různé formáty. Druhým problémem je nedostatek školeného personálu, např. problém s jazykem článků. Třetím problémem je nesprávnost nebo nekompletnost citací. Ty musí být kontrolovány a doplněny s využitím různých online bibliografických služeb. Čtvrtým problémem je identifikování disciplíny zdrojového článku a citované práce. Některé články pokrývají více než jednu disciplínu a zaměstnanci mají problémy identifikovat hlavní disciplínu. Dále bibliografická data o citovaných pracích často nabízí jen stopu k disciplíně článku a učinit správné rozhodnutí je těžké.

14.1.10. THCI projekt

Je rozvíjen od ledna 2000. Ačkoliv se vyskytly různé problémy a těžkosti, podařilo se splnit několik úkolů:

- vytvoření THCI pracujícího prostředí
- nastavení THCI konstrukčních procedur
- sepsání THCI pravidel a politiky
- implementace THCI systému struktury
- naplnění různých vyhledávacích charakteristik.

Nejvíce indexované časopisy A&HCI (Arts and Humanities Citation Index) jsou v angličtině. THCI by mohl být doplňkový citační rejstřík pro tajwanské výzkumníky s možností získat celkový přehled o tajwanském výzkumu v oblasti umění a humanitních věd. ¹²⁶

¹²⁵ CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index

¹²⁶ CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index

15. SCOPUS

SCOPUSinfo

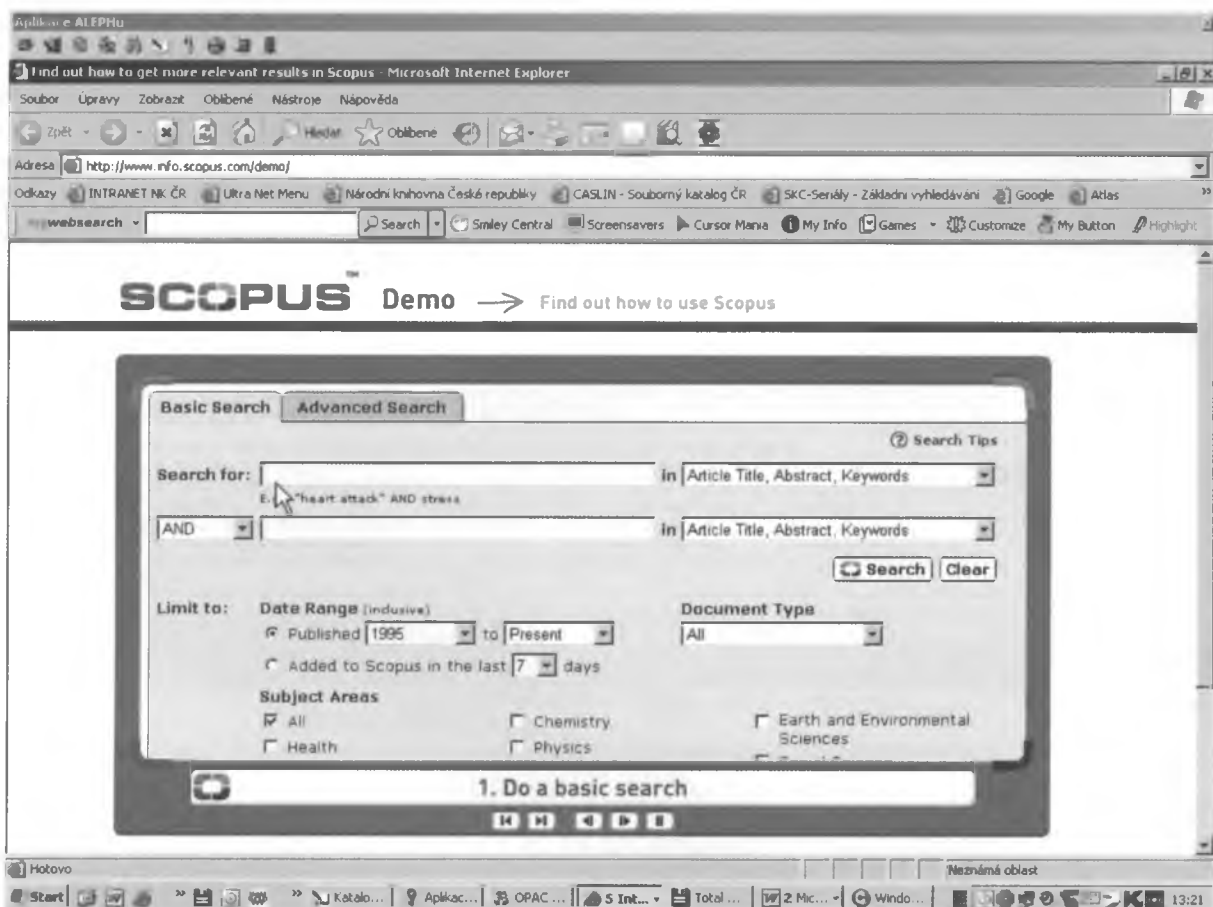
Scopus (dostupný z: <http://www.info.scopus.com> nebo www.scopus.com) je dle tvůrců největší databáze abstraktů a je vytvářena jako rejstřík. Scopus je multidisciplinární navigační nástroj, které obsahuje záznamy od poloviny 60. let 20. století a nabízející nově propojení citací napříč vědeckými abstrakty z jednoho místa.

Databáze je vytvářena a rozvíjena na základě spolupráce s výzkumníky a knihovníky. Scopus nabízí možnost snadno získat expertní výsledky. Nabízí lepší podporu v procesu vyhledávání literatury, integrované webovské vyhledávání plus jednoduché proniknutí pomocí hypertextových odkazů na plné texty článků a dalších knihovních zdrojů.

Scopus nabízí:

- Lepší pokrytí vědecké, technické, lékařské literatury a literatury z oblasti sociálních věd (14 000 titulů) než jakákoliv další databáze
- Odkazy na plné texty přímo z výsledků vyhledávání
- Abstrakty zpětně až k roku 1966, plně propojitelné odkazy od roku 2005 a denní aktualizace
- Inovativní nástroje pomáhající zlepšit výsledky vyhledávání
- Pravidelné upozorňování na články o tématu, které nás zajímá pomocí služby Alert .

Scopus umožňuje snadnější a efektivnější vyhledávání.



Pokročilé vyhledávání ¹²⁷

¹²⁷ Scopus: www.scopus.com

Find out how to get more relevant results in Scopus - Microsoft Internet Explorer

Adresa: <http://www.info.scopus.com/demo/>

SCOPUS™ Demo

Find out how to use Scopus

Results from: Scopus: 1,136 Web: 18,941 Scopus + Web: 20,077

Your query: TITLE-ABS-KEY(antimicrobial chemotherapy) AND PUBYEAR AFT 1994

Refine Results

Source Title	Author Name	Year	Document Type	Subject Area
<input type="checkbox"/> International Journal of Antimicrobial Agents (246)	<input type="checkbox"/> Jones, R.H. (25)	<input type="checkbox"/> 2004 (74)	<input type="checkbox"/> Article (893)	<input type="checkbox"/> Health (878)
<input type="checkbox"/> Journal of Antimicrobial Chemotherapy (87)	<input type="checkbox"/> Park, P. (8)	<input type="checkbox"/> 2003 (168)	<input type="checkbox"/> Review (181)	<input type="checkbox"/> Life Sciences (897)
<input type="checkbox"/> Antimicrobial Agents and Chemotherapy (72)	<input type="checkbox"/> Walker, M.A. (8)	<input type="checkbox"/> 2002 (194)	<input type="checkbox"/> Conference Paper (130)	<input type="checkbox"/> Agricultural and Biological Sciences (824)

Results: 1,136

Search within results

Document (Sort by relevance)

Date	Author(s)	Source Title	Cited By
1. 2004	Bronner, S., Montel, H., Prevost, G.	ES-MS Microbiology Reviews 28 (2) : 183-200	0
2. 2004	Yang, D., Yum, J.H., Lee, ...	Antimicrobial Agents and ...	0

2. Refine your results

Výsledky vyhledávání ¹²⁸

¹²⁸ Scopus: www.scopus.com

Find out how to get more relevant results in Scopus - Microsoft Internet Explorer

Adresa: <http://www.info.scopus.com/demo/>

SCOPUS™ Demo

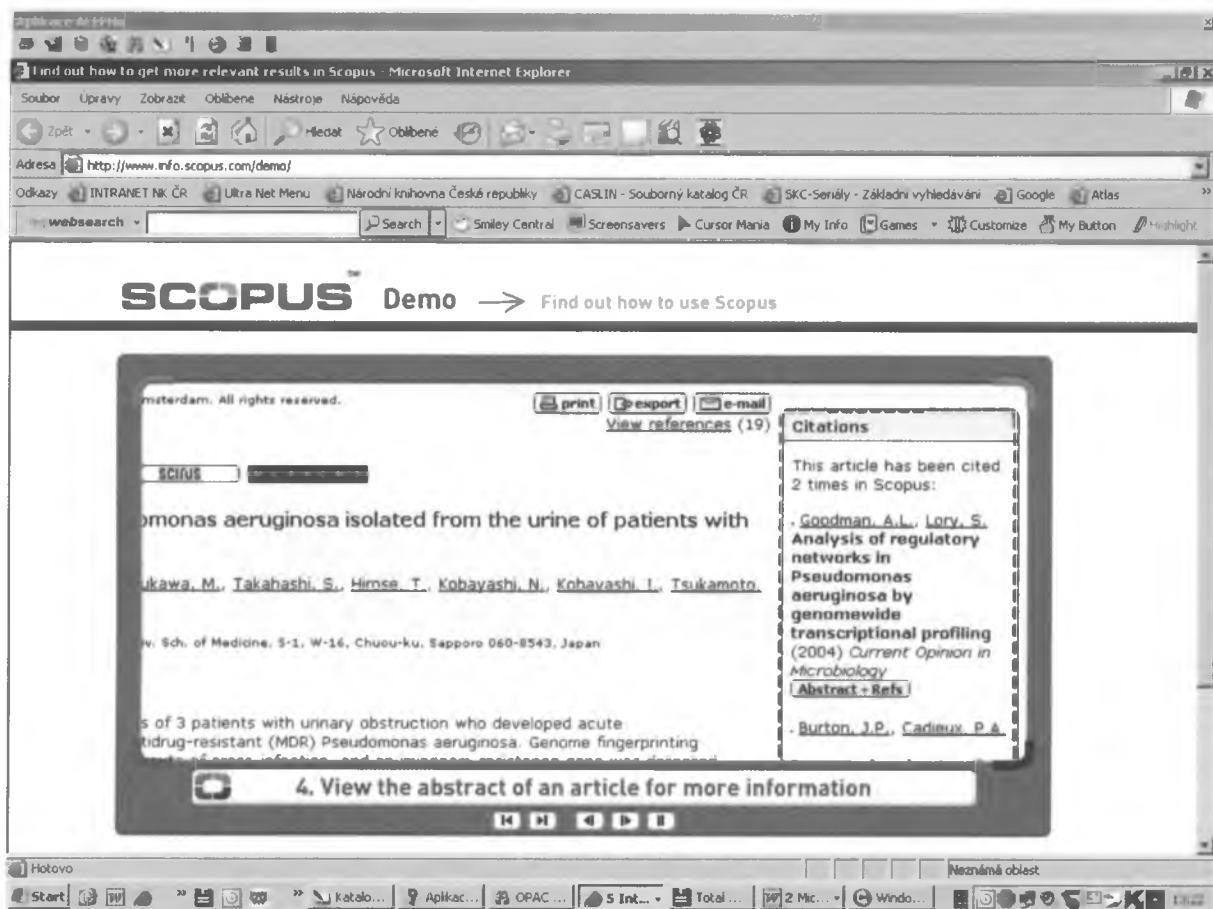
Find out how to use Scopus

1. <input type="checkbox"/>	2003	Puerperal intrauterine infection caused by <i>Edwardsiella tarda</i>	Mikamo, H. , Ninomiya, M. , Sawamura, H. , Tamaya, T.
2. <input type="checkbox"/>	2002	The magic bullet in the new millennium - Cause for optimism	Sykes, R.B.
3. <input type="checkbox"/>	2002	Multidrug-resistant <i>Pseudomonas aeruginosa</i> isolated from the urine of patients with urinary tract infection	Takeyama, K. , Kunishima, Y. , Matsukawa, M. , Takahashi, S. , Hirose, T. , Kobayashi, N. , Kobayashi, I. , Tsukamoto, T.
4. <input type="checkbox"/>	2001	Efficacy of a 14-day course of oral ciprofloxacin therapy for acute uncomplicated pyelonephritis	Takahashi, S. , Hirose, T. , Sato, T. , Kato, R. , Hisasue, S.-I. , Takagi, S. , Shimizu, T. , (...), Tsukamoto, T.
5. <input type="checkbox"/>	2001	Comparison of in-vitro activities of SCH27899 and other	Ikeima, H. , Yamamoto, T. , Kobayashi, T. , Tsukamoto, T.

3. Browse articles from a particular journal

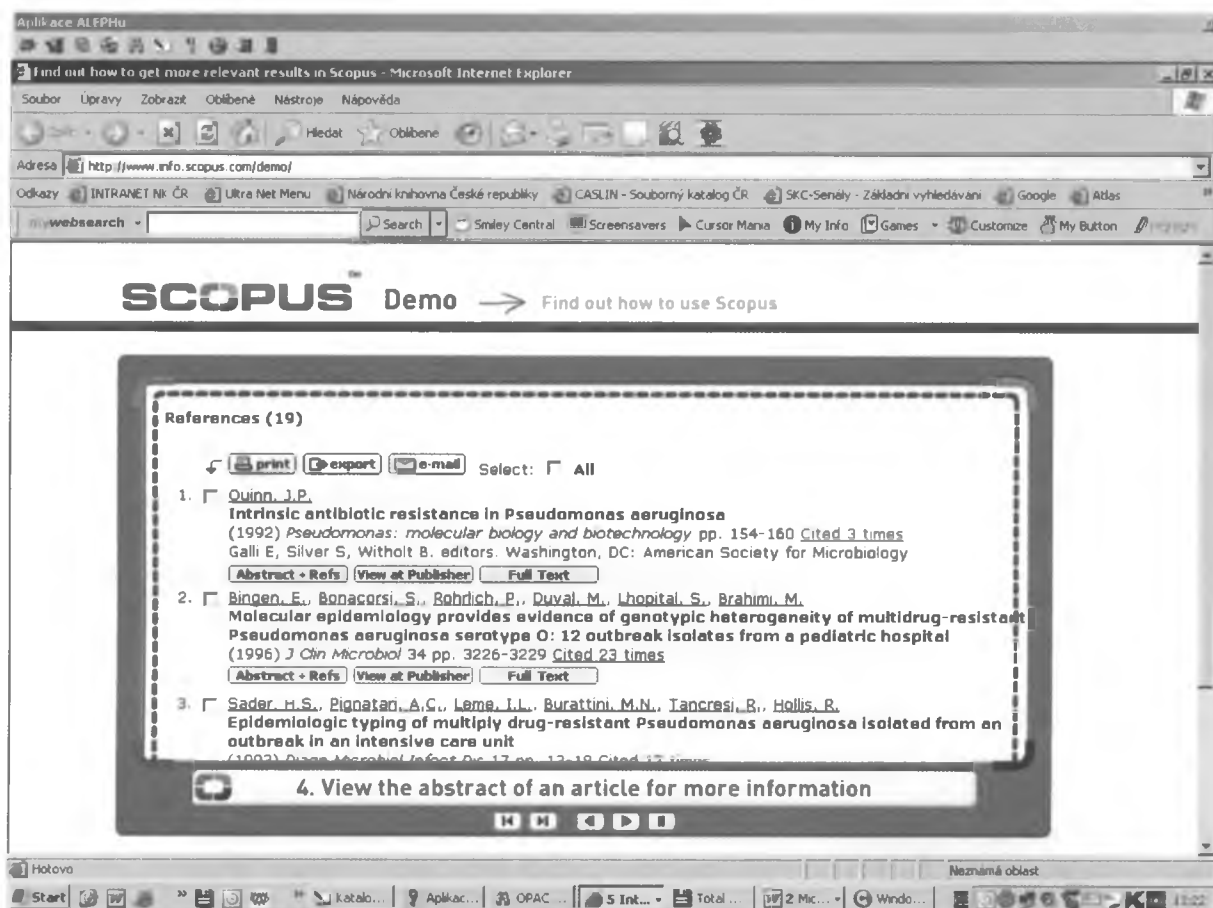
Výsledky vyhledávání, vpravo jsou vidět citace k daným článkům ¹²⁹

¹²⁹ Scopus: www.scopus.com



Podrobnější pohled na citace článku ¹³⁰

¹³⁰ Scopus: www.scopus.com



Podrobnější pohled na citace k určitému článku ¹³¹

15.1. Scopus – vlastnosti a funkce

15.1.1. Vyhledávání:

- Základní vyhledávání: začátek vyhledávání se provádí přímo z domovské stránky
- Pokročilé vyhledávání: dotaz se vloží do příkazové řádky s možností využívat booleovské operátory
- Rychlé vyhledávání: rychlé prohledávání názvu, abstraktu, autorských klíčových slov, klíčová slova z rejstříků a z polí autorů
- autorské vyhledávání: vloží se jméno autora a poté jsou nalezeny všechny články asociované s daným autorem
 - o autorské vyhledávání dovoluje specifikovat dotaz - které výsledky v seznamu se všemi variantami zápisu jména se hodí ke vloženému dotazu (např. „Smith, J“ nalezne „Smith, J“, ale také „Smith Johan“ a „Smith J.L.“)
 - o Vyhledávání jména Patureau v poli autorů nalezne jméno Patureau Mirand.
 - o Hledání Patureau Mirand nebo Patureau-Mirand nalezne Patureau Mirand nebo Patureau-Mirand, ale nenalezne Patureau nebo Mirand
- Můžeme také hledat část jména autora (např. „Moya“ najde „Moya, S.A.“, „De Moya“, „Moya-Cessa, H.“ a „Flore-Moya, A.“)

¹³¹ Scopus: www.scopus.com

- Jména se složenými elementy takové jako de la peña-olvera, f.r. jsou plně podporována
- Lepší porovnávací algoritmy zlepšily kvalitu plnotextových odkazů ze Scopusu od dřívější verze. Nedílnou součástí je databáze CrossRef a zvýšený počet časopisů spojující algoritmicky skrz Scopus vlastní bázi znalostí navyšující kvantitu odkazů.
- Scopus je založen na intuitivním a jednoduchém vyhledávání.

15.1.2. Zobrazení výsledků

- Tabulkovité zobrazení výsledků dovoluje snadno uspořádat výsledky dle data, relevance, autorů, zdrojových názvů a počtu citací (cited-by)
- Možnost předefinovat výsledky dané v přehledu výsledků na dotaz dle zdrojových titulů, jmen autorů, roku, typu dokumentu a předmětné oblasti

Vyhledávání ve výsledcích dovoluje znovu vyhledat ve všech polích.

15.1.3. Propojení na plný text

- Plnotextové články jsou na jedno kliknutí zobrazeny přímo ze seznamu výsledků, stránka s abstraktem a citace
- Různost výběru odkazů jsou dostupné včetně:
 - Odkaz k plnému textu článku na stránkách vydavatele: přes CrossRef a přes vlastní znalostní bázi Scopusu (přidaných 800 časopisům, které nejsou pokryty díky CrossRef)
 - Odkazy „Full-Text“ dovolují zkontrolovat vlastnictví knihovny a odkazy na články v plném textu, ke kterým má daná instituce oprávnění díky odkazovému analyzátoru. Toto obrazové propojování umožňuje instituci používat plné texty. Odkaz na plný text se ukazuje jen, pokud knihovna má právo získat plný text.
 - První odkazový analyzátor implementuje technologii vizuální podoby odkazu a využívá technologie LinkFindesPlus. SFX je také dostupné.
 - Jestliže oprávnění k článku v podobně plného textu není dostupné, knihovníci si mohou vybrat použití standardní OpenURL link k zobrazení alternativních služeb např. document delivery nebo odkaz na OPAC/katalog knihovny.
 - Služba dodávání dokumentů a meziknihovní výpůjční služby mohou být setříděny tak, že přizpůsobený odkaz směřuje uživatele na webovskou formu vybranou knihovnou.

15.2. Odkaz na vyhledávací historii

Relace umožňuje vytvořit přehled až 50 výsledků vyhledávání a dovoluje zobrazit přehled výsledků, editovat vyhledávání, uložit ho pro budoucí potřeby nebo si nastavit upozorňovací služby, kdy uživatel na e-mail bude dostávat informace o nových zdrojích.

15.3. Prohledávání citací a odkazů

- Prohledávání citací na články publikované od roku 1996 dodnes a umožňuje prohledat všechny předmětné oblasti
- Je možné vidět všechny citace, které daný dokument získal od dalších dokumentů pokrytých Scopusem
- Setřídít upozornění na citace dokumentů a možnost získat informaci o nových článcích, které citují konkrétní dokument
- „Related Documents“ navrací seznam článků, které sdílejí citace s dokumenty, které si uživatel zobrazil

15.4. Export výsledků

- Export výsledků vyhledávání v RIS nebo ASCII formátu
- Export výsledků a také export počtu citací (pro dané výsledky)

Obsahové pokrytí Scopusu

Scopus obsahuje databázi abstraktů se 27 miliony záznamů pokrývajících články z 14 000 titulů od více než 4 000 vydavatelů. Scopus je nejobsáhlejší z oblastí vědy, medicíny, techniky a z oblastí sociálních věd a umožňující přístup obsahující veškerou relevantní literaturu. Neustále je zkoumána další literatura, kterou by bylo možné zahrnout do databáze Scopus.

- Nejšířejí dostupné pokrytí vědecké, technické, medicínské literatury a literatury ze sociálních věd
- 14 000 titulů od 4 000 mezinárodních vydavatelů včetně 531 časopisů dostupných přes Open Access
- více než 60% titulů ze zemí jiných než USA
- abstrakty od roku 1966
- citace od roku 1996
- 81% obsahu je indexováno na základě kontrolovaných slovníků
- 100% pokrytí Medline
- včetně unikátních časopisů Medline

15.5. Pokrytí předmětu ve Scopusu

- chemie, fyzika, matematika a inženýrské obory – 4500 titulů
- vědy o zdraví a životě – 5 900 titulů (100% pokrytí Medline)
- sociální vědy, psychologie a ekonomie – 2700 titulů
- biologie, zemědělství a životní prostředí – 2 500 titulů
- obecné vědy – 50 titulů

15.6. Přidávání toho nejlepšího z webu:

Kromě toho je působivé A&I pokrytí, uživatelé mohou vyhledávat nejlepší vědecké weby přímo z rozhraní Scopus. *Scirus*, web search engine pro vědecké informace, umožňuje uživatelům přístup k více jak 167 miliónům stránek relevantních vědeckých

informací na webu, včetně autorských webovských stránek, univerzitních sídel, informací korporací a dalších zdrojů jako preprintové servery, *CogPrint*, *ArXiv.org*, OAI zdroje, stejně tak jako US Patent and Trademark Office (USPTO).

15.7. Odkazy k plným textům přes CrossRef a další

Scopus nabízí předanalyzované odkazy k plným textům článků na vydavatelských sídlech - jsou to typy odkazů „View at Publisher“ (u vydavatele). Ve většině případů tyto odkazy jsou poskytovány přes **CrossRef**, ale k nim Scopus nabízí odkazy k článkům v 813 časopisech od vydavatelů, kteří nejsou členy CrossRef. Tým specialistů se pravidelně věnuje testování a kontrole těchto odkazů, aby byla zajištěna jejich aktuálnost a funkčnost.

15.8. Odkazy na plné texty přes analyzátor odkazů

Scopus je otevřený OpenURL a pracuje s jakýmkoliv analyzátozem odkazů současně dostupným na trhu. Funkce „Full-text“ dovoluje kontrolovat vlastnictví knihovny a odkazy na základě oprávnění k přístupu k plným textům článků pomocí analyzátoru odkazů.

15.9. Jak obrazově založené odkazování pracuje

Scopus jako první kontroluje, zda požadovaná metadata (ISSN a další) jsou nabídnuta ze záznamu (záznamů) Scopusu. Poté, zatímco jsou zobrazeny výsledky ve Scopusu, je poslán požadavek k analyzátoru odkazů ke kontrole, zda uživatel má oprávnění k plnému textu nebo ne. Zatímco uživatel čeká, uvidí obrázek, který říká „Contacting Library“ (Kontaktujte knihovnu). Jestliže uživatel má oprávnění, analyzátor odkazů vrátí obrázek plného textu. Jestliže oprávnění nemá, analyzátor odkazu vrátí transparentní obrázek. Jestliže uživatel klikne na obrázek plného textu, požadavek na plný text je odeslán analyzátoru odkazů, který zpřístupní článek. Tato technologie není nová, ale jen *LinkFinderPlus* a *SFX* jsou současně schopny nabídnout uživatelům tyto inovativní volby propojení.

15.10. Scopus – knihovní integrace

Scopus je postaven na základě úzké spolupráce s vedoucími knihovnami a rozumí se, že Scopus neexistuje izolován bez své knihovny. Je proto vytvořen, aby byl integrován s lokálními zdroji a knihovním systémem a může být přizpůsoben uživatelům dle jejich potřeb.

Scopus je integrován s knihovními zdroji různými způsoby:

- odkazy na plné texty mohou být implementovány přes knihovní analyzátor odkazů
- odkazy na další služby v rámci knihovny mohou být aktivovány, např. OPAC nebo další (předplacené) databáze. Tyto odkazy nabízejí uživatelům doplňkově informace o tom, co lze najít ve Scopusu. Např. odkaz k referenčnímu časopiseckému zdroji *Ulrich's* může poskytovat uživateli informace o jednotlivém časopisu ve Scopusu.

- Služba document delivery a meziknihovní výpůjční služba může být nastavena tak, že dle uživatelského přání odkáže uživatele přímo k webovské formě vybrané knihovny.

15.11. Pokrytí oborů z oblasti umění a humanitních oborů

Scopus pokrývá podkategorie sociálních věd jako právo, vzdělávání, lingvistika a jazyky, kulturní studia stejně tak jako ekonomie a psychologie. Nicméně „jádro“ umění a humanitních věd takové jako architektura, umění, klasika, hudba, folklór, historie, literatura, filozofie (mimo orientované na vědu) a náboženství nejsou Scopusem pokryty.

Scopus pokračuje ve zkoumání uživatelských a zákaznických potřeb z důvodu rozšíření. Jestliže si trh bude přát rozšířit záběr o oblast umění a humanitních oborů, Scopus se rozšíří na tyto.

15.12. Pokrytí Scopusu v oblasti sociálních věd

Scopus v současné době pokrývá přibližně 2 700 titulů z oblasti sociálních věd, ekonomie a psychologie včetně titulů z:

- archeologie
- antropologie
- komunikace
- kulturní studia
- demografie
- vývoj
- vzdělávání
- gender studia
- geografie, plánování a vývoj
- zdraví (sociální vědy)
- lidské faktory a ergonomie
- právo
- knihovní a informační věda
- vědy o životě
- lingvistika a jazyky
- politická věda a mezinárodní vztahy
- veřejná administrativa
- výzkum bezpečnosti
- sociální a politické vědy
- doprava
- urbanistika

Scopus u většiny (85%) záznamů spolupracuje s týmem profesionálů-indexátorů. Jsou využívány kontrolované slovníky velkých databází jako jsou:

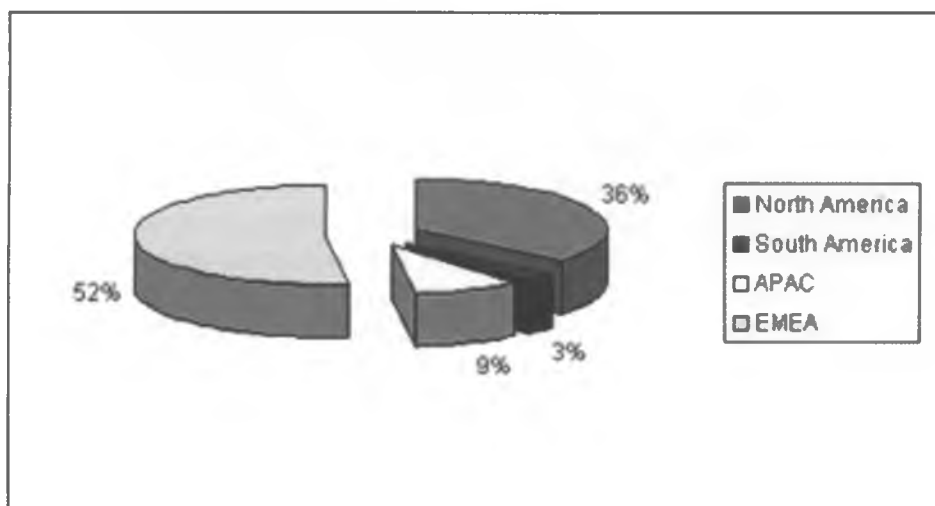
- GEOBASE Subject Index (geologie, geografie, země a životní prostředí)
- EMTREE (vědy o životě, zdraví)
- MeSH (vědy o životě, zdraví)
- FLX termíny, WTA termíny (plyny, textilní vědy)
- Regional Index (geologie, geografie, země a životní prostředí)

- Speciální indexy (biologie, vědy o životě)
- Ei thesaurus (kontrolované a nekontrolované termíny) (inženýrské technologie, fyzika).

Některé články však obsahují pouze klíčová slova nebo slova vzatá z titulu článku. Ale i zde probíhá indexace a porovnání přes kontrolní slovníky, jsou vybírány preferované termíny, odstraněna synonymie a různé podoby zápisu (spelling).

15.13. Geografické rozložení publikací ve Scopusu

Pro zajištění relevantních vědeckých informací není zanedbán v databázi globální pohled. Jsou pokryty časopisy ze všech geografických oblastí včetně neanglických titulů do té hloubky, pokud neanglická provenience obsahuje také anglické abstrakty.



15.14. Vyhledávání

15.14.1. Sort by relevance (třídění podle relevance)

Relevance je definována jako skutečnost, kdy se text dokumentu střetává s předmětem vymezeným v dotazu při vyhledávání. Vychází se z kombinace několika faktorů a uspořádání relevance preferuje dokumenty, jejichž výsledek je nejbližší zadaným kritériím. Užívaná kritéria zahrnují, jak blízko dané termíny jsou, jak jsou ve vybraných dokumentech zastoupeny, počet kolikrát jsou užity v dokumentu a kde jsou v dokumentu nalezeny. Např. pokud vyhledáváme dvě slova, a vyhledaný dokument má tato slova obě obsažená v názvu a často se objeví také v textu dokumentu, bude takovýto dokument z hlediska relevance mezi nejvýše postavenými ve výsledném seznamu – zvláště, pokud se hledaná slova již v dalších dokumentech objeví s malou četností.

Scopus nemá omezení v počtu výsledků – při vyhledávání generuje všechny záznamy.

V záznamu je zobrazeno ve výsledcích celkově 8 autorů na dokument. Jestliže dokument má více než 8 autorů, prvních 7 a poslední jsou zobrazeni. Na stránce s abstraktem jsou již zobrazena všechna jména.

15.14.2. Analyzátor odkazů

Analyzátor odkazů je systém, který získává metadata jako výstup (metadata popisující článek) a počítající odkazy k článkům popsáných výstupem. Analyzátor odkazů pracuje se znalostní bází odkazů a informace o přístupu pro jednotlivé zákazníky. Knihovny často mají některé tituly časopisů s plnotextovým vstupem dostupné přes více zdrojů. Uživatel má pak možnost definovat, z kterého zdroje chce plný text získat.

15.14.3. Scopus

Scopus je vyvíjen od roku 2002 díky intenzivní spolupráci 21 institucí a více než 300 výzkumníků z celého světa. Klíčem k jeho vývoji je rozsáhlé uživatelské testování, výsledky v navigačních nástrojích, které jsou inovativní, intuitivní a dělané pro vědeckou literaturu a tok práce vědců.¹³²

Scopus dle mého názoru by se mohl stát významným doplňkem rejstříků vytvářených Institute for Scientific Information. Díky tomu, že tento produkt vytváří firma Elsevier, která vytváří např. pro oblast medicíny databázi Embase pokrývající mnohem více evropské zdroje než např. Medline, může i v ostatních oborech přinést významné informace a podchytit literaturu, která ISI podchycena není. Především pro výzkumníky z Evropy by to mohla být významná databáze informující o člancích nebo také vědcích, kteří se zabývají daným tématem a o kterých by vědec jinak nevěděl. Mohlo by to i zlepšit komunikace a vytváření nových týmů na evropském kontinentě. Protože i když díky komunikaci např. díky Internetu máme možnost komunikovat s kolegy na druhé straně světa, pro užší spolupráci a např. vytváření výzkumných týmů přece jenom stále platí, že nejlepší jsou kolegové, kteří nejsou od nás příliš vzdáleni. A z hlediska nákladů, kam není takový problém se vydat i navštívit a konzultovat společné téma zájmu. Databáze také díky snadnému propojení nejen na plné texty, ale také k citacím daného článku jde zase o krok dál ve snadnosti práce s databází podobného typu.

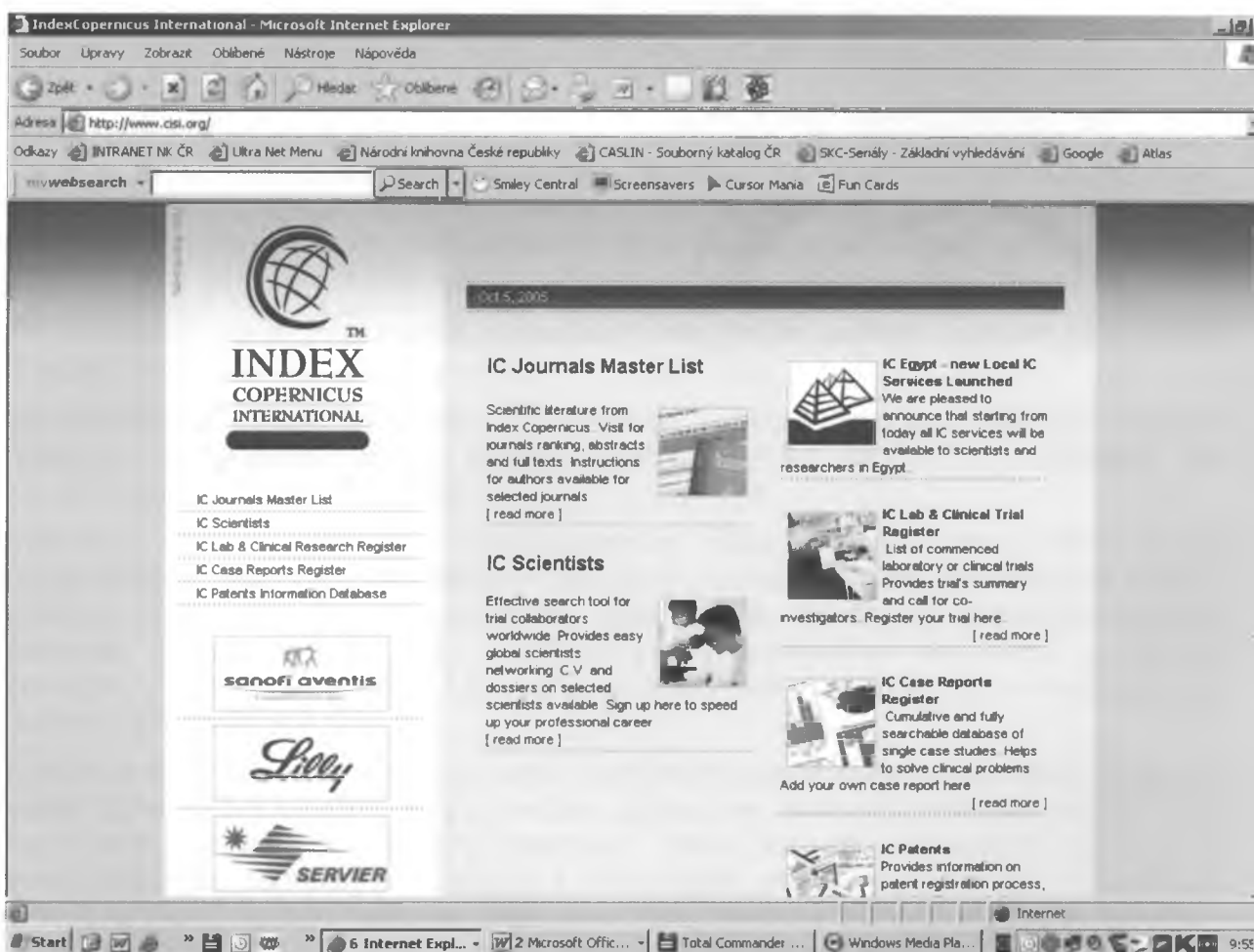
Poslední novinkou ve spojení ze Scopusem je informace, že řada knihoven a institucí má od podzimu 2005 přístup zkušebně do této databáze a to po dobu jednoho roku zdarma.

¹³² Scopus. www.scopus.com

16. Index Copernicus



Současné logo Index Copernicus



Titulní stránka Index Copernicus z října 2005¹³³

¹³³ Titulní stránka získaná z: www.cisi.org

Polský projekt **Index Copernicus** je systém hodnocení především východoevropských a středoevropských vědeckých časopisů (nejvíce časopisů z oblasti medicíny), tedy původně tento projekt byl takto zamýšlen. Dnes lze již říci, že jde o mezinárodní databázi časopisů nejen z oblasti medicíny.

Index Copernicus je vytvářen firmou **Medical Science International**, což je polské nakladatelství vydávající měsíčník *Medical Science Monitor*. Index Copernicus vznikl v roce 1998. Od roku 1999 Index Copernicus je přijímán Komitét Badań Naukowych (KBN) jako oficiální prostředek k řazení vědeckých časopisů.

Cílem indexu je pomocí hodnotícího systému zvyšovat vědeckou úroveň odborných biomedicínských periodik především z východní a střední Evropy a prostřednictvím databáze tato periodika propagovat. Řada těchto časopisů, kterých není zase tak málo, nejsou světové odbornosti známé. Jedním z důvodů je jazyková bariéra, protože většina z nich publikuje v národním jazyce a jen některé mají alespoň anglický abstrakt. Proto také nejsou indexovány v některé z velkých bibliografických databázích a většina nespĺňuje ani kritéria ISI pro zařazení do citačních rejstříků. Tím se časopisy dostávají do začarovaného kruhu. Protože nejsou sledovány citačními rejstříky a nemají impakt faktor, nejsou zajímavé pro domácí autory, pro které je citovanost jedním z hlavních měřítek hodnocení jejich úspěchu. Vědecká úroveň těchto časopisů tedy klesá, což se odráží i v jejich finanční situaci. A právě prolomení tohoto začarovaného kruhu si tvůrci Indexu Copernicu dali za cíl.

Do žebříčku se může dostat každý časopis, jehož vydavatel vyplní "přihlášku" a zašle ji spolu s třemi posledními čísly na adresu Medical Science International. Přihláška obsahuje jednak základní údaje o periodiku (název, periodičita, vydavatel, adresa atd.), jednak informace, podle kterých se provádí hodnocení (např. primární jazyk publikace, počet výtisků, procento publikovaných vědeckých prací, struktura článků aj.).

Při hodnocení, které je popsáno v tabulkách níže, může časopis celkem získat 1 000 "bodíků" (small points), které se rovnají 10 bodům TBS.

Při výpočtu ICS (Index Copernicus Score) se pak rozlišuje mezi časopisy indexovanými některou z databází ISI, u kterých platí: $ICS = 9 + TBS \wedge IF$, a časopisy ISI neindexovanými, kde $ICS = TBS$.

Informace o časopisech jsou někdy anglicky, někdy polsky a také rozsah údajů o jednotlivých titulech se liší. Některý časopis je popsán detailně (to znamená název, hodnota indexu copernicus, nakladatel, jeho adresa, popřípadě webovská stránka časopisu s naskenovanou titulní stránkou a také zařazením dle oboru) a některé časopisy obsahují jen informaci o názvu, hodnotu index copernicu a nakladatele bez dalších informací.

Systém **Index Copernicus** je rozhodně zajímavým citačním indexem, který může být velmi užitečný pro naše vědce a možná jednou se stane významným prvkem pro hodnocení českých medicínských časopisů. Dnes jsou zde především zastoupeny polské lékařské časopisy a časopisy z příbuzných oborů, ale také mnoho českých a maďarských časopisů. Některé naše časopisy i navzdory velkému množství polských medicínských časopisů si stojí velmi dobře a v řazení nejlepších 100 časopisů zauímají přední místa.

16.1. Hodnocení časopisů pro Index Copernicus

16.1.1. Total-Basic Score (TBS – celkové základní skóre)

Parametry hodnocení:

Vědecká kvalita	650 malých bodů	65%
Vydavatelská kvalita	150 malých bodů	15%
Cirkulace	100 malých bodů	10%
Frekvence- pravidelnost-stabilita trhu	50 malých bodů	5%
Technická kvalita	50 malých bodů	5%
celkem	1000 malých bodů	100%

Hodnocení časopisů v Index Copernicu

Malé body jsou předmětem pro převod na konečných 10 bodů TBS.

1000 malých bodů = 10 bodů TBS

16.2. Vědecká kvalita

Ve stanovení vědecké kvality jsou vzaty v úvahu následující parametry:

1. Indexování v mezinárodních bibliografických databázích je bráno jako nejdůležitější faktor vědecké kvality. Jsou zavedeny dvě úrovně:
 - a. základní úroveň – časopisy indexované v mezinárodních bibliografických databázích jiných než Index Medicus/MEDLINE a Current Contents – 150 bodů bez ohledu na počet získaných bodů
 - b. Index Medicus/MEDLINE úroveň (indexované v něm) = 150 bodů (Index Medicus) + 150 bodů = celkových 300 bodů (má se za to, že indexované časopisy v Index Medicus, jedné z nejdůležitějších bibliografických databází, by se mělo objevit jako maximální výsledek v dané kategorii bez ohledu na další indexované stránky).
2. Za povšimnutí stojí, že body pro indexaci v Current Contents jsou přidány v celkovém výsledku vzorce
 - a. originální výzkumné práce / celé práce publikované ročně – poměr (v procentech). Tento poměr indikuje charakter (vědecký nebo vzdělávací) časopisu
 - b. mezinárodní práce / celé práce publikované ročně – poměr (v procentech). Tento poměr odráží stupeň přijetí časopisu mezi mezinárodními autory
 - c. celkový počet prací publikovaných ročně. Tento počet ukazuje autorovu zálohu časopisu
 - d. zda časopis má mezinárodní vydavatelskou radu.

Vědecká kvalita	NEJVYŠŠÍ POČET BODŮ
Celkem Index Medicus a další indexy	300
Další indexy kromě Index Medicus a Current Contents	150
Index Medicus bez ohledu na další indexy	150
Procenta (%) originální práce publikované ročně 0% = 0 bodů > 25% = 10 bodů > 50% = 50 bodů > 75% = 75 bodů <75% = 100 bodů	100
Procenta (%) mezinárodní práce publikované ročně 0 = 0 bodů > 10 = 20 bodů > 20 = 40 bodů > 30 = 60 bodů > 40 = 80 bodů < 40 = 100 bodů	100
100 Celkový počet publikovaný ročně 0% = 0 bodů > 10% = 10 bodů > 20% = 20 bodů > 30% = 30 bodů > 40% = 40 bodů < 40% = 50 bodů	100
Mezinárodní vydavatelská rada 0% = 0 bodů > 10% = 10 bodů > 20% = 20 bodů > 30% = 30 bodů > 40% = 40 bodů < 40% = 50 bodů	50
Celkový součet pro vědeckou kvalitu	650

Hodnocení vědecké kvality časopisů

16.2.1. Vydavatelská kvalita

Vydavatelská kvalita je stupeň dodržování mezinárodních standardů pro prezentaci vědeckých materiálů přijatý **World Association of Medical Editors** (a prezentovaný v dokumentu *Uniform Requirements for Manuscripts Submitted to Bio-Medical Journals*. NEJM. 1997, vol. 336, s.309-315.) Je to mezinárodně přijatá a jednotná dohoda (Vancouver Group) pro úpravu publikací a užívání bibliografických referencí. Pro posouzení kvality časopisů jsou zkoumána další doporučení, včetně užívání klíčových slov z Index Medicus.

VYDAVATELSKÁ KVALITA		NEJVYŠŠÍ POČET BODŮ
Dodržování doporučení Vancouverské dohody		60
souhrn (max = 20 bodů)		
Žádné souhrny	= 0 bodů	
Nestrukturované	= 10 bodů	
Strukturované	= 20 bodů	
2. klíčová slova (max = 20 bodů)		
Žádná klíčová slova	= 0 bodů	
Neužívá MESH	= 10 bodů	
MeSH	= 20 bodů	
3. citace (reference) (max = 20 bodů)		
Žádné citace	= 0 bodů	
Chybějící informace	= 5 bodů	
"en block" (v bloku)	= 5 bodů	
Abecední uspořádání	= 10 bodů	
Po sobě jdoucí uspořádání	= 20 bodů	
4. instrukce pro autory (max = 30 bodů)		30
Vydavatelská politika	= 15 bodů	
Instrukce pro autory	= 15 bodů	
5. vydavatelské informace (max 30 bodů)		30
Seznam vydavatelské rady	= 15 bodů	
Adresa korespondence	= 5 bodů	
ISSN	= 5 bodů	
Jméno vydavatele a adresa	= 5 bodů	
6. prezentace reklamy (max 30 bodů)		30
Ve vědeckých člancích	= 0 bodů	
Před/ na konci článku	= 5 bodů	
Mezi sekcemi	= 10 bodů	
Jen na začátku/ na konci bloků	= 30 bodů	
Celková vydavatelská kvalita		150

Hodnocení vydavatelské kvality

16.2.2. Cirkulace/dostupnost

Cirkulace a dostupnost formuluje další skupinu parametrů. Ta se skládá z:

- jazyka publikace (lokální nebo angličtina),
- prezentace vědecké instituce nebo vědecké společnosti,
- přístup přes Internet

CIRKULACE A DOSTUPNOST		NEJVYŠŠÍ POČET BODŮ
Jazyk publikace (max. 50 bodů)		50
Lokální	= 5 bodů	
Lokální + anglické souhrny (méně než 150 slov)	= 10/20 bodů	
Plný text v lokálním jazyce	= 15 bodů	
Plný text v angličtině	= 50 bodů	
Internetový přístup (max = 50 bodů)		50
Žádný přístup	= 0 bodů	
Instrukce pro autory	= 5 bodů	
Obsah	= 5 bodů	
Shrnutí v HTML/XML	= 5 bodů	
Fulltext v PDF	= 10 bodů	
Fulltext v XML	= 10 bodů	
V angličtině	= 15 bodů	
Celkem cirkulace a dostupnost		100

Hodnocení cirkulace a dostupnosti

16.2.3. Frekvence-pravidelnost-stabilita na trhu

Tato skupina parametrů se zaměřuje na ocenění účinnosti a efektivity časopisů vydavatelské rady. Včasné publikování je důležitý element hodnocení časopisů v dobře známých a respektovaných mezinárodních bibliografických databázích jako je Index Medicus a Current Contents. Delší existence časopisu a jeho tradice jsou podporovány také.

FREKVENCE-PRAVIDELNOST- STABILITA NA TRHU		NEJVYŠŠÍ POČET BODŮ
Frekvence (počet pravidelných čísel za rok) (= max. 20 bodů)		20
Ročně, půlročně	= 5 bodů	
Čtvrtletně	= 10 bodů	
Dvoutměsíčně	= 15 bodů	
Měsíčně	= 20 bodů	
Pravidelně (max = 20 bodů)		20
Nepravidelně (časopisy, které nejsou vydávány včas mají menšení o 30 %)	= 0 bodů	
Pravidelně	= 20 bodů	
Přítomnost na trhu (max. 10 bodů)		10
> 5 let	1 bod	
> 10 let	5 bodů	
< 10 let	10 bodů	
Celková frekvence - pravidelnost – stabilita na trhu		50

Frekvence vydávání a stability na trhu

16.2.4. Technická kvalita

Příprava předtištěného elektronického textu (desk-top-publishing), tištění, barvy a kvalita papíru a formát časopisu jsou také brány v potaz.

TECHNICKÁ KVALITA		NEJVYŠŠÍ POČET BODŮ
Kvalita papíru (max = 10 bodů)		10
offset	= 0 bodů	
Křídový papír	= 10 bodů	
DTP kvalita		10 bodů
Kvalita tisku		10 bodů
Barevný tisk		10
Formát časopisu		
Celkově pro technickou kvalitu		50

Hodnocení technické kvality

16.3. Metodologie bodování Index Copernicus (Index Copernicus Score – ICS) celkové výsledky

Bodování Index Copernicus Score (ICS) výsledky pro časopisy neobsahuje přehled o ISI Master List:

$$ICS = \text{body}^{134}$$

¹³⁴ Pro časopisy, které jsou vydávány pravidelně, jsou body sníženy o 30%, $ICS = 0.7 \times \text{body}$. Index Copernicus bodování (ICS) pro časopisy, které jsou na ISI Master List: $ICS = 9 + \text{body} \wedge IF$, kde IF = hodnota impakt faktor. Takováto metodologie je ve prospěch seznamu dle Institute for Scientific Information ve Philadelphii v USA. Informace z webovské stránky rejstříku Index Copernicus: <http://www.cisi.org>.

JAROLÍMKOVÁ, Adéla, Index Copernicus.

17. Patentové citace

Patenty jsou speciálním případem citací. V patentech existují dva druhy citací: ty příležitostně poskytované vynálezcem a ty, poskytované pravidelněji patentovým průzkumníkem (člověk, který posuzuje, zda má být patent udělen). Průzkumník je porovnatelný s garantem technické zprávy. Garant rozhoduje, zda autor citoval pertinentní předešlou technologii. Vynálezce prohlašuje, že nejlepší z jeho znalostí jeho vynálezu jsou nové. Není požadováno, aby vyhledával literaturu nebo konzultoval s jeho vrstevníky literaturu za účelem potvrzení platnosti jeho prohlášení. To je na patentovém průzkumníkovi. Vyhledávání předešlých technologií průzkumníkem má často za výsledek seznam pertinentních citací. Tyto citace jsou obvykle podkladem v případě zamítnutí jednoho nebo více požadavků na patent. Seznam těchto „citovaných odkazů“, což někdy zahrnuje vlastní odkazy vynálezce, je publikován na konci každého patentu. Citace vynálezce by měly být zahrnuty, jestliže takové citace byly publikovány v jedné předepsané lokaci na začátku nebo na konci patentu. Vynálezce zařazuje tyto citace, aby ukázal stav předešlé technologie sloužící k identifikaci předešlých žádostí a ještě nevyřízené žádosti, navazování patentů atd.

Většina citací, nicméně, je poskytována průzkumníkem patentu a představuje předešlé technologie, které průzkumník používá k zamítnutí jednoho nebo více požadavků. Jestliže požadavky na patent jsou zamítnuty (a okolo 50% všech patentových žádostí spadá do této kategorie), patent není vydán. Průzkumník nemůže citovat pertinentní předběžné technologie, které ještě nebyly patentovány, k zamítnutí požadavku.

Otázkou také je, zda citovaný patent je relevantní k citujícímu patentu. Pokud máme na mysli konkrétní patent, může být podstatné vědět, zda daná technologie byla modifikována, vylepšena nebo zužitkována jakýmkoliv způsobem.

V oblasti patentů není nezajímavé vyhledat jen předešlé patenty, ale také ty, které následují.¹³⁵

Patent Citation Index je částí SCI. Myšlenka vytváření citačního indexu patentové literatury má již dlouhou historii a to od roku 1949, kdy Arthur H. Seidel navrhl v časopise *Journal of the Patent Office Society*, aby byl vytvořen soubor citačních záznamů. Soubor by obsahoval kartu pro každý vydaný patent a karta by identifikovala každý následující vydaný patent, který citoval dřívější patent z jakéhokoliv důvodu. Seidel svým návrhem zaujal Harryho C. Harta v pozdějším vydání *Journal of the Patent Office Society*. Kromě potvrzení myšlenky indexu patentových citací, Hart řekl, že navrhl takový systém dva roky dříve Patentovému úřadu, který předal návrh, se zájmem vydavateli *Shepardových citací* (*Shepard's Citations*). Bohužel v té době nedošlo k vytvoření rejstříku patentových citací.

První citační index patentové literatury byl publikován v roce 1964 v rámci Science Citation Index. Ten se lišil od současné verze *Patent Citation Index* z hlediska pokrytí zdrojových dokumentů. Nebyl jen seznamem citací patentů, které se objevily v časopisech (což je rozšíření zdrojové literatury pokryté současným *Patent Citation Index*), ale také citace, které se objevovaly v patentech, které byly vydány během roku indexování. Zdrojové patenty byly identifikovány v *Patent Citation Index* číslem a jménem vynálezce. Kompletní bibliografický popis, včetně všech vynálezců, právních zástupců, názvu patentu, klasifikačního čísla, data vydání, počtu citací v patentu, byl dostupný v rámci *Source Index* pod jménem hlavního vynálezce. Tato informace dala rešeršérovy základ pro rozhodování, zda by vyhledávání mělo pokračovat zkoumáním

¹³⁵ GARFIELD, Eugene. Patent citation indexing and the notions of novelty, similarity, and relevance

abstraktu hlavního patentového nároku v rámci *Official Gazette*, vyhledáním jeho abstraktu v *Chemical Abstracts* nebo získáním kopie patentu samotného.

Tento index se přiblížil využití plné síly citačního indexování pro vyhledávání patentové literatury. Jeho primární nedokonalost byla v pokrytí citací ve zdrojových patentech, které bylo nekompletní. Patenty poskytované průzkumníkem představují velkou část citací a jsou sbírány od počátku *SCI Patent Citation Indexu*.

17.1. Přirozenost citací

Povaha citací průzkumníka znamená důležitý příspěvek k efektivnosti citačního indexu patentové literatury. Jsou vytvářeny jako část vyhledávání dřívější technologie, která může být základem pro zamítnutí nároku, omezení dosahu aplikace nebo podporu právního cíle patentu. Dřívější technologie vyhledávání průzkumníka je vedena v rámci právního a technologického rámce a je založena nejen na tom, co je doslova shledáno při nalezení vynálezu, ale také co je implikováno. Vyhledávání předešlých technologií a vynálezů vytváří odkazy i na ty podtřídy patentové literatury, který by průzkumník patentu získal při prohledávání v rámci disciplín a specializací v časopisecké literatuře. Citační rejstřík umožňuje soustředit dohromady patenty, které spolu nesouvisejí z hlediska jejich hlavního předmětu, ale jinak mohou mít nějakou souvislost.

17.2. Užitečnost a relevance

Klíčová otázka je, zda vztah identifikovaný citačním indexem nalezne skutečně citující patent, který bude relevantní zájmům hledajícího. Neexistuje jednoznačná odpověď. Neexistuje ani objektivní měření relevance. Jistě, vynálezci a patentový průzkumníci mají problémy shodnout se, co je relevantní.

Na danou otázku je nejlepší jednoduchá odpověď, že citační index patentové literatury identifikuje vztahy mezi patenty, které nejsou identifikovány jakýmkoliv jiným způsobem a tyto vztahy umožňují rychlé vyhledání informací, což je relevantní vyhledání z hlediska ušetření času. Jistě neexistují užitečnější nástroje pro vymezení, zda technologie uveřejněná v konkrétním patentu je modifikována, zlepšena nebo využita jakýmkoliv způsobem.¹³⁶

Původ každé patentové citace je hledání zprávy. Taková zpráva typicky reprezentuje konečný produkt získaný prohlížením tisíců technicky relevantních citací, které se hlavně skládají z patentů nebo tak zvaných citací nepatentové literatury (NPL).

Individuální citace se může vztahovat k jednomu velmi specifickému technickému aspektu žádosti. Firma např. může hledat zpestření sortimentu vývojem aktivit v oblasti technologie, ve které se nikdy neangažovala. V tomto případě dokumentační vyhledávání může přinést přehled veškeré relevantní vědecké a technické informace, přenesení znalosti a „průzkumu“ celé oblasti, analyzování všech důležitých technických proměnných v disciplíně, zjištění, jaká je konkurence a jaké jsou její hlavní aktivity. Společnost takto bude mít přehled o současné situaci a vývoji názorů na budoucí trendy. Dokumentační vyhledávání nemusí být všechno zaměřeno na aspekty vztahující se k patentovatelnosti, tvoří jednoduše souhrnné zásoby technologií, které mohou být požadovány za účelem získání přehledu o specifické oblasti.

Patentové vyhledávání vždy poukazuje na specifický nárok vynálezu a cílem navrhovaným výzkumnou zprávou je poskytnout informace o technickém obsahu

¹³⁶ GARFIELD, Eugene. The application of citation indexing to the patent literature

nazývané stav techniky („state of the art“) – což je relevantní s ohledem na patentovatelnost vynálezu. Hlavním cílem citací je poskytnout základ pro vymezení dosahu právní ochrany. Tento rozsah je definován přesným slovním užitím požadavků a hlavní cíl patentového hledání je vyhledat technickou dokumentaci, jaká povoluje omezení požadavků přesně tak, co nebylo známé v době vyplňování patentové žádosti.

- PN – WO0016531 A 20000323
- AP – WO1999US19138 19990824
- PR – US19980151373 19980910
- PA – SONY ELECTRONICS INC (US)
- IN – LUDTKE HAROLD AARON
- TI – A METHOD AND SYSTEM FOR ELECTRONIC COMMUNICATION
- CT – WO9737202 A [A]; WO9935856 A [XP]; XP002125806 A [A]; XP000484151 A [A]
- CTNP – [A] JASON KRAUSE: "What the Hell is...Jini?" THE INDUSTRY STANDARD. [Online] XP002125806 Retrieved from the Internet: <URL:http://www.thestandard.com/article/display/0,1151,1153,00.html> [retrieved on 1999-04-23];
- [A] WRAY S ET AL: "NETWORKED MULTIMEDIA: THE MEDUSA ENVIRONMENT" IEEE MULTIMEDIA,US,IEEE COMPUTER SOCIETY, vol. 1, no. 4, page 54-63 XP000484151 ISSN: 1070-986X

Mezinárodní patentová žádost s citacemi ¹³⁷

Když porovnáme výzkumné zprávy EPO a USPTO, americký patentový úřad cituje asi třikrát více (patentové citace) a tři a půl krát více u nepatentových citací než EPO. Zatímco průměrný počet německých a britských patentových citací je podobný v porovnání s EPO – nepatentové citace německého úřadu jsou jen asi polovina a asi čtvrtina v britském úřadu průměrného počtu EPO. Důvodem je např. to, že při vyplnění žádosti v USA je požadováno dodat kompletní seznam stavu techniky. V USA to je právní požadavek a nesouhlas patentové žádosti může vést k následnému zrušení patentu (žádosti).

Average number of citations per application based on EPOQUE data (199010–199912)

Office	EPO	DE-PO	GB-PO	JPO	USPTO
Average no. of citations					
Patent	4.37	3.98	3.87	data not available	12.96
Non-patent	0.85	0.47	0.23		2.98

Průměrný počet citací na žádost o patent na základě dat z EPO ¹³⁸

¹³⁷ MICHEL, Jacques; BETTELS, Bernd. Patent citation analysis

¹³⁸ MICHEL, Jacques; BETTELS, Bernd. Patent citation analysis

Average number of citations per application from PCT search reports with ISA as specified; based on SDMS data (200001-200003)

Office	EPO	JPO	USPTO
Average no. of citations			
Patent	4.12	4.20	3.78
Non-patent	0.96	0.48	0.59

Průměrný počet citací na žádost o PCT (Patent Cooperation Treaty) ¹³⁹

Průměrný počet patentových citací je okolo 4 citací na patent. USPTO aktuálně cituje trochu méně patentů než Evropský a Japonský patentový úřad. Zatímco průměrný počet citací je podobný, nejsou významné rozdíly u nepatentových citací. EPO cituje dvakrát tak nepatentových citací jako JPO (Japonský patentový úřad) a asi 1.6 krát více než USPTO.

Average number of citations per application based on EPOQUE data

Office	EPO	EP	US	US
Average no. of citations	trilateral family publication	total average of all EP publicat. (for comparison)	trilateral family publication	total average of all US publicat. (for comparison)
Patent	4.18	4.37	14.26	12.96
Non patent	1.47	0.85	6.29	2.98

Průměrný počet citací na žádost dle EPOQUE dat ¹⁴⁰
 (Pro patentové rodiny – patentová rodina je skupina patentů, týkající se stejného předmětu nebo technologie)

USPTO cituje 3-4 krát více než EPO jak patenty tak nepatentové publikace. Průměrný počet citací na publikaci člena patentové rodiny v Evropě a Japonsku je vyšší, než můžeme spočítat z celkového počtu všech amerických patentů, to je 14.26 na žádosti z patentové rodiny v porovnání s 12.96 celkově. V EPO je tato mezi 4.18-4.37. Citace nepatentové literatury pro žádosti patentové rodiny jsou dvakrát více častější pro EPO a USPTO v porovnání k celkovému počtu výzkumných zpráv EPO a USPTO.

17.3. Jazyk patentových citací

Průměrný patentový průzkumník – vědec s technickým pozadím nebo inženýr – není schopný porozumět technické literatuře v nekonečném počtu jazyků. Průměrný evropský nebo americký průzkumník patentu by pravděpodobně nevstupoval do databáze, která může být jen prohledatelná v japonštině. Ohromující podíl vlastních národních citací je patrný u JPO a USPTO, více než v 90% případů. U EPO více než 90% celkových patentových citací jsou evropské, americké, mezinárodní, německé

¹³⁹ MICHEL, Jacques; BETTELS, Bernd. Patent citation analysis

¹⁴⁰ MICHEL, Jacques; BETTELS, Bernd. Patent citation analysis

nebo britské citace. Hlavním jazykem citací jsou dle důležitosti angličtina, němčina a francouzština.¹⁴¹

ISA	CC cit.	EP	JP	US	WO	GB	DE	other
EPO		23.3%	5.2%	30.9%	16.6%	6.2%	13.1%	4.8%
JPO		0.8%	94.7%	1.3%	2.9%	0.0%	0.0%	0.3%
USPTO		1.0%	1.7%	90.2%	1.7%	1.0%	0.7%	3.7%

Citace dle jazyka¹⁴²

KEELY, Louise C. Using Patents in Growth Models.

Patentové citace mají být velmi užitečný způsob měření kvality patentů. Trajtenberg váží patenty počtem citací, kde citace jsou brány jako zástupci pro hodnotu patentu. Patenty vážené jejich hodnotou (např. citacemi) jsou shledány jako souvztažné s výpočty sociálního nadbytku zboží. Trajtenberg užíváním metodologie také tvrdí, že vztah mezi patentem a výzkumem a vývojem je do velké míry soudobý. Trajtenberg uzavírá tím, že patentové počty mohou být zpracovávány jako typ výstupu vynálezu, spíše než vstupu, což je v souladu s výsledky týkající se vztahu mezi patentovými počty a firemní hodnotou.

Zjistilo se, že citačně vážený kapitál má pozitivní souvztažnost s firemní hodnotou, která předčí vahou, nevážené citační počty a kapitál vážených citací mají pozitivní efekt na kontrolování kapitálu výzkumu a vývoje. Kapitál výzkumu a vývoje má větší souvztažnost s firemní hodnotou než další měření, pokud každé z nich stojí samotně.

17.4. Budoucnost

Budoucí citace mají větší efekt na firemní hodnotu než minulé citace, ačkoliv minulé citace jsou souvztažné s firemní hodnotou. Užití citací jako měření patentové kvality je konzistentní se zjištěními dalších vědců. Ti používají přehled dat např. o vynálezech v USA a Německu a jsou schopni užívat citace jako závislou proměnnou vracející se k patentové hodnotě. Zjistilo se, že hodnotnější patenty jsou, v průměru, také vysoce citované. Tyto výsledky mohou být interpretovány jako empirické ospravedlnění pro použití patentových citací jako váhy pro patentovou hodnotu. Zdá se, že kvalita vážící kapitál citacemi, nám dovoluje vidět, že důležitý kapitál z hlediska technologické akumulace přispívá k hodnotě firmy. Tyto výsledky kontrastují naprosto s vlivem neváženého patentového kapitálu, který má malý nebo žádný vliv na firemní hodnotu, zvláště u patentů prudce překrucující závěry, které mohou být čerpány z empirických výsledků. Není ještě jasné, zda užití citací nám dovoluje rozlišovat mezi patenty jako formy odměny přivlastnění, jako signál o prvenství firmy. Nicméně, interpretování patentů, vážených z hlediska kvality jako měření inovativního výstupu se jeví méně nejisté, než užití pouhých patentových počtů.

Patentová politika se nemění významně, starší patentové chování se neopakuje s ohledem na patentování amerických univerzit a federálních laboratoří. Henderson,

¹⁴¹ MICHEL, Jacques; BETTELS, Bernd. Patent citation analysis

¹⁴² MICHEL, Jacques; BETTELS, Bernd. Patent citation analysis

Jaffe a Trajtenberg zkoumali výrazné zvýšení v patentování a výzkumu a vývoji na univerzitách mezi léty 1965 a 1988. Patentová intenzita, měřená jako míra patentů, je díky financování výzkumu a vývoje zvýšená. Používají patentové citace jako zástupce významu patentu a všeobecnosti jako důkaz, že univerzity zvyšují patenty co do počtu, ale upadají v průměru významnosti a generality.

Citace mohou být používány k porovnání relativního významu patentů a okrajově může z příspěvku citací být usuzováno o zisku. Další způsob měření kvality patentů je přes obnovu a aplikaci výběru patentování firem.¹⁴³

Gittelman a Kogut provedli analýzu publikací a patentů 116 biotechnologických firem za období 1988-1995. V modelu, který spojuje vědecké schopnosti a patentové citace, se ukazuje, že vědecké myšlenky nejsou jednoduše proměňovány do vynálezů; důležité vědecké myšlenky a vlivné patenty provázejí rozdílné a konfliktní vývojové logiky. Objem publikací je spojen s patentovanými inovacemi; spolupráce s univerzitami a s vědeckou silou a vedou k velmi výkonným patentům. Nicméně inovace s velkým vlivem jsou negativně spojeny s důležitými vědeckými pracemi. Tyto výsledky vedou ke konfliktu mezi vědou a inovací; vědci musí přispívat k oběma, zatímco obývají jednu epistémickou komunitu.

Role malých, výzkumně aktivních firem je vytvářet zdroje znalostí; to funguje jako organizační mechanismus pro kombinování schopností mnohostranných vědců uvnitř a mimo hranice firmy; a je potřeba řídit výběr vědy vytvářející hodnotné technické vynálezy.

Jedna důležitá lekce sociologie vědy je, že vytváření vědeckých znalostí je aktivita, která je institucionálně tvořená a organizovaná. Kariéry vědců byly spojeny s jejich úspěchem při publikování výsledků v prestižních časopisech a odolávání následné veřejné kritice. Jak se věda rozvíjí, je roztržena do odlišných komunit, s rozdílnými identitami, časopisy a modely výzkumu a hodnocení.

Sociologie vědy, ačkoliv velmi obohacená, všeobecně souhlasí s pohledem na vědu, která je zakotvená v odlišných komunitách. Jak Merton a ekonom jako Dasgupta a David a Stephan poznamenali, tyto normy vytváří podněty, které jsou použitelné do té míry, že profesionální zařazení je spojeno s úsilím, nadměrnost duplicitního úsilí je součástí a vědci touží velmi rozšiřovat výsledky za účelem získání respektované reputace. Současný publikační systém podněcuje rapidní rozšiřování znalostí v rámci vědeckých komunit a organizuje jednotlivé příspěvky, které se stávají předmětem soutěžení. De Solla Price poznamenal „věda je spiknutí za účelem shromáždit schopnosti mnoha mužů a věda je větší radikální spiknutí, strukturuje toto sdružování tak, že vzniklý celek tohoto druhu znalostí může růst mnohem více než by jakýkoliv jednotlivec byl schopen“. Věda je „manufakturovaná“ v laboratořích, v nichž vědci hledají moc a podporu pro přesvědčení každého o jejich důležité pozici. Publikovaná práce je legitimní nástroj přesvědčení a symbol úspěchu.

Věda je „manufakturována“ v kontextu akademických komunit, firmy mohou užitečně aplikovat vědecké znalosti k rozvoji nových technologií. Argument, že věda by mohla řídit komerční vynálezy, byla hlavní obrana za účelem získání veřejné podpory národních vědeckých infrastruktur. Firmy mohou mít prospěch z vědeckých znalostí, nicméně, věda není přístupná sériovým výstupům, ale je vytvářena vědci situovanými v rámci těchto vědeckých komunit. Užitečná rovnice vědy jako výstup do technologie je problematická, jelikož vědecké výstupy jsou také viděny ne jen jako vytvářející technologie, ale také manufakturuje vědecké výstupy hodnocené dalšími vědci.

¹⁴³ KEELY, Louise C. Using Patents in Growth Models

Logika vědeckého objevování se nedrží stejné logiky, která vládne vývoji nových technologií a tyto konfliktní logiky představují potencionální problémy pro vynálezy založené na vědě. Komunity, ve kterých vědecké myšlenky cirkulují a logiky, díky nimž jsou vybírány, znamenají, že hodnoty kalkulace ve vědě a v průmyslu jsou rozdílné. Vynálezy staví na znalostech vědy, ale věda, která je „dobrá“ pro vynálezy je hnána kupředu logikou, která je rozdílná od té, která zaměstnává vědeckou komunitou, aby stanovila „drahocennost“ nebo „důležitost“ vědy. Zatímco průmysl může potřebovat vědeckou prozíravost, aby rozhodovala technologické problémy nebo našla nové projekty, firmy nepřispívají přímo svými příspěvky k významnosti nebo polemikám vědeckých otázek.

Perspektiva komunity zahrnuje lpění na toku vědeckých znalostí směrem k firmám. Literatura o motivech firem, aby publikovali jejich výzkum ve vědeckých časopisech má nesporně nebo explicitně potvrdit centrální význam vytváření vazeb do této komunity. Např. výzkum ve farmaceutických a biotechnologických firmách odhaluje, hlavně, význam osobních spojů mezi komerčními firmami a univerzitami. Bohatstvím biotechnologických firem jsou spoje na jejich angažmá „hvězdných vědců“ a jejich lokaci ve výzkumných regionech. Významní vědci si často udržují zaměstnání na univerzitě, přičemž podíl na výzkumu ve výzkumných centrech je přirozeným a nezbytným požadavkem úspěšných biotechnologických firem. V řadě studií ve farmaceutice a biotechnologii spolupráce s výzkumníky, kteří působí také mimo firmu a jejich vnitřní vědecká orientace vedou k vyšší výzkumné produktivitě.

Stern v roce 1999 argumentoval, že výzkumně orientovaní vědci mají podstatnou „chuť“ pro výzkum a publikování. Dovednosti těchto vědců jsou potřebné pro přerod výzkumu do vývoje produktů. Firmy jim dovolují se angažovat ve výzkumu a publikovat jako formu platu při výměně pro tyto další aktivity; tyto vědci získávají nižší mzdu, než vědci, kterým nedovolují publikovat. Sternova práce vzbuzuje zajímavé implikace, že vztah komerčních vynálezů a vědeckých znalostí je problematický pro vědce, jejichž aktivity jsou směřovány do komerčních výstupů, ale jejichž identita zůstává zapuštěna v hodnotách a systému odměňování vědeckých komunit.

Výzkum v biologických vědách se spoléhá na menší týmy a je méně kapitálově intenzivní, spolupráce na publikování, ačkoliv společném, je charakterizována méně autory. Pravidla týkající se empirického hodnocení a publikování jsou silná kulturní vyjádření kolektivní organizace individuálních příspěvků ve vývoji vědeckých znalostí.

Citační stopy jsou bibliometrické fosilie díky nimž se měří opakování úspěchu myšlenky. Tyto fosilní záznamy dovolují výzkum relativního úspěchu myšlenky a jejího vlivu v následné práci. Citace prací jsou důležitý způsob, jak vědci hodnotí jejich relativní stálost, díky které si vyměňují dárky, uznávají prestiž, a hledají převahu jejich argumentů. Především patentové citace, to je citace vytvořená pozdějšími patenty k patentům dříve vydaným, podobně ukazují stopy významu komerčních vynálezů, ačkoliv proces díky němuž jsou vytvářeny, je méně uctívá k statutu a efektům reputace. Campbell tvrdí, že akademické obory dělají pokroky díky „vývojové epistemologii“ v níž favorizované myšlenky jsou rozhlášeny a nepříznivé jsou ztraceny. Sociologie vědy ukazuje, že články, které nejsou citovány během pěti let, jsou pravděpodobně zapomenuty. Citační modely k patentům ukazují podobný časový rámec.

Rozdílnost ve vývojové logice, která vytváří práci a patentové citace naznačuje problémy, kterým čelí soukromé firmy, když se snaží převést znalosti, vytvářené ve vědeckém prostředí, do hodnotných technologií. Vědecká a komerční snaha se odlišuje v rozdílných citačních stopách vytvářených charakteristickými pravidly, které jsou určovány logikou, díky níž dobrá práce nebo hodnotný patent je vybrán a reprodukován.

Jestliže citace k vědeckému výzkumu v patentech se drží stejných norem jako citace v jiných pracích, poté vlivné patenty by mohly také citovat vlivné práce. Nicméně, vývojová logika, která vybírá „lepší“ patenty je rozdílná, než ta která vybírá vlivnější práce. Vývojová dynamika pro patenty odráží faktory spojující požadavky trhu, technologické příležitosti a legální požadavky na vlastnická práva. Požadavky trhu zvyšují úsilí ke komercionalizaci a proto patent, více v jednotlivých sektorech, takto favorizuje jen ty, které se k němu vztahují a relevantní patenty. Z důvodu rozdílnosti v těchto dynamikách výběru, nepředpokládáme, že vlivné práce povedou k vlivným patentům. Vědci neodměňují práce jen pro jejich technologický slib či slib uplatnění na trhu, oni je odměňují z důvodů relevance k jejich vlastní epistemické komunitě.

Firmy si přejí, aby jejich vědci se angažovali na vnější spolupráci, aby zlepšili svou produktivitu a dosáhli vlastnická práva k výzkumu vytvářeného v univerzitních a veřejných laboratořích. Vědci spolupracují na publikacích, ale oni tak činí ve shodě s normami vědy.

Shrnutí – návrhy:

1. návrh. Práce a patenty projevují podobné citační modely jako vědecké práce.
2. návrh. Vývojové logiky, které vybírají hodnotné vědecké práce a hodnotné patenty jsou odlišné a protože tomu tak je, vlivné práce s velkou pravděpodobností nevedou k vlivným patentům více než jiné druhy prací.
3. návrh. Výběr spolupráce je ovlivněn statutem ve vědecké komunitě.
4. návrh. Vědci, kteří se účastní přenosu výzkumu do komerční sféry na úrovni objevu a inovace jsou schopni srovnat tyto dvě konfliktní logiky efektivněji než ti, kteří se specializují ve vědě nebo technologii.

Biotechnologický průmysl v USA je charakterizován velkou rozptýleností znalostí a intenzivní soutěživostí v technologiích. Biotechnologické firmy jsou aktivně angažovány v publikování ve vědecké literatuře. Proto v průmyslu tak velmi spoléhají na komercionalizaci základních vědeckých objevů, vztah mezi vědeckými znalostmi a výstupy vynálezů jsou zvláště silné.

Profesionální a poznávací dělení mezi jednotlivci angažovanými ve vědeckém výzkumu a angažovaností v komercionalizaci je mnohem méně ostré v biotechnologii než v dalších oblastech technologie, např. mikroelektronice. Zucker a další ukazují, že specializované biotechnologické firmy jsou formovány s úmyslem ukořistit znalosti držené akademickými vědci poblíž „hranic“ objevení znalostí. Rozběh firmy umožňuje jim získat ekonomickou hodnotu z jejich hodnotných znalostí.

Powell s dalšími ukazuje, že výkon firmy se zvyšuje s intenzitou mezifiremní spolupráce. Shan a další zjistil, že sítě mají vlastnosti samoreplikování, spolupráce si žádá síť stavěnou na předešlých modelech vztahů. Spolupráce se objevuje jako výsledek vytváření užitečných znalostí v trvanlivé síti vztahů pro vytváření patentovatelných znalostí.

Biotechnologické počátky jsou charakterizovány spoléháním na vědecké znalosti, které se nacházejí v prostředí univerzit a akademických laboratoří. Jejich vlastní vědci, zatímco soutěží v produkování hodnotných technologií, se také aktivně angažují na vytváření znalostí.

Vzorové firmy produkují mnohem více publikací, než patentů. Spolupráce s externími organizacemi je definována charakteristikou výzkumných aktivit vzorových firem. Měří se spolupráce na výzkumu reprezentovaná články, v nichž firmy a organizace mimo

firmu jsou jmenovány jako institucionální afiliace jednoho nebo více autorů. Sdílení publikace na základě spolupráce je okolo 70% celkového počtu publikovaných článků.

Významná většina sdílených výzkumů je mezi biotechnologickými firmami na jedné straně a univerzitami na straně druhé, spolupráce typu firma-firma je jen malá část z celku. Z tohoto lze odhadnout, že 90% výzkumných partnerů byly univerzity nebo další výzkumné instituce (vládní laboratoře, nemocnice nebo výzkumné instituce).

17.5. Předešlé patentové citace

Předešlé citační počty kolikrát patent („citovaný patent“) je zahrnut v předešlé dovednosti a objevuje se v následujících patentech. Patentové citace obsahují informace o důležitosti patentových technologií a mohou být také užity jako zástupce ekonomické hodnoty vynálezce. V biotechnologii, kde patenty jsou klíčovým významem a znamenají návrat k vynálezu, citační poměry pravděpodobněji více než v jiných oblastech obsahují informaci o technologické a ekonomické hodnotě daného vynálezu.

Citace k danému patentu mohou přibývat z důvodů, které neodrážejí jeho význam, ale spíše jeho výběr: starší patenty pravděpodobněji budou citovány více, ale nejsou nezbytně významnější než mladší patenty. Efekty technologické oblasti mohou také ovlivnit citační míru: patenty v některých oblastech mohou být citovány více než patenty v jiných oblastech, kde jsou citovanější patenty. Na druhé straně patenty v méně citovaných oborech mohou mít vyšší číslo a mohou být citovány následujícími patenty, protože zde jsou méně citovanými patenty. V těchto případech, důvody pro citování pravděpodobně nesouvisejí s významem patentu, který jsme se rozhodli měřit. Proto, je potřeba pečlivě obsáhnout proměnné, které jsou kontrolovány z hlediska věku patentu a technologické oblasti, tak jako další charakteristiky firmy a citovaného patentu, které by mohly ovlivnit frekvenci s níž je patent citován.

Firmy, které hledají vědecká zjištění, která by mohli integrovat do jejich výzkumu, pravděpodobněji budou citovat vědecká zjištění v jejich patentech, proto mají více nepatentových citací dřívějších technologií. Vědecká intenzita je „zadní“ citační počet, měřený jako počet kolikrát citují patentové citace nepatentové literatury. Deng a další zjistil, že vědecká intenzita je pozitivně spojená s následujícím finančním výkonem pro skupinu technologie – a dané vědecky založenou společností.

Starší firmy mají více času akumulovat znalostní bázi, kterou mohou aplikovat ve vynálezech, nicméně, mohou reprezentovat znalosti staršího data než mladší firmy.

	Průměr	Medián	Maximum	Minimum	Standardní odchylka
Citace k patentům	12.3	7	463	0	21.5
Věk patentu	8.88	9	13	5	2
Velikost patentové rodiny	6.14	6	39	1	5.32
Počet vynálezců patentu	3.05	3	20	1	2.12
Věk firmy	16.9	17	25	7	3.98
Strategie farmaceutické firmy	0.69	1	1	0	0.46
Patenty léků	0.28	0	1	0	0.45
Testování patentů	0.08	0	1	0	0.28

Zástupci publikování firmy	0.39	1	1	0	0.26
Objem firemního publikování	139	31	1.395	0	260
% znovu publikování firmou	0.66	0.7	1	0	0.24
% spojení firmy s vydavateli patentů	0.6	0.63	1	0	0.21
Průměr firmy citování publikací	-0.29	-0.19	5.63	-3.36	1.12
Firemní vědecká intenzita	39.9	28.8	373	0	32.7
Patentová vědecká intenzita	37.9	17	1.675	0	82.8
% patentů spojení vydavatelů patentů	0.69	0.86	1	0	0.37

Souhrnná statistika ¹⁴⁴

¹⁴⁴ GITTELMAN, Michelle; KOGUT, Bruce. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns.

Proměnná	1	2	3	4	5	6	7	8	9	10	11	12			
Citace patentům	1														
Věk patentu	0.36	1													
Velikost patentové rodiny	0.33	0.47	1												
Počet vynálezců patentu	0.24	0.12	0.26	1											
Věk firmy	0.08	0.17	0.05	0.05	1										
Strategie farmaceutické firmy	0.12	0.04	0.11	0.11	0.24	1									
Patenty léků	0.06	0.12	0.22	0.08	0.11	0.25	1								
Testování patentů	0.00	-0.05	-0.07	-0.03	-0.11	-0.22	-0.20	1							
Objem firemního publikování	-0.04	-0.22	-0.06	0.01	0.55	0.28	0.08	-0.04	1						
% znovu publikování firmou	0.09	0.05	0.06	0.09	0.15	0.24	0.06	0.09	0.07	1					
% spojení firmy s vydavateli patentů	0.08	0.05	-0.01	-0.05	0.22	0.26	0.09	-0.16	0.32	-0.09	1				
Průměr firmy citování publikací	0.07	0.20	0.11	0.05	0.29	0.33	0.09	-0.07	0.33	0.25	0.14	1			
Firemní vědecká intenzita	0.09	-0.01	0.09	0.10	0.40	0.32	-0.08	0.47	0.03	0.24	0.19	0.19	1		
Patentová vědecká intenzita	0.07	-0.08	0.16	0.14	0.14	0.11	0.11	-0.04	0.18	0.02	0.06	0.06	0.06	1	
% patentů spojení vydavatelů patentů	0.09	0.07	-0.03	-0.07	0.11	0.06	0.05	-0.09	0.14	-0.10	0.47	0.47	0.47	0.47	1

Souvztažnost dvou proměnných ¹⁴⁵

¹⁴⁵ GITTELMAN, Michelle; KOGUT, Bruce. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns.

17.6. Modely publikačních efektů patentových citací

Modely 1-4 (následující tabulka) zahrnují kontrolu proměnných a přidávají různé vědecké investice proměnných. Model 1 zahrnuje jen kontrolu proměnných. Jak bylo předpokládáno, starší patenty a patenty z velkých patentových rodin získaly více citací. Firmy kódované jako ty, které pracují na vývoji léků, získaly více citací pro své patenty. Na druhé straně, patenty léků nezískaly významně více citací, ačkoliv patenty v kategoriích patentů při testování ano. Výsledky nejsou nezbytně protikladné: firmy s farmaceutickou strategií také patentovaly v kategoriích při testování.

model	1. jen kontrola proměnných		2. jen zástupci firemních publikací		3. spolupráce a vědecká měření firem, které publikují		4. citace k publikacím bez dalších vědeckých indikátorů na úrovni firem	
	β	S.E.	β	S.E.	β	S.E.	β	S.E.
vzorek	Všechny firmy s > 1 patentem		Všechny firmy s > 1 patentem		Všechny firmy s > 1 patentem a > 1 publikací		Všechny firmy s > 1 patentem a > 1 publikací	
Proměnná	β	S.E.	β	S.E.	β	S.E.	β	S.E.
Konstanta	-2.64***	0.34	-2.80***	0.37	-3.67***	0.42	-3.35***	0.40
Logaritmus (věk patentu)	2.02***	0.16	2.05***	0.16	2.38***	0.19	2.19***	0.19
Velikost patentové rodiny	0.05***	0.01	0.05***	0.01	0.03***	0.01	0.04***	0.01
Počet patentů vynálezců	0.06***	0.01	0.06***	0.01	0.08***	0.02	0.08***	0.02
Věk firmy	-0.00	0.01	-0.00	0.01	-0.02	0.01	0.00	0.01
Strategie farmaceutické firmy	0.21***	0.07	0.21***	0.07	0.23***	0.09	0.33***	0.08
Patenty léků	-0.05	0.07	-0.05	0.07	-0.06	0.07	-0.06	0.07
Testované patenty	0.23***	0.11	0.23**	0.11	0.19	0.12	0.24**	0.12
Zástupce firemních publikací			0.13	0.12				
Logaritmus (objem publikací)					0.05**	0.03		
% znovu publikování firmou					0.23	0.15		
Firemní vědecká					0.00***	0.00		

intenzita								
Průměr citací firmy k publikacím^{a, b}					-0.14***	0.04	-0.11***	0.03
Další parametr rozptylu	-0.014***	0.05	-0.14***	0.05	-0.15***	0.05	-0.13***	0.05
N	1.120		1.120		934		934	
Logaritmus pravděpodobnosti	-3.748		-3.748		-3.044		-3.052	

*p < 0.10, ** p< 0.05, *** p< 0.01

a hodnota 1 je přidána ke každému pozorování

b kumulované k roku plnění patentu

Tabulka : Negativní modely dvou proměnných efektivnosti kvality vědecky schopných firem při produkování vysoce citovaných patentů.¹⁴⁶

Model 2 přidává fiktivní proměnnou, aby ukázal, zda firma má jakékoliv publikace k souboru dat ke sledovanému patentu. To je pozitivní, avšak ne významné. Tento překvapující výsledek může indikovat, že věda nemá významný vliv na výstupy vynálezů; nebo to může znamenat, že existují skupiny firem v rámci vzorku a u některých firem věda nemá vliv na vynálezy.

Vědecká intenzita je pozitivní a velmi významná ($p < 0.01$). Patenty vycházející s vědeckých znalostí jsou pravděpodobněji citovány, proto, pravděpodobněji, vytvářejí další vynálezy a hodnoty. Tento výsledek podporuje poznámka, že integrující vědecký výzkum s vynálezy má pozitivní dopad na výkon firem v oblasti patentů a rozšiřuje se o dřívější zjištění, že vědecká intenzita je spojena s hodnotou firmy.

Důležitosti studie je citační průměr firmy na publikaci. To má negativní efekt na citační míru ($p < 0.01$). Vysoce citované práce jsou spojené s méně citovanými patenty. Tento negativní projev implikuje, že vytváření vysoce kvalitních publikací aktuálně snižuje patentovou aktivitu. Toto je výsledek, který říká, že úspěšné patenty a úspěšné práce sledují rozdílné logiky výběru, a že tyto logiky jsou v protikladu.

Souvztažnost dvou proměnných mezi velmi citovanými patenty a velmi citovanými pracemi je pozitivní v 0.36 a zpětný obrat k významným pracím na důležité patenty, ukazuje významný a pozitivní vztah. Nicméně, věk patentu je zahrnut, vztah se obrací k negativnímu a významnému; pravděpodobnost testu ukazuje pomocná proměnná významně zlepšující souhlas. Teoreticky, to je předpokládáno: starší patenty citují starší patenty a znamená to více předcházejících citací. Takto souvztažnost dvou proměnných je nepravá, odrážející věk. Velmi citované články vytvářejí mnohem více patentů, než kontrolní skupina. To je pravda pro skupinu patentů stejně tak jako patenty firem, které jsou autory článku (autopatenty – podobně jako autocitace). Význam vědy je přilákat patentové aktivity, u firem, které vytvářejí vynálezy a u dalších organizací.

¹⁴⁶ GITTELMAN, Michelle; KOGUT, Bruce. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns.

Firmy soustřeďují jejich patentové aktivity okolo významných vědeckých znalostí. To je pravda, pokud význam vědy je měřen v absolutním smyslu (pořadí vzorku) nebo relativní u výstupů firemního vlastnictví vědeckého výzkumu (vzorek firmy).

Patenty, které citují významné články více citují samy sebe: hodnotná věda vede k hodnotným vynálezům.

Skupina	Řazené vzorky	Vzorky firem
Vysoce citované články	232 (46)	194 (29)
Kontrolní skupina článků	44 (6)	58 (13)

Počet patentů citující články: vysoce citované články versus kontrolní skupina (autopatenty v závorkách).¹⁴⁷

Vysoce citované články jsou negativně spojeny s patentovými citacemi ($p < 0.01$). U vzorku firem je projev pozitivní, ale významný ($p < 0.68$). Zjištění jsou silně podnětné, že velmi citované patenty nestaví na hodnotné vědě a, vskutku, významná věda vede k vynálezům, které chudě působí.

Proměnná	Řazený vzorek		Firemní vzorek	
	koeficient	S.E.	koeficient	S.E.
Přerušení	-2.88***	0.49	-3.17***	0.51
Logaritmus (věk patentu)	2.29***	0.23	1.98***	0.19
Vědecká intenzita patentu	0.30***	0.09	0.30***	0.09
autopatenty	0.27	0.27	0.49*	0.28
Citace vysoce citovaného článku = 1	-0.81***	0.33	0.38***	0.14
Parametr rozptýlení	0.67***	0.13	0.38***	0.14
Logaritmus pravděpodobnosti	-452		-444	
N	229		212	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Negativní dvoučlenné modely citací k patentům citujícím vzorové článku¹⁴⁸

¹⁴⁷ GITTELMAN, Michelle; KOGUT, Bruce. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns.

¹⁴⁸ GITTELMAN, Michelle; KOGUT, Bruce. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns.

Firmy jsou silně řízené příspěvím vědců pracujících v akademických institucích, protože málo těchto velkých skupin se kdy objevilo u patentů pro firmu.

Měření přesahu mezi vědci, kteří publikují a patentem na úrovni firmy je důležitý indikátor stupně, s nímž firma je schopna úspěšně přenést výzkum do vynálezu.

	5. efekt firem spojených s vydavateli patentů		6. efekt patentů spojených s vydavateli patentů		7. model se standardními chybami upravený pro shlukování firem		8. model s firmami fixovaným efekty (firemní zástupci se nezobrazují)		9. model ohebnost	
	Všechny firmy s >1 patentem a > 1 publikací	S.E.	Všechny firmy s >1 patentem a > 1 publikací	S.E.	Všechny firmy s >1 patentem a > 1 publikací	S.E.	Všechny firmy s >1 patentem a > 1 publikací	S.E.	Všechny firmy s >1 patentem a > 1 publikací	S.E.
proměnná	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.
konstanta	-3.82***	0.43	-4.00***	0.44	-3.98***	0.61				
Logaritmus (věk patentu)	2.33***	0.20	2.38***	0.20	2.34***	0.12	3.11***	0.34	5.05***	
Velikost patentové rodiny	0.03***	0.01	0.03***	0.01	0.03***	0.01	0.04***	0.01	0.18***	
Počet patentů vynálezů	0.08***	0.02	0.08***	0.02	0.08***	0.02	0.10***	0.12	0.23***	
Věk firmy	-0.02	0.01	-0.01	0.01	-0.01	0.02	0.33	0.44	-0.21	
Firemní farmaceutická strategie	0.20**	0.09	0.25***	0.09	0.25	0.14	-0.52	0.85	0.18*	
Patent léků	-0.08	0.08	-0.08	0.08	-0.07	0.09	-0.09	0.08	-0.02	
Testování patentu	0.22*	0.12	0.21*	0.12	0.21	0.16	0.31***	0.12	0.02	
Logaritmus (objem publikací) a,b	0.04	0.03	0.04	0.03	0.04	0.22	0.18*	0.10	0.15	
% znovu publikování firmou	0.29*	0.15	0.26*	0.15	0.26	0.22	0.40	0.36	0.17	
Firemní vědecká intenzita	0.00***	0.00								
Průměr citací firmy k publikacím ^{a, b}	-0.14***	0.04	-0.13***	0.04	-0.013***	0.05	-0.15*	0.08	-0.05***	
Další parametr rozptylu	-0.15	0.05	-0.16***	0.05	-0.16	0.07	-0.46***	0.06		
N	934		934		934		934		934	
Logaritmus pravděpodobnosti	-3.042		-3.040		-3.040		-2.916			
% znovu publikování	0.29*	0.15	0.26*	0.15	0.26	0.22	0.40	0.36	0.17	

firmou										
Firemní vědecká intenzita	0.00***	0.00								
Průměr citací firmy k publikacím ^b	-0.14***	0.04	-0.13***	0.04	-0.013***	0.05	-0.15*	0.08	-0.05***	
Patentová vědecká intenzita			0.07***	0.03	0.07*	0.04	0.05**	0.03	0.19*	

Negativní dvoučlenný model efektivity mostu vědců na produkci vysoce citovaných patentů¹⁴⁹

Firmy s přístupem k dokonalým zdrojům znalostí nebo dovedností by měly provádět toto se slabšími zdroji nebo dovednostmi. Modely ve studii neposkytují silnou podporu pro tuto hypotézu. Velmi citované patenty jsou spojené s vědeckou intenzitou a firemními efekty, ale ne s ostrím vědy; jsou spojeny s vědci, kteří publikují, ale jen málo spojeny s objemem publikací. Bylo zjištěno, že schopnost vytvářet excelentní vědu má silně negativní dopad na míru citací patentů. Modely indikují, že investování do vědeckého výzkumu vytváří smíšené výsledky a vztah mezi vědou a vynálezem je komplexnější, než lze jednoduše předpokládat.

Místo hladkého vnitřního přenosu mezi vědeckými schopnostmi firmy, výsledky indikují velmi rozdílný proces zahrnutý v získávání vědeckých znalostí a vytváření velmi vlivných vynálezů. Negativní vztah mezi vědeckými schopnostmi patentových aktivit naznačuje problematické rozpojení mezi vědeckými znalostmi firmy a její schopností vytvářet velmi vlivné vynálezy. Vědecké myšlenky nejsou jednoduchými výstupy do vynálezů; významné vědecké myšlenky a vlivné vynálezy následují rozdílné a zřejmě konfliktní vývojové logiky. To vyvolává otázku: kdo investuje kapitál firmám ve vědeckém výzkumu, když tyto investice se nezdají, že by se vyplatily v podobě velmi citovaných patentů?

Existují dva faktory jevící se jako důležité v předpovídání patentových citací. Velmi vlivné vynálezy jsou založeny na vědecké literatuře a jsou vytvářeny lidmi, kteří investují a dělají výzkum. Tyto faktory nejsou nezávislé jeden na druhém. Spojení patent-vydavatel může splnit důležitou funkci identifikování a aplikování vědeckého výzkumu, který firma by mohla většinou vytěžit z jejích projektů. Tato funkce zahrnuje identifikování, tak jako přístup externích výzkumníků v oblasti, kteří pravděpodobně přinesou nové nebo komplementární znalosti firmě. Jiný způsobem je vytvoření mostu mezi vědeckými znalostmi a vynálezem, což se zdá být závislé na přístupu k jednotlivcům, kteří působí v obou aktivitách, spíše než schopnosti vytvářet hodnotné vědecké znalosti samotné. Práce a patenty nenásledují stejný výběr logik a dále, vědci produkují obojí. Firmy doplňují vědce, kteří mohou úspěšně tvořit most mezi těmito logikami a poskytnout podněty, které podpoří jejich duální aktivity. V tomto ohledu, zjištění implikují charakteristiky na úrovni firmu, které Cockburn se spolupracovníky shledal jako důležité.

Vědecké znalosti a patenty mají spojitost, ale dobré publikace a dobré patenty ne. To může být jednoduše vysvětleno tak, že dva artefakty nejsou vybrány stejnou vývojovou logikou. Jinými slovy, patentové citace jsou filtrovány společným vlivem technického

¹⁴⁹ GITTELMAN, Michelle; KOGUT, Bruce. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns.

bohatství a vlivu trhu. To jsou velmi rozdílná vývojová kritéria, než která jsou viděna ve světě publikování.

Toto filtrování technologicky hodnotných patentů výběrem dynamik mezi patentovými citacemi znamená, že existují technologické a tržní komponenty k patentování. Jisté patenty bohatší technologické povahy znamenají pozdější výhody pro příbuzné technické znalosti, které podporují více vynálezských aktivit v této oblasti a proto, více patentů. Citování počátečních (předešlých) patentů, které otevírají tuto cestu technologickým vynálezům. To je zpětná vazba, která tesá cestu v patentových modelech. Modely patentových citací nepotvrzují, co Merton nazývá Matthewovým efektem v prestiži vědy, tedy přitahování citací a zdrojů; odrážejí technické a tržní příležitosti.

Firma, která má skvělé schopnosti dělat vědecký výzkum, té se nedaří dobře při vytváření prodejních vynálezů, jak indikuje Sternova analýza vědecké mzdy. Nicméně, mají reputaci pro vykonávání "dobré" vědy, mohou přitahovat druhy lidí, firma potřebuje inovovat. Firemní heterogenita v aktivitách center vynálezů, které mají schopnost překládat znalosti vytvářené v rámci epistemické komunity vědy do znalostí, které trh ohodnotí. Vědci, kteří publikují a vynalézají jsou pomocní při vytváření mostů mezi vědeckými znalostmi a významnými technologiemi. Heterogenita při aktivitě vynálezů vychází ze schopností firmy k přístupu a vytváření schopnosti dělat vědu, zatímco přemosťují vývojovou logiku, která vybírá mezi jejími výstupy. Tato role směřuje k potencionálním rozdílům ve schopnostech firem doplnit a řídit intelektuální kapitál, jak zjistili ve svých studiích Cockburna a Henderson.¹⁵⁰

Existují rozdílné modely citačního chování u patentů, což náleží rozdílným oblastem průmyslu. Patenty v některých oborech průmyslu pravděpodobně mají mezifiremní a meziprůmyslové přesahy, ale existují obory průmyslu s vnitřními firemními modely citování patentů nebo modely v rámci oboru průmyslu. Úspěšná spolupráce v oblasti průmyslu závisí na citování patentu.

Stimulování firem, aby spolupracovaly na výzkumu a vývoji, sociální plánování posunuje způsob jejich výzkumu a vývoje a chování produkce ze soutěživé na méně soutěživou s výsledkem lepších výsledků.

17.7. Přesahy znalostí: patentová „stopa“

Rozhodnutí patentovat jistý vynález je „strategické rozhodnutí“. Jestliže se firma rozhodne požádat o patent, rozeznává potencionální hodnotu vynálezu. Samozřejmě to neznámá, že nepatentované znalosti jsou bezcenné, ale mělo by to obhájit, že patentované znalosti jsou s největší pravděpodobností komercializovatelné. Hlavním cílem patentu je chránit „individuální“ činnost vynálezu a udělit vynálezci dočasná práva na exkluzivní prospěch jeho inovativní myšlenky. Ale jak objem „vynálezu investováním“ se zvyšuje v čase, patenty musejí rozvíjet opatření, také práva investora, který zaměstnal vynálezce a/nebo poskytl prostředky výzkumu. Patenty se stávají ochráncem ne jen individuální činnosti tvoření, ale jsou také výsledkem přímého a řízeného investování do výzkumu a vývoje.

Shledán jako čistý legální dokument, patent se mění na nástroj strategického soutěživého chování. Firmy staví jejich portfolia intelektuálního vlastnictví, obchodují s patenty, prodávají licence, vytvářejí patentová sdružení pro další firmy. V některých

¹⁵⁰ GITTELMAN, Michelle; KOGUT, Bruce. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns.

oborech průmyslu patenty mají rozhodující strategický význam. Ve farmaceutickém průmyslu např. není dostatečné patentovat jen jednu molekulární strukturu pro efektivní ochranu vynálezu. Malé molekulární změny stejně aktivních komponentů musejí být také patentované, takto to firmy v chemickém a farmaceutickém průmyslu aplikují pro četné patenty, aby ochránily jejich vynálezecké úsilí a investice. Firmy v dalších oborech průmyslu se také stávají aktivnější v patentování.

Patenty zahrnují významnou část komerčně hodnotných znalostí, to je racionální uvažování výhod zužitkování patentových dat v analýze.

Patenty znamenají firemní strategie chování v oblasti výzkumu a vývoje. Obsah patentu se skládá z informací ověřených a předaných později kontrolnímu orgánu. Patentová citace je přezkoušený důkaz předešlých znalostí užitých vynálezcem (vynálezci), kteří dostanou daný patent. Tyto předešlé znalosti eventuelně pocházejí ze stejné patentové domény. Patentové citace ukazují přesah znalostí jednou chráněných (to je uznaných jako potencionálně významné) souborů znalostí k jiným.

Studium patentových citací má své omezení. Výhody a nevýhody užití patentových citačních dat jsou diskutovány Grilichesem a Jaffem. Patentové citace jsou spojené s patentovou procedurou samotnou. Zachycují jen toky znalostí, které se vyskytují mezi patentovaným „kouskem“ vynálezu, takto nedoceneným aktuálním rozsahem přesahu znalostí. Další významy přenosu znalostí nejsou zachyceny patentovými citacemi. To znamená: koupě zboží s technologiemi, které se toho týkají, zaměstnání inženýrů a dalšího kreativního personálu z dalších firem a institucí, dobrovolná výměna znalostí na konferencích a ve vědeckých publikacích atd. Co se týká nepatentových citačních způsobů výměny znalostí je nezbytné zdůraznit, že jen patentové citace jsou velkým rozšířením a dokončením jako reprezentace takovéto výměny. Znalosti neformálně požadované nebo nepřímo, se pravděpodobně stávají objektem diskuzí s dalšími ekonomickými činiteli. Takové debaty jsou společné v praxi obchodu a poskytují podstatnou část dat, pokud jsou použity pro analýzu výměny znalostí patentů. Patentové informace jsou lépe ochráněny proti takovému narušení, protože je jasně naznačeno vlastnictví konkrétního kousku znalosti, což je chráněno zákonem. Patentové diskuze jsou také možné, ale obvykle jsou rozhodnuty rychle autoritativními institucemi.

Existuje více nedostatků v patentových citačních datech. Průzkumník patentu má právo přidat další citace, které on či ona shledá jako aplikovatelné v daném případě, dokonce i když vynálezce nemá informace o dalších přidaných citacích vynálezů. Přidání nových citace patentovým průzkumníky je široce uplatňované v USPTO (United States Patent and Trademark Office) a EPO (European Patent Office), průzkumníci jsou schopni to dělat také. Např. u amerických patentů takto není možné efektivně rozlišit mezi „originální“ citací a citací přidanou průzkumníkem. Má to výhody včetně velkého objemu dostupných dat, jasnost patentového požadavku učiní z patentové citace dobrý objekt pro analýzy přenosu znalostí.

Patentové citace poskytují měření přesahu znalosti, což je rozdílné od dalších konvenčních měření.

Práce *Knowledge spillovers in Belgium* se zaměřuje na patentové citace belgických firem. Patentové citace jsou uspořádány do sekcí:

- geografická distribuce citací
- na firmu orientované rozložení patentů a citací
- struktura „zpoždění“ citací mezi citujícím a citovaným patentem
- distribuce citací mezi různými druhy průmyslu

17.8. Geografická distribuce citací

Základní geografické rozložení patentů belgickými žadateli. Tabulka ukazuje seznam 10 zemí, ze kterých pochází nejvíce originálních citovaných patentů. Seznam se skládá se zemí s více jak 1% celkových citací a pokrývá tak 96.2% citací.

Patenty z USA jsou nejčastěji citované belgickými firmami, což dovoluje předpokládat existenci velmi silného „transatlantického“ toku znalostí. Japonský přesah znalostí je také dost silný. Další pozice jsou zabrány zeměmi Evropské unie (EU) a Českou republikou, která v době výzkumu byla kandidátem EU. Předpoklad „geografické blízkosti“ není silně podporován nasbíranými informacemi: domácí patenty nejsou nejvíce citované. Ovšem Belgie je velmi malá země ekonomicky otevřená USA a Japonsku.

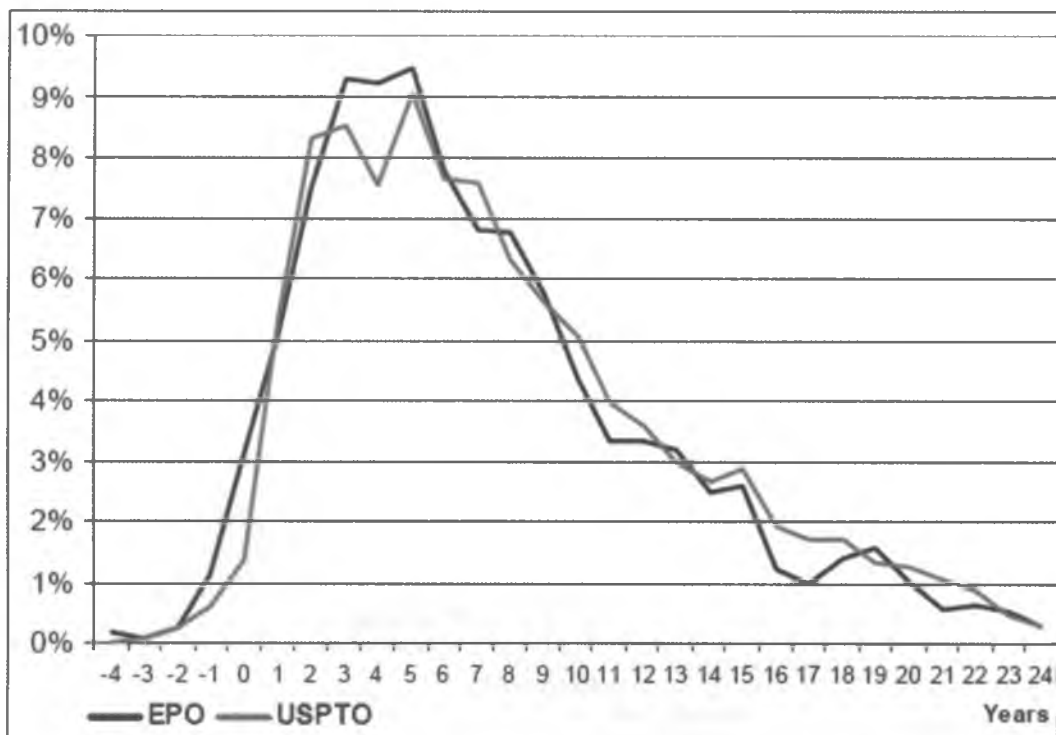
	Country	USPTO	EPO	Total
1	United States of America	42.15%	35.98%	40.62%
2	Japan	18.64%	17.11%	18.26%
3	Belgium	17.18%	20.42%	17.98%
4	Germany	6.07%	7.36%	6.39%
5	France	3.17%	4.15%	3.41%
6	Great Britain	3.19%	3.88%	3.36%
7	Italy	1.63%	2.15%	1.76%
8	Czech Republic	1.93%	0.85%	1.65%
9	Switzerland	1.55%	1.60%	1.56%
10	The Netherlands	0.90%	1.95%	1.16%
	Other	3.59%	4.55%	3.84%

Geografické rozmístění patentových citací belgickými firmami pro patenty udělené EPO a USPTO v letech 1996-2000. ¹⁵¹

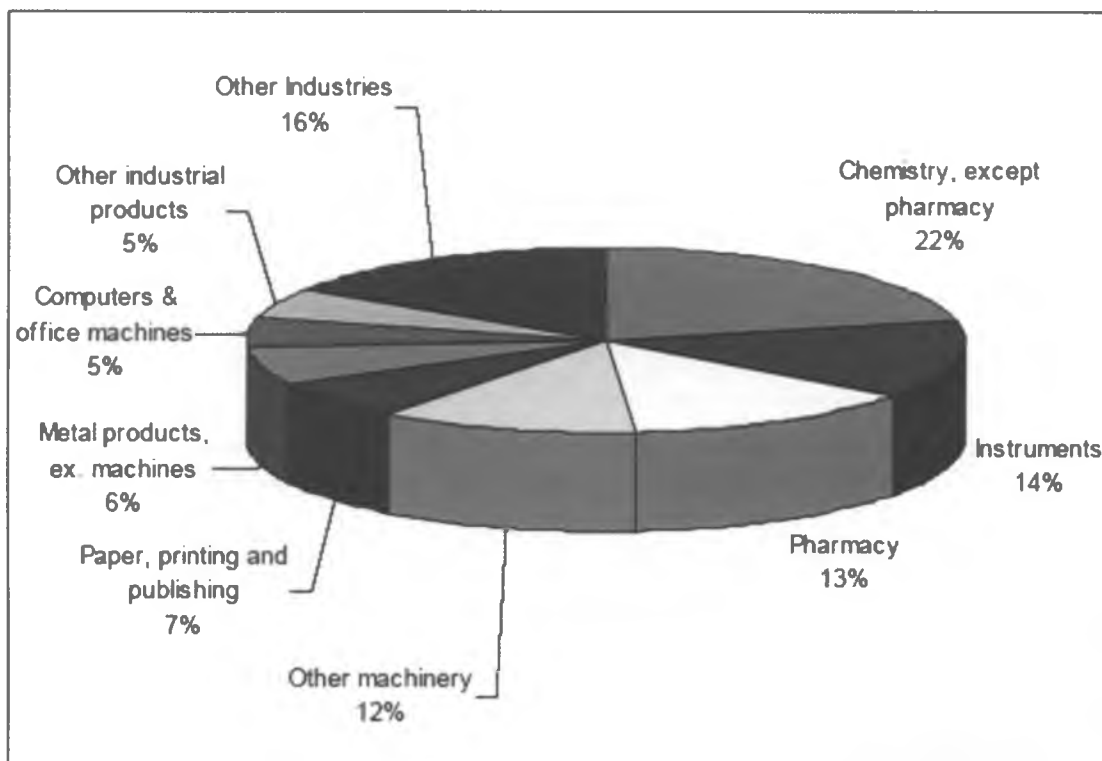
17.9. Struktura „zpoždění citací“ mezi citujícími a citovanými patenty

Data jsou založena na zpoždění a informují o časové struktuře přesahu znalosti. Obrázek ilustruje rozmístění citovaných patentů za různé roky. Základní tvar rozmístění je velmi podobně tvarovaný, jak vypočetli také o citační frekvenci Jaffe a Trajtenberg ve své studii. Obrázek ukazuje, že nedávné patenty (relativně k datu citujícího patentu) jsou s větší pravděpodobností citované než starší. Specifikum procesu průzkumu patentu aktuálně dovoluje pro (malý podíl) negativní citace opoždění hodnot, než se objeví to, jak jeden patent může citovat publikovanou žádost dalšího patentu, který je udělen později než citující patent sám nebo kdy citovaný patent je znovu publikován.

¹⁵¹ LUKACH, Ruslan; PLASMANS, Joseph. Knowledge spillovers in Belgium: evidence from the firm's patent citation behavior



Časové zpoždění založené na belgických patentech udělených EPO a USPTO během let 1996-2000. ¹⁵²

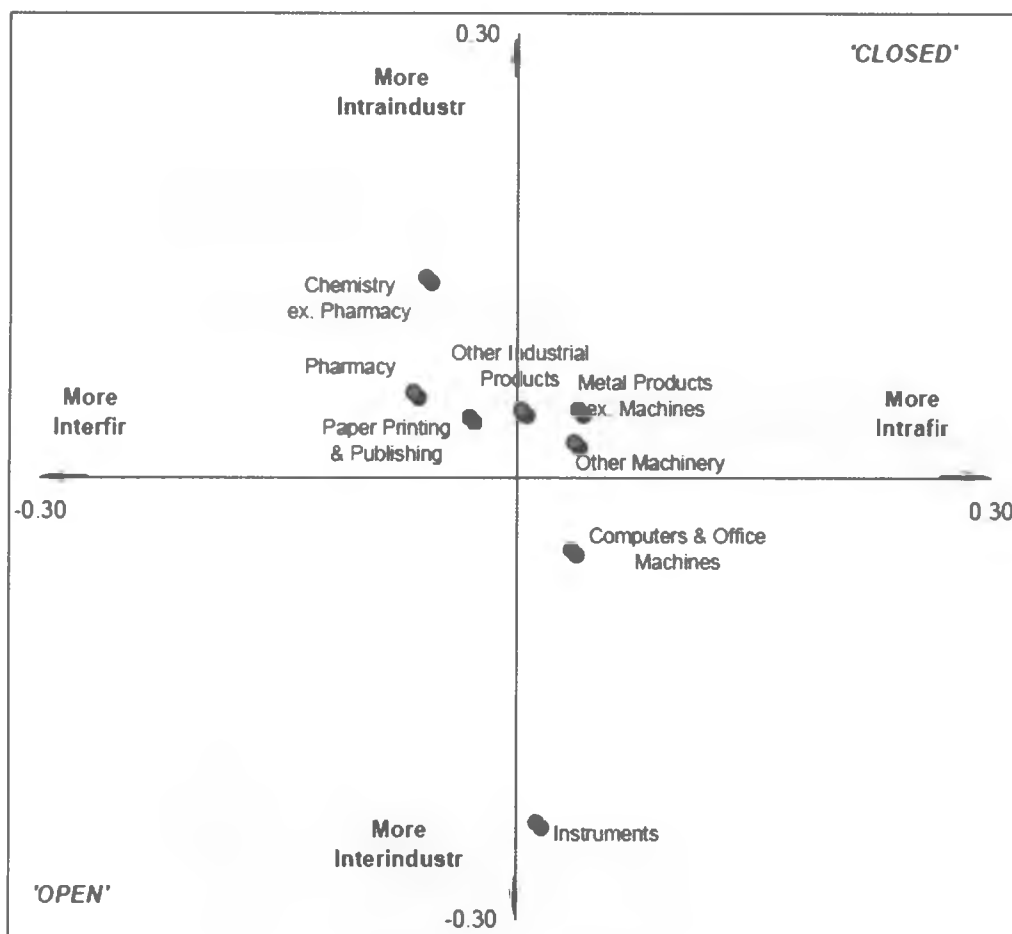


Procenta citací jednotlivými obory průmyslu za léta 1996-2000 ¹⁵³

¹⁵² LUKACH, Ruslan; PLASMANS, Joseph. Knowledge spillovers in Belgium: evidence from the firm's patent citation behavior

17.10. Citované patenty v rámci jedné firmy/ v rámci jednoho oboru průmyslu

Snahou bylo získat výsledky o přesazích znalostí týkající se pozice pro jednotlivé obory průmyslu se vztahem k pravděpodobnosti citací v rámci jedné firmy a citací v rámci jednoho oboru průmyslu. Horizontální osa je koeficient pro stejnou firmu a vertikální osa je pro koeficient pro stejný obor průmyslu. Model popisuje změny v pravděpodobnosti patentové citace ve smyslu návratu.



Pozice oborů průmyslu ve vztahu citace v rámci jedné firmy a přesah znalostí v rámci oborů průmyslu (Založeno na Weibullově modelu) ¹⁵⁴

Vlevo dole obrázku jsou obory průmyslu, které mají větší sklon přesahu znalostí v rámci jedné firmy a mezi obory průmyslu (pravděpodobnost poklesu citací pro patenty náležící stejné firmě a třídě průmyslu). Tyto obory průmyslu nazýváme „otevřené“. Nahoře vpravo obrázku jsou více uzavřené obory průmyslu, které jsou více nakloněny mezifirmním a meziprůmyslovým citacím (citace je pravděpodobnější, jestliže patentová dvojice pochází ze stejného oboru průmyslu a je vlastněná stejným vlastníkem). Vpravo dole je kombinace vyšší pravděpodobnosti citací mezi různými obory průmyslu, ale přesah znalostí v rámci jedné firmy, což, např. může být případ

¹⁵³ LUKACH, Ruslan; PLASMANS, Joseph. Knowledge spillovers in Belgium: evidence from the firm's patent citation behavior

¹⁵⁴ LUKACH, Ruslan; PLASMANS, Joseph. Knowledge spillovers in Belgium: evidence from the firm's patent citation behavior

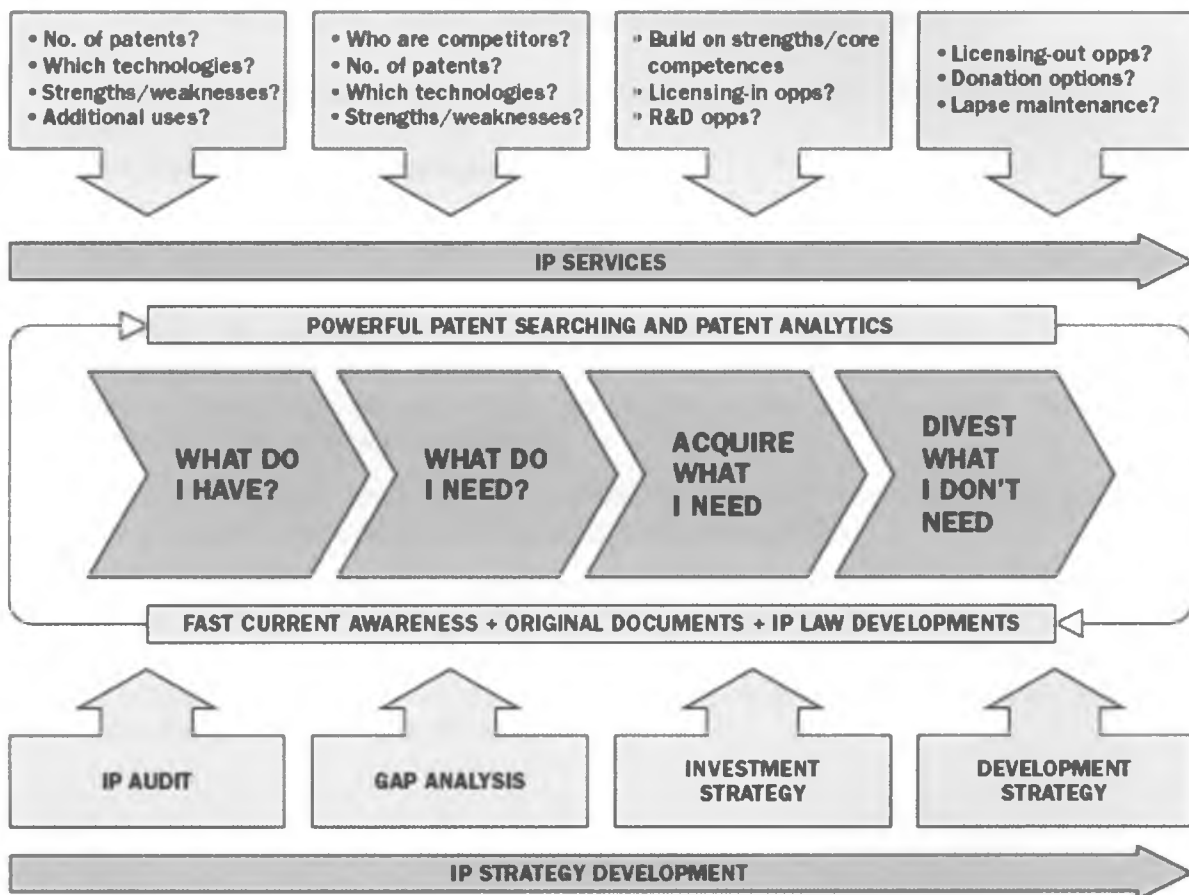
v komplexních technologiích. Nahoře vlevo je kombinace citací v rámci jednoho oboru průmyslu a mezi firmami.¹⁵⁵

Je potřeba mít nástroje a techniky, které pomohou porozumět portfoliu obsahu, jak a kde jdou dohromady s kompetencemi organizace a jaké příležitosti na trhu jsou využity technologiemi. Existuje potřeba identifikovat mezery, kde doplňková technologie může být licencována při identifikování technologie, která se netýká jádra, kde know-how může být licencováno nebo zbyveno a dojde k navrácení peněz, tím se zabývá patent mining (vytěžování patentů).

Organizace by se měla zaměřit na oblast patentu a začít auditem intelektuálního vlastnictví organizace:

- jaké patenty organizace má v portfoliu
- jaké technologie zahrnují
- mohou identifikovat silné stránky organizace (hlavní kompetence)?
- Existují nějaké jasné slabé stránky?

Effective patent portfolio management strategy¹



Efektivní patentové portfolio strategie managementu¹⁵⁶

¹⁵⁵ LUKACH, Ruslan; PLASMANS, Joseph. Knowledge spillovers in Belgium: evidence from the firm's patent citation behavior

¹⁵⁶ STEMBRIDGE, Bob; CORISH, Breda. Patent data mining and effective patent portfolio management.

Není jednoduchá otázka stanovit si vlastnická nebo licenční práva. Zahrnuje to ohodnocení významu technologie a jejích pravděpodobných trhů. Zahrnuje to také ohodnocení síly ochrany nabízené specifickými patenty a zda jsou zamýšleny v relevantních oblastech ochrany technologií.

Je nutné si vymezit, co je vynález a co patent. Např. americká společnost TiVo má 32 unikátních vynálezů a 122 individuálních patentových publikací v různých patentových jurisdikcích.

Firma také musí identifikovat potřeby firmy, jak se dostat přes konkurenci.

Je nutné:

- identifikovat konkurenty
- kolik patentů má konkurence ve svém portfoliu
- z jakých oblastí technologie patenty jsou
- jaké jsou silné stránky konkurence
- zda kladou firmě nějaké překážky

Dále je nutné pokusit se identifikovat díry, které je možné vyplnit řízeným růstem.

- kde může firma zvýšit investice do výzkumu a vývoje jako svou silnou stránku
- jaké klíčové vynálezce by měla společnost získat k zesílení konkurenční výhody?
- Jaké licence by měla společnost mít v technologii za účelem omezení slabých stránek?
- Kde jsou silné stránky konkurence?
- Jaké jsou překážky?
- Existuje příležitost vytvořit nějaké spojení?

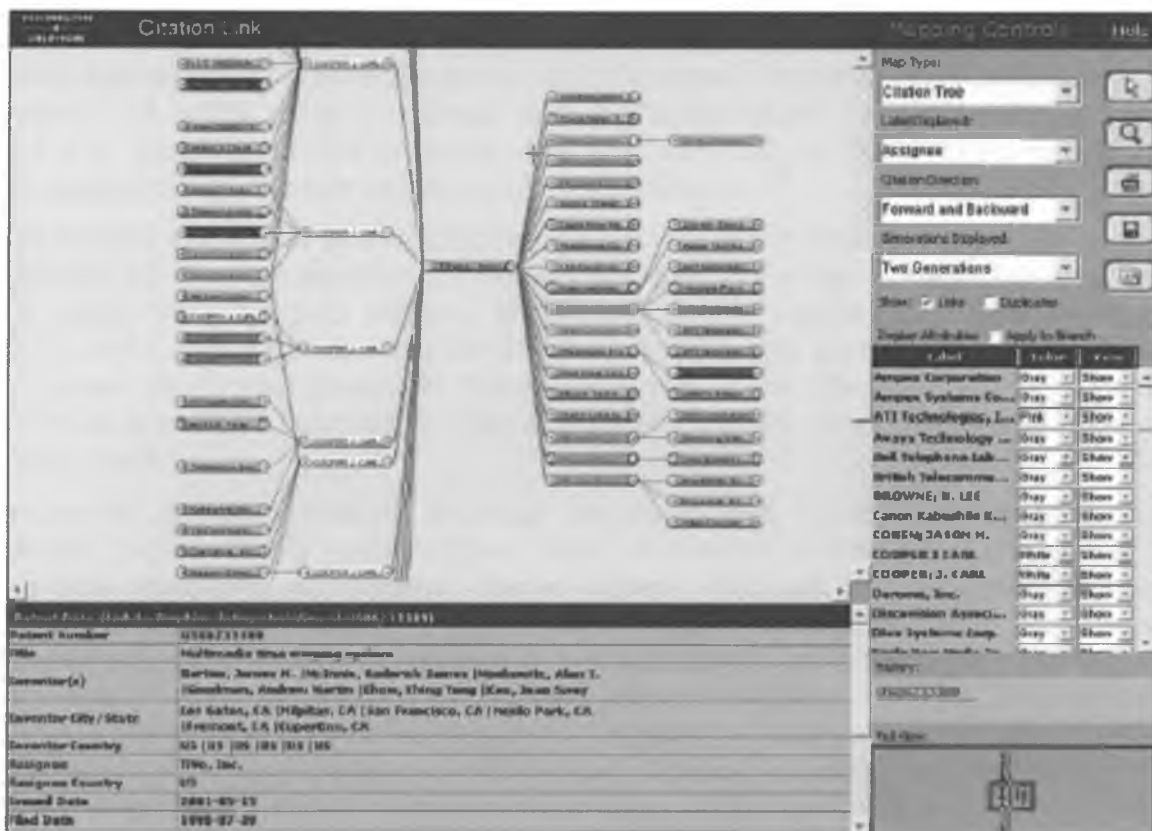
Je nutné se zaměřit na jednotlivce, kteří jsou zahrnuti ve vývoji technologie, mít nějaké klíčové zaměstnance, kteří jsou významní pro vývoj firmy a nových technologií. A případně najít a získat významné vynálezce. K tomu také může sloužit patentová analýza, k vyhledání významných vynálezců.

Licencování patentovaných technologií je důležitá součást intelektuálního přínosu managementu a klíčový zisk vytěžování dat patentového portfolia.

Můžeme využít patentovou analýzu k identifikování těch společností, které patentují ve stejné oblasti technologie a tak identifikovat firmy, které mohou být zajímavé pro získání licencí. Můžeme najít užší spojení mezi cílovou organizací a dalšími hráči na trhu díky užití citací zaměřených na patenty, které reprezentují společný bod zájmu. Můžeme s patentových zájmů vytvořit citační mapu.¹⁵⁷

¹⁵⁷ STEMBRIDGE, Bob; CORISH, Breda. Patent data mining and effective patent portfolio management.

TiVo patent citation map¹⁵⁸



Patentová citační mapa ¹⁵⁸

Přínos znalostí na základě měření ekonomické hodnoty (zisk, tržní hodnota,...) při měření vynálezů (R&D výdaje (= Research and development neboli výzkum a vývoj), patenty,...). Obtížné je měřit vstup do vynálezů ne výstup. Patenty nám poskytují málo dalších informací v rámci R&D

17.11. Mohou být citace užitečné?

Kapitál R&D je důležitý, vysvětluje proměnnou pro tržní hodnotu (v USA a Velké Británii). Patenty také, ale jsou mnohem méně důležité, informují jen málo o oblasti R&D.

Citace jsou používány jako zástupce pro shromažďování znalostí a popisu výzkumných trajektorií. Jsou to omezené důkazy, které mohou korespondovat s čímkoliv „ekonomickým“. Souvztažnost s tržní hodnotou firmy může pomoci hodnotit užití citací v ekonomické analýze.

Citace ve výzkumných pracích slouží jako odkazy k dřívějším technologiím, buď k patentům nebo jiné vědecké literatuře, na které současné patenty stavějí nebo ji využívají. Některé jsou přidávány průzkumníky patentového úřadu, např. USPTO . Některé jsou přidávány k vynálezu, ne však vynálezcem. Některé jsou přidávány, aby se vyvarovalo porušení (omezení rozsahu, ochrana proti žádostem) . Některé jsou přidávány pro „učení“ patentu (jako přehledová články).

¹⁵⁸ STEMBRIDGE, Bob; CORISH, Breda. Patent data mining and effective patent portfolio management.

17.12. Některé skutečnosti o citacích

Dřívější práce nalézající hodnotnější patenty jsou citovány více. Jedna čtvrtina patentů nezíská žádné citace. 0.01% získá více než 100 citací. Prodleva mezi patentem a jeho citováním má určitý sklon (prodlevu) a to nejčastěji okolo 3.5 let. Většina citací se objeví v 10 letech, ale může to nastat větší opoždění (30 let). Počet citací na patent se zvýšil nedávno s příchodem počítačového vyhledávání.¹⁵⁹

Práce *Market value and patent citations* se zabývá ekonomickým významem patentů, především na citacích založených měření užívající hodnocení finančního trhu firem, které vlastní patenty. Bylo zjištěno, že citační zásoby patentů jsou více souvztažné s hodnocením trhu, než zásoba patentů jako taková a tato skutečnost je platná hlavně pro význam ve smyslu postavení firem, které mají velmi citované patenty. Bylo také zjištěno že autocitace jsou cenné dvakrát více jak normální citace, zvláště ty, týkající se menších firem.

Mikroúroveň dat o patentech zahrnuje detailní oblast technologií, citace k dalším patentům, počet nároků, geografickou lokaci a mnoho dalších informací, které jsou stále více dostupné ve strojem čitelné formě. Tyto informace poskytují bohaté technologické, geografické a institucionální detaily, patentová data jsou veřejně dostupná pro všechny druhy výzkumných institucí (firmy, univerzity a další nevýdělečné organizace a vládní laboratoře) ve virtuální podobě v každé zemi. Na obecné úrovni, ekonomové používají data a/nebo patenty váženě na základě citací těchto patentů, které měří tvůrčí výstup organizací nebo geografických jednotek; ekonomové používají intenzitu citací nebo měření významu nebo vlivu entity vynálezů; a používají celkový tok citací jako zástupce pro toky znalostí, aby našli ty, kteří hromadí znalostí napříč organizačními, technologickými a geografickými hranicemi.

V práci *Market value and patent citations* se hledá síla základu pro použití patentu a patentových citačních dat zkoumáním rozsahu zásob tržní hodnoty firem, což je souvztažný jev s jejich zásobou patentů a patentových citací. Trh akcií investorů obsahuje racionální očekávání rozšíření současné hodnoty budoucí zisku firmy, což se liší od její zásoby znalostí. Proto měření založená na patentech má vztah s hodnotou na trhu a reprezentuje důkaz, že zastupují soukromou ekonomickou hodnotu zásob znalostí firmy.

Vynálezy, identifikované jako citace či odkazy zahrnují předešlé patenty a další publikované materiály, které identifikují nebo popisují aspekty relevantních technologií, které byly známy veřejnosti dříve. Citace identifikují dřívější technologie, praxi, která je nezbytně vyloučená z vlastnictví práv udělené patentem. Takto dohromady s jazykem patentového nároku, který přesně popisuje, co je patentovaným vynálezem, co nebylo nikdy dříve učiněno – citace pomáhají vymezit vlastnické právo, které patent reprezentuje. Patentové citace mohou být identifikovány vynálezcem samotným, osobou, vyhledáváním prováděné patentovým zástupcem vynálezce nebo patentovým průzkumníkem, který přezkoumává aplikaci pro patentový úřad.

Na patentové citace lze nahlížet jako na poskytování přímého vlivu technologií a rozmístění znalosti, v jednom technologickém vynálezu explicitně identifikuje několik dalších jako vytváření technologického stavu techniky, na kterém staví. Naneštěstí,

¹⁵⁹ HALL, Bronwyn H.; JAFFE, Adam; TRAJTENBERG, Manuel.

tento optimistický pohled je někdy zamlžen realitou, která je značně „rušena“ v patentových citačních datech. Přirozenost a rozsah tohoto rušení závisí na úsilí s jakým jsou patentová data zapsána. Někteří autoři používají tato data k rozšíření otázky zahrnující prostorové rozmístění, toky znalostí mezi firmami ve výzkumném konsorciu a rozmístění veřejného výzkumu. Při využívání citací jako evidence rozmístění nebo alespoň toků znalostí, z citovaných vynálezců k citujícím, je jasný problém, že mnoho citací je přidáno právními zástupci vynálezců patentu nebo patentovými průzkumníky a mohou reprezentovat vynálezy, které byly zcela neznámé citujícímu vynálezci. Na druhé straně, ve využívání citací získaných patentem jako indikátor významu patentu, vliv nebo dokonce ekonomická hodnota, citace, které jsou identifikovány stranami dalšími, než je citující vynálezce, mohou dobře rozvádět hodnotné informace o velikosti technologické „stopy“ citovaného patentu. To je, pokud patent sleduje oblast v technologickém prostoru, která je později často posuzována, že sousedí s oblastmi, které jsou patentované v budoucnu, to indikuje, že citovaný patent je důležitý, pro kohokoliv, kdo se rozhodne, že citace je nezbytná.

Jeden průzkum vynálezců se zabýval některými těmito problémy. Přibližně 160 držitelů patentů odpovědělo otázky okolo jejich vynálezů, vztah jejich vynálezů k patentům, které byly citovány jejich patenty a vztah k látce patentů, které byly technologicky podobné citovaným patentům, ale které nebyly citovány. Citované patenty a patenty necitované nebyly rozlišovány v přehledu dotazníku, ačkoliv je možné, že vynálezci znali nebo vyhledali i ty, které později necitovali. Výsledky potvrdili, že citace jsou ruchem měření toku znalostí, ale také navrhuji, že mají podstatný informační obsah. Celkově, přibližně polovina všech citací se nezdá, že by odpovídala jakémukoliv druhu toku znalostí; vskutku, v podstatném zlomku případů, vynálezci usoudili, že dva patenty dokonce neměly blízko k sobě navzájem. Ve stejnou dobu, odpovědi odhalující statisticky a kvantitativně významné rozdíly mezi citovanými patenty a necitovanými s ohledem, zda citující vynálezce cítil, zda se něco dověděl z citovaného patentu, když se o něm učil, jak se učil a z čeho se učil. Kvalitativně, zdá se, že asi jedna polovina z citací odpovídá některému druhu vlivu citovaného vynálezu u citujícího vynálezce a asi jedna čtvrtina odpovídá docela bohatému toku znalostí, docela významnému vlivu nebo obojí.¹⁶⁰

Existuje také malý počet studií, které „hodnotí“ užití citačních dat k měření ekonomického vlivu a to tak, že citace mají souvztažnost s nepatentově orientovanými měřeními hodnot. Trajtenbert shromáždil patenty vztahující se ke třídě lékařských nástrojů (počítačová tomografie nebo CAT skenery) a vztah toku patentů v čase s úmyslem odhadnout sociální přebytek přisuzovaného k vynálezům skeneru. Zajímavé práce Lanjouwa a Schankermana také užívají citace, společně s dalšími atributy patentu (počet nároků a počet rozdílných zemí, ve kterých vynález byl patentován) jako zástupce patentové kvality. Zjistili, že měření patentové „kvality“ založené na těchto více indikátorech má význam při předpovídání, jaké patenty budou obnoveny a jaké budou předmětem soudního jednání. Z toho usuzovali, že tyto měření kvality jsou významně spojeny se soukromou hodnotou patentů. S ohledem k univerzitním patentům, Shane zjistil, že mnoho velmi citovaných M.I. T. patentů bude pravděpodobně předmětem licencí a také s větší pravděpodobností základem pro začátek nové firmy. Sampat a Ziedonis zkoumali vztah mezi citacemi a výnosy z licencí u univerzitních patentů.

¹⁶⁰ Patenty mohou obsahovat citace na dřívější patenty a také citace nepatentových publikací.

17.13. Tržní hodnota a patenty

Výzkum, který užívá patenty při srovnání tržní hodnoty jsou trochu limitovány, protože se setkávají s obtížemi při vytváření firemních sad dat, které obsahují patentová data. Je velmi těžké vystopovat měření toků znalostí na základě citací, tento druh hodnocení je obtížný.

Griliches, Hall a Pakes ukazují, že z jednoho důvodu patenty nemohou projevit velmi mnoho souvztažnosti s měřeními jako R&D nebo dochází k tomu, že představují extrémně rušivé měření základních ekonomických hodnot vynálezů s nimiž jsou asociovány. To je protože rozložení hodnoty patentovaných vynálezů je známo jako extrémně vychýlené, to je, několik patentů je velmi hodnotných a mnoho dalších je téměř bez ceny.

Scherer se jako jeden z prvních zabýval tímto problémem. Počet patentů, které firma má je chudé zastupování součtu hodnoty těchto patentů, neměli bychom předpokládat, že souvztažnost bude vysoká. Jestliže počet citací získaných patentem indikuje jeho hodnotu, potom zatížené patentové počty citační intenzity by měli zmírnit asymetrii a zvýšit informační obsah patentů. Rozložení citací je velmi šikmé, navrhuje snad, že to může být odrazem hodnoty rozložení.

Firmy přihlašují patenty s různých důvodů: ochrana výhradních marketingových práv produktu k vynálezu, získání peněz, které budou použity pro výměnu technologií s dalšími firmami, slouží jako orientační bod pro produktivitu výzkumného personálu. Firmy v různých oblastech technologií mají podstatně rozdílné sklony k patentům.

Problémem je, že patenty jsou používány jako měření inženýrské produktivity a jako peníze pro výměny technologií licencování, některé patenty vlastněné firmou mohou reprezentovat stejný druh „suché díry“ ve smyslu, že ukazují technologické cesty, které nemusejí být produktivní.

Myšlenka užívání následných citací k patentu jako měření patentové hodnoty vychází z argumentu, že hodnotné technologické znalosti v rámci firmy mají tendenci vytvářet patenty, na nichž budoucí výzkumníci budou stavět (a proto citují), jestliže pracují na vlastním vynálezu.

Protože citace mohou být vytvořeny kdykoliv po aplikaci patentu, je otázkou, zda citační váhy jsou založeny na všech citacích k patentu aplikované pro tento rok nebo dříve, nebo zda bychom měli užívat jen citace, které se již vyskytly. To je, vytváření zastoupení citací pro hodnocení patentu, který je znám v čase, kdy je patent aplikován nebo vytvořit zástupce pro budoucí hodnotu vynálezu, pro nějž hodnota současného trhu firmy je jen nestranná předpověď?

Existuje alespoň příklad, kdy jedna citace patentu, který je 50 let starý, se vztahuje k patentu, který byl přihlášen v roce 1921 a udělen v roce 1992. Tak dlouhé zpoždění udělení obvykle jsou výsledkem pokračování procesu, jenž dovolují patentová pravidla, pod kterými vynálezce může modifikovat aplikaci patentu, která ponechává aplikační data originálu.

Patenty mají souvztažnost s R&D aktivitou na firemní úrovni, měří něco, co je odlišné od R&D, „úspěch“ ve vynálezecké aktivitě nebo snad úspěch v přidělení obnovy takové aktivity.

17.14. Kdy citace jsou informace?

Existuje rozdíl mezi tržní hodnotou minulých a budoucích citací. Zda zahrneme zásobu citací samotnou nebo použijeme plný model s R&D a patenty, koeficient zásoby založený jen na budoucích citacích je rovný nebo větší než koeficient založený na všech citacích a koeficient minulé citační zásoby je negativní a okrajově významný nebo nevýznamný. Zjevný aspekt je, že budoucí citace jsou více souvztažné než minulé citace s předpokládanou výnosností patentového práva.

Protože existují dvě zásoby, minulé a budoucí, jsou vysoce souvztažné měření stejné základní kvantity, toto zjištění nezbytně neimplikuje, že citace jsou bezcenné pro předpovídání hodnoty přínosu znalostí asociované s patenty nebo předpokládaný tok zisku z tohoto přínosu. Zásoba minulých citací firmy by mohla být skvělou předpovědí budoucích citací, které jsou předpokládány pro patentové portfolio, dokonce i když není tak dobré jako znalost budoucích citací, když předpokládáme tržní hodnotu firmy.

Ačkoliv budoucí citace jsou silnějším indikátorem tržní hodnoty patentového portfolia ve vlastnictví firmy, minulé citace také pomáhají v předpovídání budoucích výnosů.

Je zajímavé zkoumat otázku správné formy rozložení citačního hodnocení. Je skutečnost, že výnos firemních patentů méně citacemi než je průměrná hodnota, znamená, že toto R&D je neproduktivní? Jak se toto hodnocení změní pro firmu s patenty, která má vysoké citační výnosy? Pro firmy s méně patenty než je medián průměrného počtu citací na patent, není rozdíl v tom, jak hodně jejich medián padá; firmy s 4-6 citacemi na patent nemají vyšší hodnotu než firmy s méně než 4 citacemi. Nicméně, firmy, které mají průměr více než medián počtu citací na patent, u nich dochází k výraznému zvýšení jejich tržní hodnoty a je to přibližně lineární křivka. Firmy s více jak 20 citacemi na patent, mají tržní hodnotu o 54% vyšší než by se předpokládalo z jejich R&D kapitálu a jejich patentové zásoby.

Zdá se, že větší firmy jsou především v oblasti elektroniky veliké, zatímco ty v oblasti farmacie, které mají průměrně vysokou míru citací jsou pravděpodobněji menší biotechnologické firmy.

Některé citace, které se objevují u patentů jsou k patentům, které byly přiřazeny ke stejné firmě, jako je citující patent. To označujeme jako autocitace. Protože citace jsou vytvářeny procesem kompilace zahrnutí vynálezce, právní zástupce patentu a patentového průzkumníka, není zcela jasné, jakou interpretaci dává tato autocitace.

Jednou z možností je, že se objevuje jednoduše proto, že patenty stejné firmy jsou dobře známé společnosti nebo protože vynálezcovu přání je přiznat je kolegům. Jestliže je to tak, autocitace by měly být ekonomicky méně významné než další citace. Na druhé straně, firmy citující jejich vlastní patenty by si měly být vědomy kumulativního charakteru vynálezu a „zvyšující zisky“ majetku hromaděním znalostí, zejména v rámci úzkých oborů nebo technologické trajektorie. Autocitace by mohly navrhovat, že firma má silnou konkurenční pozici v konkrétní technologii a je v pozici, aby mohla rozšířit některé znalosti o svém postavení vytvořené na základě jejího vývoje. To bude zahrnovat, že firma má nižší náklady, protože potřebuje méně často získat technologie od dalších a firma tak získá vyšší zisky nebo nájemné za její aktivity bez vstupů podpory. Přítomnost autocitací může také indikovat úspěšné dotace nashromážděného vlivu firmy

Zatímco citace od dalších mohou indikovat kumulativní vliv, který je získán od dalších firem. Jestliže je to tak, potom *soukromá* hodnota autocitace by měla být větší než z dalších citací.

Rozšíření počtu patentů firmy s citační intenzitou informací vytváří zástupce pro zásobu znalostí firmy, které jsou značně více relevantní hodnota než jednoduchý počet patentů sám. Jak zdůraznil Sam Korum, i když citace jsou rozumný informační signál úspěchu, neznamená to, že budou více souvztažné s hodnotou R&D, protože optimalizování firem zvýší jejich R&D jako reakci na úspěch. Zásoba citací je také spojena s významnou přírůstkovou tržní hodnotou po kontrolování firemního R&D. Firmy s hodně citovanými patenty (více než 20 citací na patent), představuje pravděpodobně velké rozdíly tržní hodnoty, zvýšení hodnoty o 50%, relativně k firmě se stejným R&D a patentovou zásobou ale s mediánem citační intenzity.

Citace obsahují užitečné přírůstkové informace o hodnotě, citace jsou nejhodnotnější když:

1. tržní hodnota je souvztažná k významnému rozsahu s podílem eventuelních citací, které nemohou být předpovídány na základě minulých citací
2. tržní hodnota je pozitivně souvztažná s podílem citací k patentům firmy, které firma sama vytvořila.

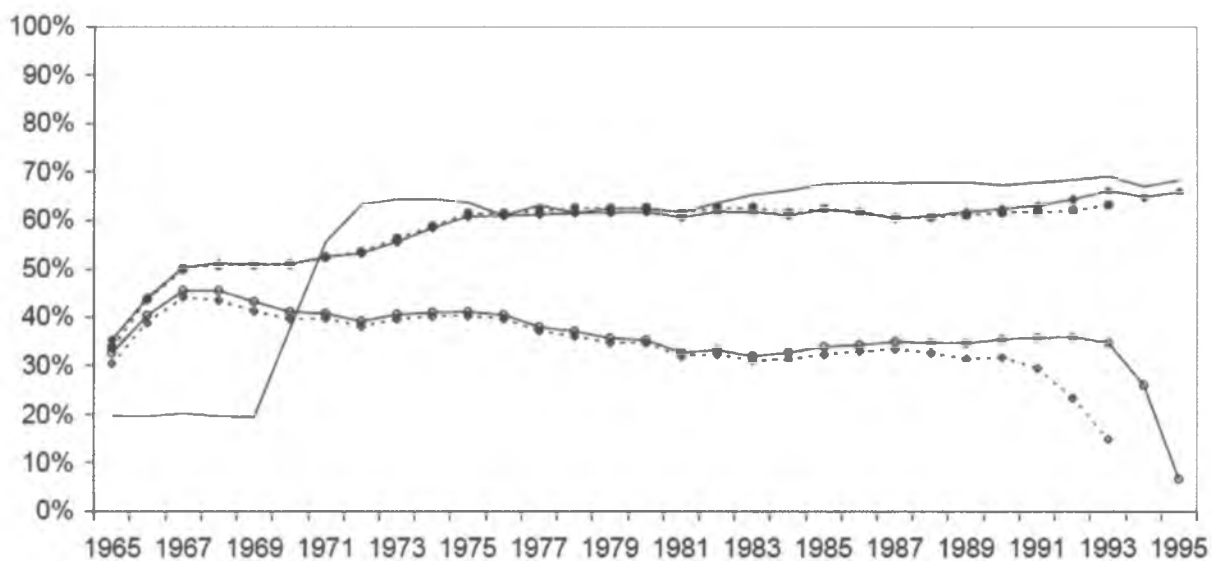
První zjištění navrhuje, že trh „již zná“ hodně o kvalitě vynálezů, což bude konec konců potvrzeno přírůstkem budoucích citací, které jsou „nepředpokládané“ ve smyslu nepředpovídání jen na základě minulých *citačních* informací. Tento výsledek jasně požaduje další prozkoumání. První by bylo užitečné prozkoumat užití funkční formy nebo normalizace, což by dovolovala relativní hodnotu minulých a budoucích citací v porovnání příměji, spíše než jen zeptání se, jaké citace přidají více k R druhé mocnině. Kromě toho, se můžeme zeptat, jak mnoho let cenných citací můžeme vidět, aby to představovalo většinu z citací. Je 10 let dost? Jaký zlomek z toho „známe“ (ve smyslu souvztažnosti s tržní hodnotou) díky znalosti životnosti citací po 5 letech? Mělo by se také prozkoumat, zda tento výsledek je řízen distribucí, jakou známe, je asociované hodně z hodnoty. Jinými slovy, v jakém rozsahu je možné předpokládat, že patent v zásadě vzato získá více jak 20 citací založené jen na prvních několika letech patentů? Je těžké předpokládat skutečně velké výherce, které získají bez předpokladu část z celku citací významně?

Druhé zjištění, že autocitace jsou do značné míry pozitivní pro hodnotu, odhaluje velmi zajímavou cestu výzkumu. Variabilní autocitace nám poskytuje jakési okno do technologické soutěže, ve smyslu, že nás může informovat o rozsahu, jak firmy rozšiřují šířitele znalostí nebo sílu jejich konkurenční pozice tváří v tvář dalších firem v jejich průmyslu.

Další změny ve výsledcích zahrnují více průzkum formy vztahu hodnoty citací. Má význam u citovaných patentů změna v čase se změnami v patentovém režimu? Pod dobu let citační intenzity nad 20 citací získávají jak malé tak větší firmy. Bylo by užitečné vyřešit, zda tyto firmy se liší jedna od druhé a také rozšířit s jakými výsledky mají vztah k *průměru* citační intenzity více než 20 citací, co je za patenty v pozadí.

Další aspekty citačního chování, které by měly být relevantní. Jedna možnost je *všeobecnost*. Trajtenberg, Jaffe a Henderson navrhli měření „všeobecnosti“ definované jako (1 minus) Herfindahl Index koncentrace citací na patentové třídě. Myšlenka je, že citace, které jsou rozptýlené v rámci velkého počtu technologických oblastí, jsou více „obecné“ a naopak. Ovlivnění tržní hodnoty firmy, ačkoliv, člověk by mohl předpokládat následující: pro firmy, které se zaměřují na úzké oblasti aktivity, větší všeobecnost je

špatná, jelikož firma nebude schopna přisvojit si znalosti z dalších oblastí. Pro konglomeráty, protiklad může být pravdou.¹⁶¹



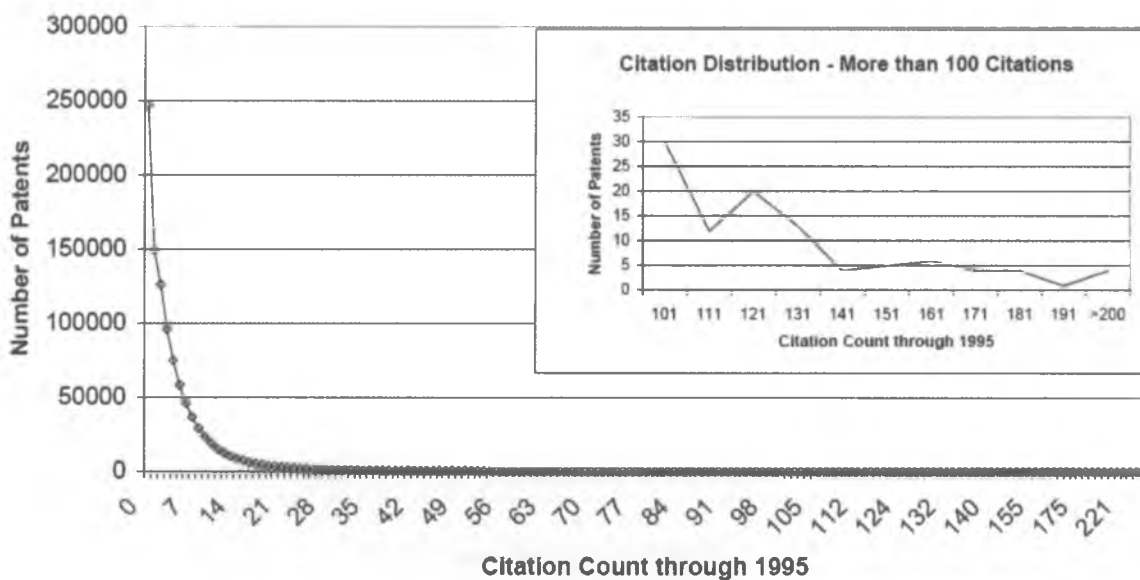
Obrázek Americká výroba – příklad 4 846 firem¹⁶²

Procenta značí podíl firem v procentech

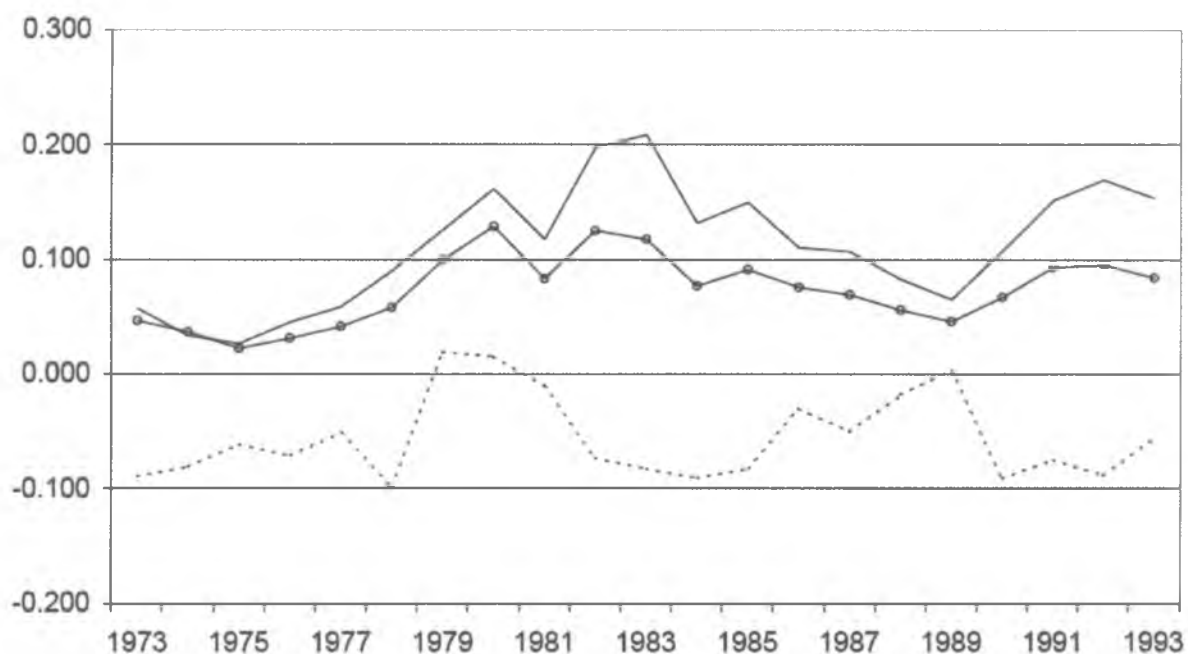
- % s R&D zásobou
- % se zásobou patentů
- - -□ - - % se závislou proměnnou
- - -◇ - - % se zásobou citací
- % současné patenty

¹⁶¹ HALL, Bronwyn H.; JAFFE, Adam; TRAJTENBERG, Manuel. *Market value and patent citations: a first look.*

¹⁶² HALL, Bronwyn H.; JAFFE, Adam; TRAJTENBERG, Manuel. *Market value and patent citations: a first look.*



Rozložení citací pro patenty v roce 1995
(udává počet citací a počet patentů)¹⁶³



Firmy, které se podílí na výzkumu a vývoji (R&D) – rozdělení citací
v minulosti a do budoucnosti¹⁶⁴

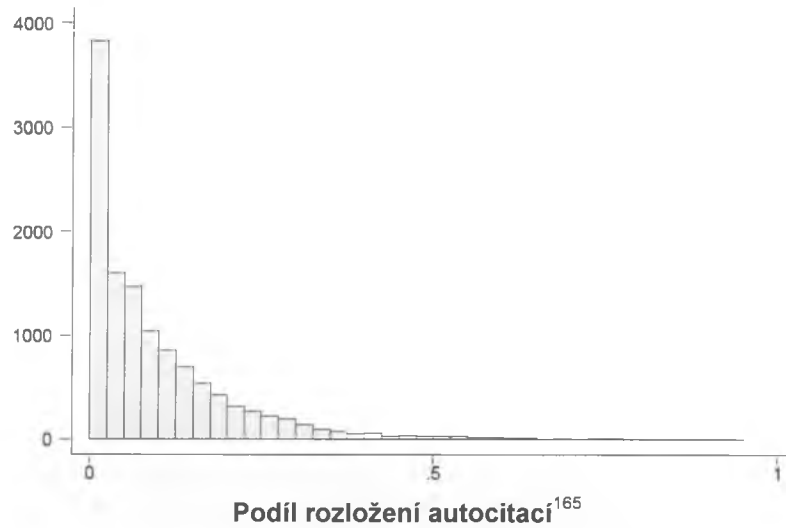
Levý sloupec – relativní hodnota

- budoucí podíl citací
- podíl citací v minulosti

¹⁶³ HALL, Bronwyn H.; JAFFE, Adam; TRAJTENBERG, Manuel. *Market value and patent citations: a first look.*

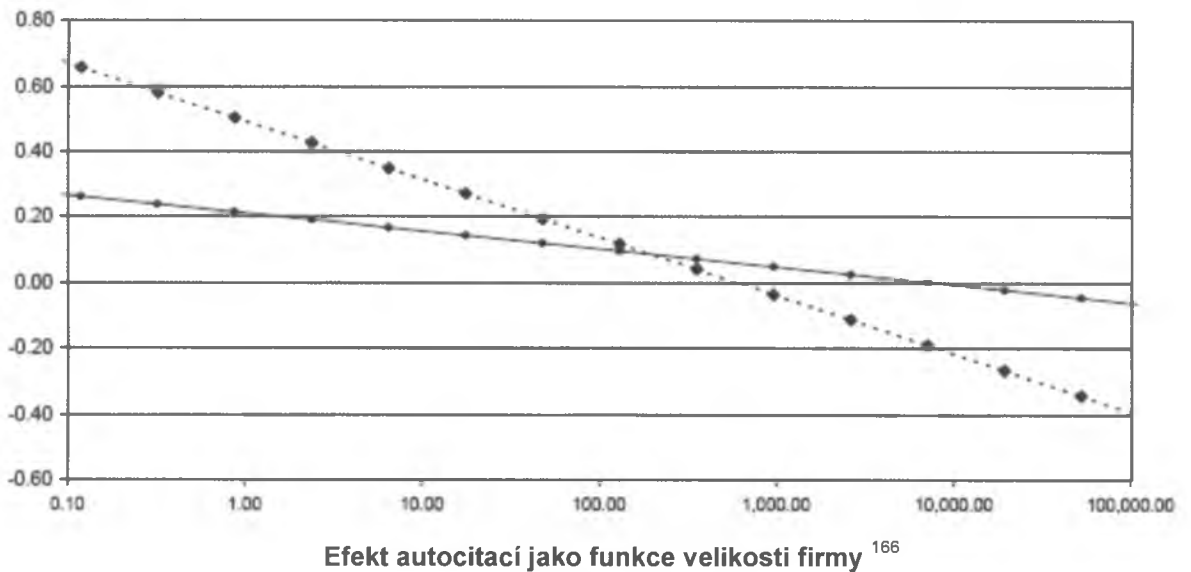
¹⁶⁴ HALL, Bronwyn H.; JAFFE, Adam; TRAJTENBERG, Manuel. *Market value and patent citations: a first look.*

—●— celkový podíl citací



Levý sloupec – frekvence citací

Dole – podíl autocitací



Levý sloupec – fiktivní koeficient

Dole – prodej v milionech dolarů

|-♦-|- fiktivní (autocitace = 0)

—●— koeficient vlastních patentů

¹⁶⁵ HALL, Bronwyn H.; JAFFE, Adam; TRAJTENBERG, Manuel. *Market value and patent citations: a first look.*

¹⁶⁶ HALL, Bronwyn H.; JAFFE, Adam; TRAJTENBERG, Manuel. *Market value and patent citations: a first look.*

Někteří kritici říkají, že díky malému počtu nepatentových citací v softwarových patentech, se neobjevuje jako samozřejmost tento druh citací a vynálezy nezískávají citace a tyto patenty by měly být brány jako neplatné. Navíc, „postavení patentování technologií“ je označeno velkým počtem patentů v technických oblastech minimální revolučním vynálezem.

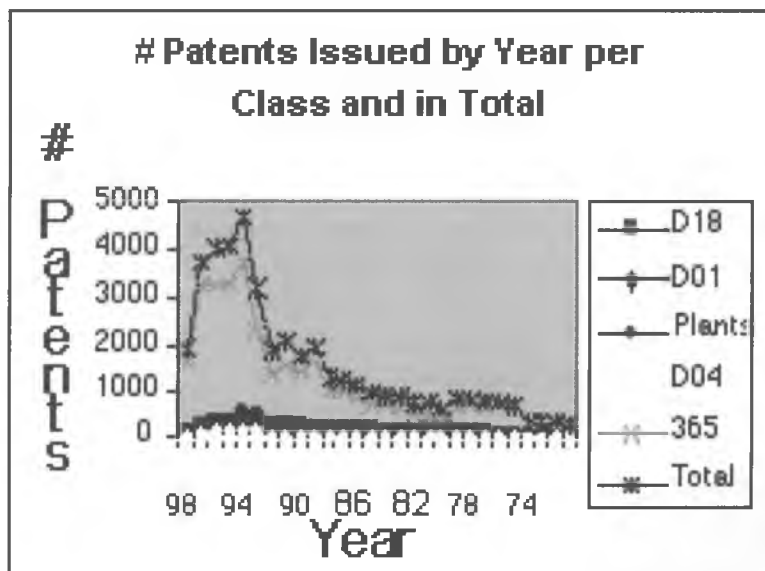
Studie *A Survey of Non-Patent References in Selected Patent Classes* se zabývá výzkumem praxe nepatentových citací v patentech za dobu posledních 27 let v oblastech netechnické povahy. Navíc, růst v počtu vydaných patentů a počtu citací jakýmkoliv dřívějším způsobem bylo odhadnuto. Byly porovnány hlavní kritizované oblasti softwarových patentů.

17.15. Zjištění

Softwarová data pro porovnání byla vzata z kolekce tříd, primárně třídy 364 a 395 jak je předložil Gregory Aharonian, Internet Patent News Service v *'1997 Software Patent Statistics - Feb. to July' and '1995 US Software Patent Statistics'*.

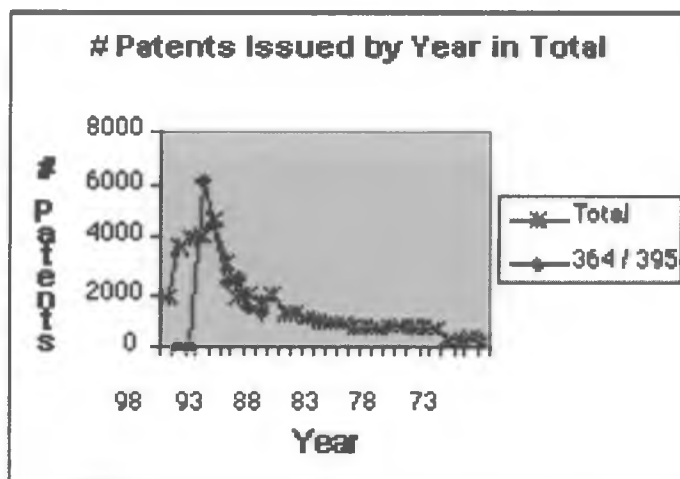
Počet patentů vydaných ročně

Ohromný růst v počtu patentů jako výsledek v softwarové aréně je poznamenán v člancích Aharoniana. Tento růst byl zaznamenán jako že nemá žádný vztah k minimální revoluční inovaci v softwarovém průmyslu. Když míra softwarového růstu je porovnána s růstem míry nalezené ve studiích tříd, je zřejmé, že míra růstu není spoutána s oblastí softwarových patentů, ale zdá se být v trendem podobným dalším třídám patentů. Některé studované třídy jsou diskutabilní v oblastech dokonce méně revolučních vynálezů.



Vydané patenty za rok na třídu a celkově ¹⁶⁷

¹⁶⁷ BLANK, Lynne. *A Survey of Non-Patent References in Selected Patent Classes*



Patenty vydané za rok celkově ¹⁶⁸

Ukazuje se, že existuje nárůst ve vydávání patentů ve všech studovaných třídách, zvláště na konci 80. let. Tento růst trochu zpomalil v 90. letech.

Tento trend může být zakořeněn v ekonomice spíše než ve vědě a vynálezech. Vynálezci, kteří vyvíjejí a komercializují produkty v konkurenčním prostředí trhu pravděpodobně chtějí ochránit jejich investice. Jednou z významných forem ochrany je patentování. Patent může zajistit hledisko na trhu nebo alespoň zdravé popředí startu komerčního úspěchu. Je docela pravděpodobné, že vzestup v počtu patentů není spojen s radostí pionýrského patentu, ale spíše radost ze získání peněz.

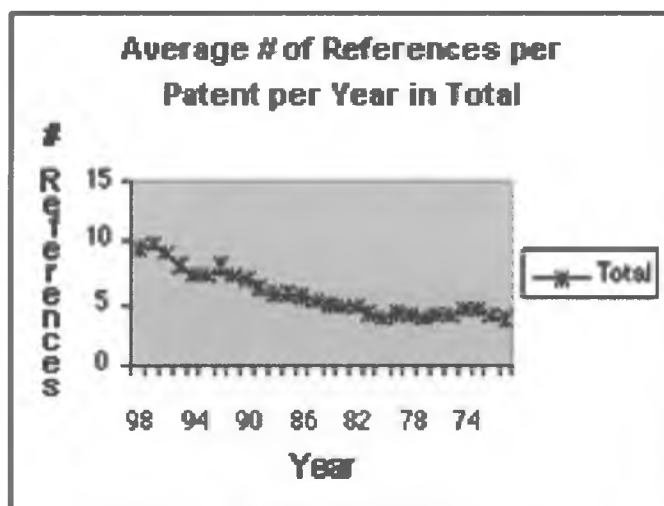
Počet citací citovaných na patent

Studované třídy poskytují průměrný počet citací citovaných na patent běžně v oblasti 9. Data indikují, že průměrná citace na patent byla zhruba 4 v roce 1972. A postupem času došlo k nárůstu průměrného počtu citací na patent.

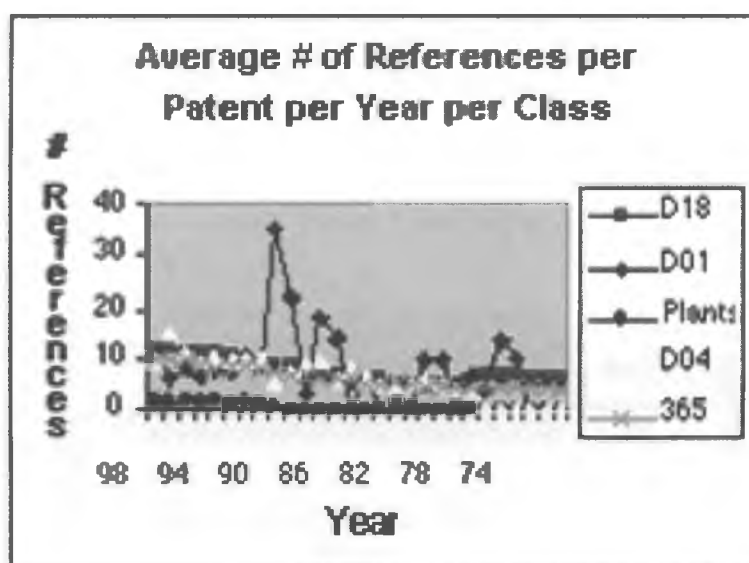
Tento výsledek není překvapivý. Informace o všech druzích jsou více dostupné díky Internetu a dalším formám elektronických a dalších medií. Tato míra zvýšení citací nemůže být zvýšena tak rychle jak by někteří žádali, ale kritika si musí vzpomenout, že informace nekalkulují. Mnoho obchodníků vytvářející patentování stále musí vydělávat. Nejsou požadovány všechny informace o tématu, právě to, což je materiální. Množství informací musí být řízeno a drženo v rovnováze, aby poskytovalo vhodný počet informací, které nebudou zabraňovat patentu zařazení že je příliš nákladný.

Ve skutečnosti, toto zvýšení průměrného počtu citací na patent může být indikátor DOSTATKU vynálezů v oblasti. To může znamenat, že právě průkopnické patenty by byly unikátně krátce citacemi dřívějších technik. Klesání míry citací nebo třídy s malými počty citací mohou být tzv. horkou postelí vynálezu.

¹⁶⁸ BLANK, Lynne. *A Survey of Non-Patent References in Selected Patent Classes*



Průměr citací na patent za rok celkově ¹⁶⁹

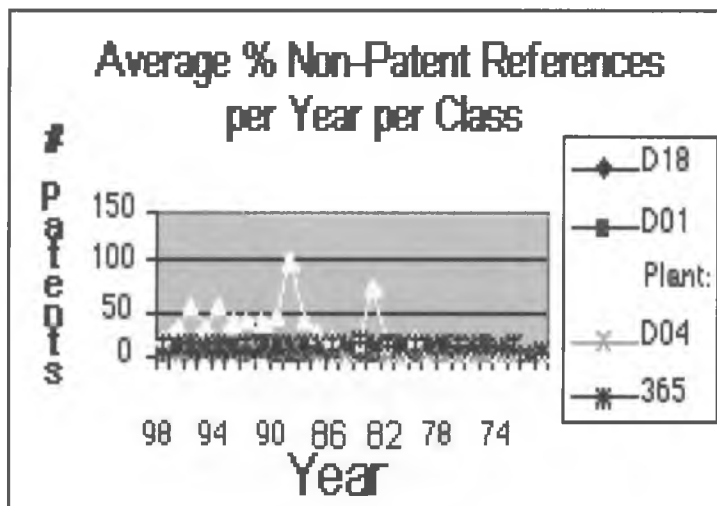


Průměr citací na patent na rok a třídu ¹⁷⁰

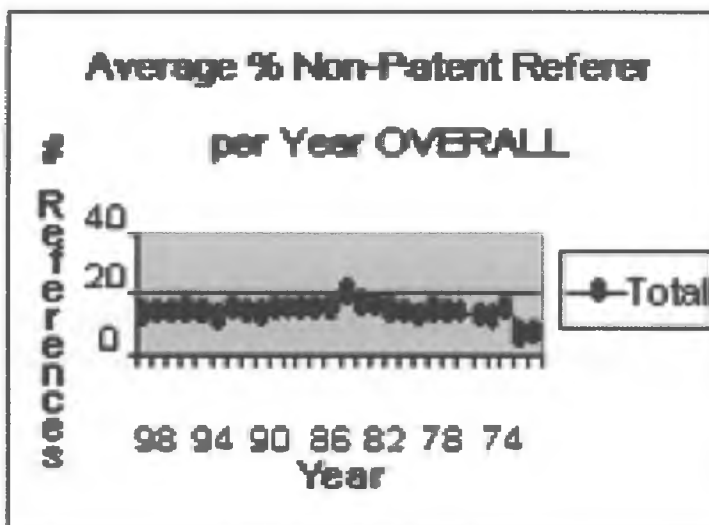
¹⁶⁹ BLANK, Lynne. *A Survey of Non-Patent References in Selected Patent Classes*

¹⁷⁰ BLANK, Lynne. *A Survey of Non-Patent References in Selected Patent Classes*

Procento citovaných nepatentových citací



Průměr procent nepatentových citací na rok a třídu ¹⁷¹



Průměr procent nepatentových citací na rok vřezahrnující ¹⁷²

Zajímavě je, že průměr nepatentových citací se zdá být relativně konstantní během období studie, pohybující se okolo 12-14%.

Aharoninovy články indikují, že průměrný počet nepatentových citací v softwarových patentech je 3.1. Jestliže to předpokládá, že % nepatentových citací dřívějších citací na softwarové třídy je konzistentní s daty ve studii, průměrný počet dřívějších technik citací v softwarových patentech je v průměru okolo 22 citací. Porovnání s patenty ve studii, to je neuvěřitelný počet – a pisatelé softwarových patentů by měli být pochváleni.

Studované třídy ve studii *A Survey of Non-Patent References in Selected Patent Classes* demonstrovali několik zajímavých bodů. Za prvé, existuje *tržní* zvýšení v počtu

¹⁷¹ BLANK, Lynne. *A Survey of Non-Patent References in Selected Patent Classes*

¹⁷² BLANK, Lynne. *A Survey of Non-Patent References in Selected Patent Classes*

patentů vydaných ročně. Poté nastal vrchol na konci 80. let, míra počtu citací se v 90. letech jeví jako klesající, ačkoliv více patentů je vydáno ročně nyní než v 70. letech a v první polovině 80. let. Za druhé, počet citací citovaných průměrně se zvyšuje v čase, z průměru 4 citací na patent na začátku 70. let k průměru 9 citací na patent v 90. letech. Konečně % citací citovaných, jenž jsou založeny na nepatentových dřívějších technikách se jeví jako relativně konstantní za dobu 27 let a to okolo 12-14%.¹⁷³

Platnost předpokladu: „dřívěji citované publikace a/nebo následné citující publikace budou mít vztah k tématu vynálezu, z hlediska vyhledávání technik patentového citování může být velmi hodnotné vyhledat dřívější technologie, zkontrolovat možná porušení, rozhodnutí průzkumníků a může to být cenné také pro competitive intelligence (informace o konkurenci).

Např. databázové centrum STN poskytuje možnost vyhledávat patentové citace v:

IFIPAT a USPATFULL pro vyhledávání amerických patentů,

PATOSEP a EUROPATFULL pro vyhledávání evropských publikací o patentech,

PATOSDE a PATDPA pro německé publikace

Také **Derwent Patent Citation Index (DPCI)** byl vytvořen speciálně pro tento typ vyhledávání.

Hlavní užití vyhledávání citací

Předešlé technologie

- zlepšení vyhledávání vyčerpávajících předmětů nebo vyhledávání kódování
- vyhledávání dokumentů, které se týkají zvláště relevantních dokumentů pod kontrolou

Competitive Intellingence

- jaké společnosti citují významnou skupinu patentů

Spory

- jaké společnosti citují portfolio naší společnosti
- setřídění současného povědomí

Rozhodnutí

- jaké citace byly citovány dalšími patentovými úřady
- užití odůvodnění nebo vyvrácení postoje

Derwent Patent Citation Index

- spuštěn 15. dubna 1995
- jméno souboru DPCI v STN

¹⁷³ BLANK, Lynne. *A Survey of Non-Patent References in Selected Patent Classes*

- obsahuje citované patenty
- pokrytí:
 - o týdně od roku 1994
 - o citace průzkumníků patentů
 - o retrospektiva od poloviny 70. let 20. století pro patenty americké, britské, evropské a světové
 - o zdroje od 16 patentových autorit od r. 1994
 - o citace od vynálezců (od roku 1994)

Geografické pokrytí

- Rakousko, Nizozemí
- Austrálie, Nový Zéland
- Belgie, Jižní Afrika
- Kanada, Švédsko
- Evropa, Švýcarsko
- Francie, Velká Británie,
- Německo, USA
- Japonsko, WIPO

Hlavní použití

- předešlé technologie
- právní spory
- Competitive Intelligence
- rozhodnutí

Užití

Předešlé technologie

- tradiční vyhledávání
- hodnocení
- vyhledání citací k „dobrým“ odpovědím v DPCI
- studium citovaných dokumentů, možno přestěhování počtu citovaných dokumentů zpět do WPINDEX/WPDS pro další hodnocení

=> FILE WPIDS

COST IN U.S. DOLLARS	SINCE FILE ENTRY	TOTAL SESSION
FULL ESTIMATED COST	0.21	0.21

FILE 'WPIDS' ENTERED AT 12:31:52 ON 16 JAN 1998
COPYRIGHT (C) 1998 DERWENT INFORMATION LTD

FILE LAST UPDATED: 12 JAN 1998 <19980112/UP>
>>>UPDATE WEEKS:
MOST RECENT DERWENT WEEK 199802 <199802/DW>
DERWENT WEEK FOR CHEMICAL CODING: 199748
DERWENT WEEK FOR POLYMER INDEXING: 199750
DERWENT WORLD PATENTS INDEX SUBSCRIBER FILE, COVERS 1963 TO DATE
>>> D COST AND SET NOTICE DO NOT REFLECT SUBSCRIBER DISCOUNTS -
SEE HELP COST FOR DETAILS <<<
>>> CHANGES TO DWPI COVERAGE - SEE NEWS <<<

=> 8 MICROWAVE(W) (COOKER OR OVEN) AND HUMID?

38882 MICROWAVE
8058 COOKER
29789 OVEN
6300 MICROWAVE (W) (COOKER OR OVEN)
39259 HUMID?
L1 124 MICROWAVE (W) (COOKER OR OVEN) AND HUMID?

=> SELECT L1 AN 1-

E1 THROUGH E124 ASSIGNED

=> FILE DPCI

COST IN U.S. DOLLARS	SINCE FILE	TOTAL
FULL ESTIMATED COST	ENTRY	SESSION
	5.10	5.31

FILE 'DPCI' ENTERED AT 12:33:16 ON 16 JAN 1998
COPYRIGHT (C) 1998 DERWENT INFORMATION LTD

FILE LAST UPDATED: 13 JAN 1998 <19980113/UP>
MOST RECENT DERWENT DPCI WEEK 199801
PATENTS CITATION INDEX, COVERS 1973 TO DATE
>>> LEARNING FILE LDPCI AVAILABLE <<<

=> S E1-E124

1 71-68662S/AN
1 73-51849U/AN
1 76-L8369X/AN
1 77-H2126Y/AN
1 79-A8405B/AN

**Note that the AN's are identical
between WPINDEX and DPCI.**

L2 66 (71-68662S/AN OR 73-51849U/AN OR 76-L8369X/AN OR 77-H2126Y/AN
OR 79-A8405B/AN OR 79-C0735B/AN OR 79-C0736B/AN. . .

=> S 94-287542/AN

L3 1 94-287542/AN

**The first portion of a DPCI record
contains the same bibliographic and
patent family information as WPI.**

=> D L3 ALL

L3 ANSWER 1 OF 1 DPCI COPYRIGHT 1998 DERWENT INFORMATION LTD
AN ***94-287542*** [36] DPCI
DNN N94-226452
TI Automatic thawing device for microwave oven - controls output level
of electromagnetic wave of magnetron in response to output signal of
temperature, humidity or gas sensor to automatically and
appropriately thaw frozen food.
DC Q74 T01 X25 X27
IN CHAI, E S; GONG, C S; KIM, S T
PA (GLDS) GOLDSTAR CO LTD; (GLDS) LG ELECTRONICS INC; (LGEL-N) LG
ELECTRONICS INC; (GLDS) KINSEISHA KK
CYC 10
PI EP-616487 A2 940921 (9436)* EN 13 pp H05B-006-68
R: DE ES FR GB SE
BR9401193 A 941025 (9444) G05B-013-02
CA2115406 A 940920 (9444) F24C-007-02
JP06300267 A 941028 (9503) 9 pp F24C-007-02
EP-616487 A3 941012 (9533)
US5436433 A 950725 (9535) 13 pp H05B-006-68
CN1094805 A 941109 (9544) F24C-007-02
ADT EP-616487 A2 94EP-0400420 940228; BR9401193 A 94BR-0001193 940317;
CA2115406 A 94CA-2115406 940210; JP06300267 A 94JP-0042404 940314;
US5436433 A 94US-0177864 940104; CN1094805 A 94CN-0102948 940318

- AN je identické mezi WPINDEX a DPCI

- První část záznamu DPCI obsahuje stejné bibliografické informace a informace o patentové rodině jako WPI

PRAI 93KR-0015287 930806; 93KR-0004226 930319
IC ICM F24C-007-02; G05B-013-02; H05B-006-68
ICS F24C-007-08; H05B-006-80
FS EPI GMPI

EXF EXAMINER'S FIELD OF SEARCH UPE: 960611

NCL US5436433 A 950725
000/009.325; 219/703; 219/707;

Citation counters may give a hint as to the importance of the invention. E. g. landmark inventions may have been cited by many subsequent citing publications.

IC EP-616487 A3 941012
H05B

CTCS CITATION COUNTERS

PNC.DI	2	Cited Patents Count (by inventor)
PNC.DX	9	Cited Patents Count (by examiner)
IAC.DI	1	Cited Issuing Authority Count (by inventor)
IAC.DX	3	Cited Issuing Authority Count (by examiner)
PNC.GI	0	Citing Patents Count (by inventor)
PNC.GX	2	Citing Patents Count (by examiner)
IAC.GI	0	Citing Issuing Authority Count (by inventor)
IAC.GX	1	Citing Issuing Authority Count (by examiner)
CRC.I	0	Cited Literature References Count (by inventor)
CRC.X	0	Cited Literature References Count (by examiner)

CDP CITED PATENTS UPD: 960611

Cited by Inventor

CITING PATENT	CITED PATENT	ACCNO
EP-616487 A2	KR-896080 KR-921987	

Cited by Examiner

CITING PATENT	CAT	CITED PATENT	ACCNO
EP-616487	A3 X	EP-166997 A	86-009003/02
		PA: (MATU) MATSUSHITA ELEC IND CO LTD	
		IN: UEDA, S	
	X	EP-268329 A	88-141803/21
		PA: (PHIG) PHILIPS GLOEILAMPENFAB NV; (IREI) IRE IND	
		RIUNITE EUROMESTICI; (PHIG) PHILIPS NORDEN AB	
		IN: FIOROLI, M; IDEBRO, M G; KOKKELER, F; RISMAN, P	
		O G; VIGANO, L; RISMAN, P	

The CITED PATENTS section provides earlier cited patents for each citing family member of this record.

You can search for a PN in the CITED PATENTS section (eg => S US5123456/PN.D or /RPN) to retrieve newer citing records.

- počty citací mohou poskytnout nápovědu jako důležitou část vynálezu. Např. vynálezy mohou být citovány mnoha následnými citujícími publikacemi
- **CITED PATENTS** (Citované patenty) poskytuje dřívěji citované patenty pro každého člena citující rodiny ze záznamu
- Můžeme vyhledávat pro PN v sekci **CITED PATENTS** (eg => S US5123456/PN.D or /RPN) vyhledáním nověji citujícího záznamu.

US5436433 A EP-166997 A 86-009003/02
 PA: (MATU) MATSUSHITA ELEC IND CO LTD
 IN: UEDA, S
 EP-268329 A 88-141803/21
 PA: (PHIG) PHILIPS GLOEILAMPENFAB NV; (IREI) IRE IND
 RIUNITE EURODOMESTICI; (PHIG) PHILIPS NORDEN AB
 IN: FIOROLI, M; IDEBRO, M G; KOKKELER, F; RISMAN, P
 O G; VIGANO, L; RISMAN, P
 JP63226541 A 88-310193/44
 PA: (MITQ) MITSUBISHI DENKI KR; (MITS-N) MITSUBISHI
 DENKI MOMUKI
 US4255639 A 81-C8623D/13
 PA: (SHAF) SHARP KK
 IN: KAWABATA, T; MAKITA, M; MASUZAWA, S
 US4453066 A 83-B8349K/06
 PA: (MATU) MATSUSHITA ELEC IND CO LTD
 IN: MORI, F
 US4599503 A 86-009003/02
 PA: (MATU) MATSUSHITA ELEC IND CO LTD
 IN: UEDA, S
 US4841111 A 88-141803/21
 PA: (PHIG) PHILIPS GLOEILAMPENFAB NV; (IREI) IRE IND
 RIUNITE EURODOMESTICI; (PHIG) PHILIPS NORDEN AB
 IN: FIOROLI, M; IDEBRO, M G; KOKKELER, F; RISMAN, P
 O G; VIGANO, L; RISMAN, P
 US4874914 A 89-229280/32
 PA: (MICR-N) MICROWAVE OVENS LTD
 IN: EKE, K I
 US4998001 A 90-092918/13
 PA: (IREI-N) IRE IND RIUNITE EUR; (WHIR) WHIRLPOOL
 INT BV; (IREI) IRE IND RIUNITE EURODOMESTICI
 IN: CIGARINI, E; CREPALDI, L; KOKKELER, F; ORLANDI,
 G

CGP CITING PATENTS

UPG: 961216

Conversely, the CITING PATENTS section provides newer citing patents for each of the cited family members in this record.

Cited by Examiner

CITED PATENT	CAT	CITING PATENT	ACCNO
US5436433	A	US5545881	A 96-279946/29
		PA: (GLDS) LG ELECTRONICS INC; (GLDS) KINSEISHA KK	
		IN: CHAI, E S; LEE, K H	
		US5552584	A 96-032111/04
		PA: (WHIR) WHIRLPOOL EURO BV; (WHIR) WHIRLPOOL CORP	
		IN: IDEBRO, M G	

Naopak, **CITING PATENTS** (citující patenty) poskytují nověji citované patenty pro každého ze členů citované rodiny v záznamu.

Užití

Zasahování do práv do kontroly

Pokud potřebujeme vědět, zda zasahování co práv se může objevit u více významných patentů v daném portfoliu, užívá se DPCI jako následující:

- vyhledání patentů dle zájmu
- vytištění citujících patentů a jejich studium
- založení SDI služby pro monitorování všech patentů týdně

Příklad

=> FILE DPCI

=> S EP 415569/PN

L1 1 EP 415569/PN
(EP415569/PN)

=> D L1 ALL

AN 91-067136 [10] DPCI
DNC C91-028370
TI Agrochemical compsns. for use on water - contg. solid carrier,
active agent and low-density oil.
DC C03
IN BIRD, N R; RAMSAY, G; WARRINGTON, R P
PA (ICIL) IMPERIAL CHEM IND PLC; (WARR-I) WARRINGTON R P; (ZENE)
ZENECA LTD

CYC 22

PI EP 415569 A 910306 (9110)*
R: AT BE CH DE ES FR GB GR IT LI LU NL SE

AU 9060220 A 910307 (9117)

PT 95163 A 910522 (9124)

CA 2023547 A 910302 (9131)

BR 9003989 A 910903 (9140)

JP 03193705 A 910823 (9140)

CN 1049771 A 910313 (9146)

HU 57536 T 911230 (9206)

EP 415569 A3 921028 (9341)

AU 642151 B 931014 (9348)

IL 95305 A 941128 (9504)

US 5466693 A 951114 (9551)

Individual PN's or AN's numbers may be searched in DPCI, or a large set of records may be crossed into DPCI using SmartSELECT. E. g. => S L# <AN> where L# is an answer set in WPINDEX.

A01N025-04
A01N025-12
6 pp A01N043-54

- individuální PN nebo AN čísla mohou být prohledány v DPCI nebo větší sady záznamů mohou být přeneseny do SmartSELECT. Např. => S L# <AN> kde L# je odpověď sady v WPINDEX

ADT EP 415569 A EP 90-308478 900801; JP 03193705 A JP 90-228511 900831;
 EP 415569 A3 EP 90-308478 900801; AU 642151 B AU 90-60220 900806; IL
 95305 A IL 90-95305 900806; US 5466693 A Cont of US 90-570588
 900821, US 93-125702 930923
 FDT AU 642151 B Previous Publ. AU 9060220
 PRAI GB 89-19833 890901; GB 90-7995 900409; GB 90-16783 900731
 IC A01N025-12; A01N043-54
 ICM A01N025-04; A01N025-12; A01N043-54
 ICS A01N025-14
 FS CPI

EXF EXAMINER'S FIELD OF SEARCH UPE: 960816

NCL US 5466693 A 951114
 424/409; 424/502; 514/269; 514/951; 514/952

CTCS CITATION COUNTERS

PNC.DI	0	Cited Patents Count (by inventor)
PNC.DX	20	Cited Patents Count (by examiner)
IAC.DI	0	Cited Issuing Authority Count (by inventor)
IAC.DX	7	Cited Issuing Authority Count (by examiner)
PNC.GI	3	Citing Patents Count (by inventor)
PNC.GX	8	Citing Patents Count (by examiner)
IAC.GI	2	Citing Issuing Authority Count (by inventor)
IAC.GX	3	Citing Issuing Authority Count (by examiner)
CRC.I	0	Cited Literature References Count (by inventor)
CRC.X	11	Cited Literature References Count (by examiner)

CDP CITED PATENTS UPD: 960816

Cited by Examiner

CITING PATENT	CAT	CITED PATENT	ACCNO
EP 415569	A	No Citations	
EP 415569	A3	EP 113857	84-159432/26
		PA: (FARB) BAYER AG	
		IN: HAUSMANN, H; SCHMIDT, R R; VOEGE, H	
		GB 2011788	79-53121B/29
		PA: (ELIL) LILLY & CO ELI; (TAKE) TAKEDA CHEM IND	
		LTD	
		IN: KAWAKAMI, T; MATSUURA, K; OSUGI, T	
		GB 2184946	87-187074/27
		PA: (NOVO) NOVO IND AS	
		IN: BRDBURY, R S; MOST, B H; QUINLAN, R J	
		JP 2025403	90-071763/10
		PA: (AMCY) AMERICAN CYANAMID CO; (KUNI-N) KUNIMINE	
		KOGYO KK	

**These earlier
 patents were
 cited by
 EP415569**

US 5466693 A

**This US equivalent
to the EP patent
cited these earlier
patents.**

JP 62175409 87-253855/36
PA: (HOKK) HOKKO CHEM IND CO LTD;
JP 62198602 87-286910/41

PA: (SANQ) SANKEI CHEMICAL KK
JP 63017802 88-061248/09

PA: (MITK) MITSUI TOATSU CHEM INC;
AU 7166581
CA 2008701 A 90-248282/33

PA: (ICIL) IMPERIAL CHEM IND PLC; (ZENE) ZENECA LTD;
(CLOU-I) CLOUGH J M; (ICIL) ICI PLC
IN: CLOUGH, J M; GODFREY, A; RICHARD, C; CHEETHAM,
R; GODFREY, C R A; STREETING, I T
EP 113857 A 84-159432/26

PA: (FARB) BAYER AG
IN: HAUSMANN, H; SCHMIDT, R R; VOEGE, H
EP 189377 A 86-198665/31

PA: (TAKE) TAKEDA CHEM IND LTD;
IN: ASAKA, A; GOTOU, Y; OBAYASHI, H; TANABAYSH, C
EP 206987 A 86-341264/52

PA: (CIBA) CIBA GEIGY AG
IN: BESSE, J J; SOMLO, J
EP 248554 A 87-343475/49

PA: (ICIL) IMPERIAL CHEM IND PLC
IN: BARTON, J E D; COLLINS, D J; SLATER, J W
FR 2611435 A 88-301472/43

PA: (OPIT-N) OPITNA STANCIA POMO;
IN: ALEXIEV, A B; MILUSHEVA, I R
FR 2611436 A 88-301473/43

PA: (OPIT-N) OPITNA STANCIA POMO
IN: ALEXIEV, A B; MILUSHEVA, I R
GB 2184946 A 87-187074/27

PA: (NOVO) NOVO IND AS
IN: BRDBURY, R S; MOST, B H; QUINLAN, R J
JP 8817802
US 4080191 A 78-31876A/17

PA: (SANO) SANDOZ LTD
IN: HARVEY, J C
US 4886656 A 86-198665/31

PA: (TAKE) TAKEDA CHEM IND LTD
IN: ASAKA, A; GOTOU, Y; OBAYASHI, H; TANABAYSH, C
US 5145856 A 90-248282/33

PA: (ICIL) IMPERIAL CHEM IND PLC; (ZENE) ZENECA LTD;
(CLOU-I) CLOUGH J M; (ICIL) ICI PLC
IN: CLOUGH, J M; GODFREY, A; RICHARD, C; CHEETHAM,
R; GODFREY, C R A; STREETING, I T

- americké ekvivalentní patenty k evropským citované těmito dřívějšími patenty

Citations by Examiner

CITING PATENT	CAT	CITED LITERATURE
US 5466693	A	The Merck Index, 10th Edition, Windholz, editor, Merck & Co., Inc, Rahway N.J., 1983, p. 980
US 5466693	A	J55141401 (Nihon Noyaku KK) (Abstract only), 20/04/1979
US 5466693	A	J55154902 (Nippon Toketsu Kans) (Abstract only) 23/04/1979
US 5466693	A	J56030901 (Nihon Noyaku KK) (Abstract only), 23/08/1979
US 5466693	A	J62188602 (Sankei Chemicals Company Limited) (Abstract only) 02/09/1987
US 5466693	A	J63041401 (Higashimura et al.) (Abstract only) 22/02/1988
US 5466693	A	J73015614 (Tokyo Organic Chemical Inds. Ltd.) (Abstract only) 14/02/1970
US 5466693	A	Chem. Abstracts 112:32163c (Walcerz et al.), 1990
US 5466693	A	Chem. Abstracts 112:50640k (Puritch et al.), 1990
US 5466693	A	Chem. Abstracts 112:50643p (Idemitsu Kosan Co., Ltd.), 1990
US 5466693	A	Chem. Abstracts 103:208952q (Sekiguchi et al.), 1985

CGP CITING PATENTS UPG: 971009

These newer citing patents, e.g. WO9500303, (they cited EP415569) may be close in art and may need to evaluated.

Cited by Inventor

CITED PATENT	CITING PATENT	ACCNO
EP 415569	A	DE 4320495 A1 95-031116/05 PA: (FARB) BAYER AG IN: BUSCHHAUS, H; HEUER, L; KUGLER, M; KUNISCH, F; SCHRAGE, H
		DE 4409040 A1 95-329214/43 PA: (FARB) BAYER AG IN: BUSCHHAUS, H; KUGLER, M; KUNISCH, F; UHR, H
		WO 9500303 A1 95-051838/07 PA: (FARB) BAYER AG IN: BUSCHHAUS, H; HEUER, L; KUGLER, M; KUNISCH, F; SCHRAGE, H

Cited by Examiner

CITED PATENT	CAT	CITING PATENT	ACCNO
CA 2023547	A	US 5580544 A	97-041723/04
		PA: (USRU) UNIROYAL CHEM CO INC; (USRU) UNIROYAL CHEM LTD	
		IN: DAO, D C; DRENNAN, C H; HALLATT, W L; HIBBETT, H L	
EP 415569		EP 499823 A3	92-286002/35
		PA: (BADI) BASF AG; (GRAM-I) GRAMMENOS W	
		IN: AMMERMANN, E; DOETZER, R; HARREUS, A; HELLEND AHL, B; LORENZ, G; SAUTER, H; GRAMMENOS, W	
		US 5254717 A	92-286002/35
		PA: (BADI) BASF AG; (GRAM-I) GRAMMENOS W	
		IN: AMMERMANN, E; DOETZER, R; HARREUS, A; HELLEND AHL, B; LORENZ, G; SAUTER, H; GRAMMENOS, W	
		WO 9322921 A1	93-386097/48
		PA: (ZENE) ZENECA LTD	
		IN: BROWNELL, K H; GODWIN, J R; HEANEY, S P; SKIDMORE, A M; BROWNE L, K H	
EP 415569	A	EP 499823 B1	92-286002/35
		PA: (BADI) BASF AG; (GRAM-I) GRAMMENOS W	
		IN: AMMERMANN, E; DOETZER, R; HARREUS, A; HELLEND AHL, B; LORENZ, G; SAUTER, H; GRAMMENOS, W	
		EP 641163 B1	93-386097/48
		PA: (ZENE) ZENECA LTD	
		IN: BROWNELL, K H; GODWIN, J R; HEANEY, S P; SKIDMORE, A M; BROWNE L, K H	
		PX EP 684225 A1	96-000972/01
		PA: (BADI) BASF AG;	
		IN: AMMERMANN, E; DOETZER, R; GRAMMENOS, W; HARREUS, A; HELLEND AHL, B; LORENZ, G; SAUTER, H	
		EP 684225 B1	96-000972/01
		PA: (BADI) BASF AG	
		IN: AMMERMANN, E; DOETZER, R; GRAMMENOS, W; HARREUS, A; HELLEND AHL, B; LORENZ, G; SAUTER, H	

Více o nověji citujících patentech je možné zjistit vyhledáním citujících patentů v World Patents Index

FILE 'WPIDS' ENTERED AT 09:52:15 ON 23 FEB 1998
COPYRIGHT (C) 1998 DERWENT INFORMATION LTD

=> S WO 9500303/PN

L2 1 WO 9500303/PN
(WO9500303/PN)

=> D L2 ALL

L2 ANSWER 1 OF 1 WPIDS COPYRIGHT 1998 DERWENT

AN 95-051838 [07] WPIDS

CR 95-031116 [05]

DNN N95-040732 DNC C95-023730

TI Wood preservative, esp. for fungus and insect control - comprises synergistic mixt. of hexaconazole and/or metconazole with insecticide and opt. other azole fungicide.

DC C02 C03 D22 E13 E19 F09 G02 P63

IN BUSCHHAUS, H; HEUER, L; KUGLER, M; KUNISCH, F; SCHRAGE, H

PA (FARB) BAYER AG

CYC 40

PI WO 9500303 A1 950105 (9507)* DE 29 pp B27K003-50 <--

RW: AT BE CH DE DK ES FR GB GR IE IT LU MC NL OA PT SE

W: AU BB BG BR BY CA CN CZ FI HU JP KR KZ LK NO NZ PL RO RU SK

UA US

AU 9471231 A 950117 (9521) B27K003-50

DE 4406819 A1 950907 (9541) 8 pp B27K003-34

FI 9506113 A 951219 (9611) B27K000-00

NO 9505107 A 951215 (9615) B27K003-50

EP 705160 A1 960410 (9619) DE B27K003-50

R: AT BE CH DE DK ES FR GB IT LI NL PT SE

BR 9407120 A 960903 (9641) B27K003-50

JP 08509437 W 961008 (9705) 28 pp B27K003-50

ADT WO 9500303 A1 WO 94-EP1868 940608; AU 9471231 A AU 94-71231 940608;

DE 4406819 A1 DE 94-4406819 940302; FI 9506113 A WO 94-EP1868

940608, FI 95-6113 951219; NO 9505107 A WO 94-EP1868 940608, NO

95-5107 951215; EP 705160 A1 EP 94-920437 940608, WO 94-EP1868

940608; BR 9407120 A BR 94-7120 940608, WO 94-EP1868 940608; JP

08509437 W WO 94-EP1868 940608, JP 95-502383 940608

FDT AU 9471231 A Based on WO 9500303; EP 705160 A1 Based on WO 9500303;

BR 9407120 A Based on WO 9500303; JP 08509437 W Based on WO 9500303

PRAI DE 94-4406819 940302; DE 93-4320495 930621

REP 2.Jnl.Ref ; EP 385076; EP 458060; EP 514644; EP 533017; EP 591764;

WO 9302557; WO 9406293

IC ICM B27K000-00; B27K003-34; B27K003-50

ICS A01N043-653; C09D005-14

To crossover all the citing patents automatically just select the PN.G and re-search in the PN index in WPIDS, or use SmartSELECT to do this in one step, => S L1 <PN.G>/PN (you have extracted citing patent numbers, PN.G, and re-searched them in the PN field).

AB WO 9500303 A UPAB: 971125
 Wood preservative contains: a) alpha -butyl- alpha
 -(2,4-dichlorophenyl)-1H-1,2,4-triazole-1-ethanol (hexaconazole) (I)
 and/or 5-(4-chlorophenyl)methyl)-2,2-dimethyl-1-
 (1H-1,2,4-triazol-1-ylmethyl)cyclopentanol (metconazole) (II); b) at
 least 1 synergistic insecticide; and opt. c) at least 1 additional
 azole fungicide.
 (I) and (II) as well as component c) may be present as addn.
 salt or metal salt complex.
 USE - Preservative is esp. effective against slime fungi,
 wood-discolouring and -destroying fungi, e.g. Ascomycetes
 Deuteromycetes, Zygomycetes and Basidiomycetes, and insects, e.g.
 Hymenoptera, termites and partic. long-horned beetles (Cerambycidae,
 Lyctidae, Bostrychidae and Anobiidae). It may be formulated as a
 finished agent or as a concentrate and may be used to protect wood
 and wood prods. such as timber used in construction, e.g. rafters,
 and joinery, railway sleepers, bridge members, landing stages,
 wooden vehicles, crates, pallets, containers, telegraph poles, wood
 cladding, fences, gates and chipboard. Application is pref. by
 impregnation e.g. using vacuum or pressure.
 ADVANTAGE - Preservative is more effective than known
 microbicides. It has a broad activity spectrum, good stability and
 high long term activity, and penetrates well into wood. Further, the
 activity of the fungicide is not negatively influenced by the
 insecticide.
 Dwg.0/0
 FS CPI GMPI
 FA AB; DCN
 MC CPI: C04-A07C; C05-B01M; C05-B01N; C10-H02F; C14-A06; C14-B04B;
 C14-S09; D09-A01; E07-D13C; F05-B01

Citace obrázku viz ¹⁷⁵

Derwent Patent Citation Index (DPCI) je vytvářený Thomson Derwent a poskytuje přístup k 54.5 milionům patentů a literatuře citací nalezených v 5.5 milionech patentových rodin. Citace pocházejí přibližně z 20 000 – 30 000 dokumentů od 6 hlavních patentových autorit.

Každý záznam v databázi popisuje patentovou rodinu pro jednotlivý vynález. Patentové citace odkazované průzkumníkem v patentovém dokumentu jsou nazývány „citované“ patenty („cited“ patents) v záznamu DPCI. Pokud citace odkazuje na starší vynález/patent, je to také přidáno do staršího záznamu rodiny jako „citující“ patent („citing“ patent). Záznam DPCI poskytuje pohled retrospektivní technologie pro vynález (citovaný patent) a jeho vliv na pozdější technologii (citující patenty).

Citace jsou poskytovány pro členy rodiny (základní a ekvivalentní) přidané k souboru z následujících patentů vydávající autority: European Patent Office (EP), Německo (DE), Japonsko (JP) (jen chemické a elektrické patenty); Patent Cooperation Treaty (PCTWO), Velká Británie (GB) a Spojené státy Americké (US).

Za období pokryté aktualizacemi Derwent 9418 až 9719 DPCI obsahuje citovaná patentová data z Rakouska, Austrálie, Belgie, Kanady, Francie, Nizozemí, Nového Zélandu, Jižní Afriky, Švédska a Švýcarska. Autorské citace získané ze „základního“ patentu v každé rodině jsou zahrnuty. Tato data zůstávají v souboru a jsou prohledatelná.

¹⁷⁵ Case study: patent citations

V rámci DIALOG vyhledávání tzv. Onesearch (což je vyhledávání, které umožňuje vyhledávat najednou více databázi v jednom balíčku (např. všechny medicínské databáze) se používá:

17.16. PATCITES – vyhledává soubory s patentovými citacemi

PATENTS – vyhledává patenty

Vzorové záznamy:

DIALOG(R) File 342:Derwent Patents Citation Index

(c) 1998 Derwent Info Ltd. All rts. reserv.

AX=, DX=00543555 WPI Acc No: 88-355270/50

/TI Electro-optical voltage detector involving slit or fibre - deduces voltage

at apex of prism from relative deflection of transmitted beam with spatial

intensity distribution

CK=, CO=, PA= Patent Assignee: (HAMM) HAMAMATSU PHOTONICS KK

AU=, IV= Author (Inventor): TAKAHASHI H; AOSHIMA S; NAKAMURA T; TSUCHIYA Y

Patent Family:

Patent No Kind Date Examiner Field of Search

PC=, PN=, PD= EP 294815 A 881214 (BASIC)

DE 3887006 G 940224

EP 294815 B1 940112

US 4982152 A 910101

US 5164667 A 921117

DW=Derwent Week (Basic): 8850

AC=, AN=, AD= Priority Data: JP 87144982 (870610)

AC=, AN=, AD= Applications: US 201205 (880602); DE 3887006 (880609); EP 88109229

(880609); US 619913 (901130)

Designated States

DS= (Regional): DE; GB

DC=Derwent Class: S01; U11; V07

IC=Int Pat Class: G01R-001/067; G01R-015/07

NP=Number of Patents: 005

NC=Number of Countries: 003

NT=Number of Cited Patents: 004

NR=Number of Cited Literature References: 001

NG=Number of Citing Patents: 004

Citované patenty

Family Member Cited Patent Cat WPI Acc No Assignee/Inventor

By Examiner:

PC=, PN= US 4982152 A US 3447855 A

CT= US 4982152 A US 4618819 A 86-298431/45 (UYRP) UNIV ROCHESTER/
MOUROU G; VALDMANIS J A

RX= US 5164667 A US 4631402 A 84-208333/34 (HITA) HITACHI LTD;

CO= (HITD) HITACHI CABLE LTD/NAGATSUMA K; TAKAGI K;
MATSUMURA H

US 5164667 A US 4683421 A 87-030927/05 (WESE) WESTINGHOUSE

IV= ELECTRIC CORP/MILLER R C; ASARS J A

Odkazy Citovaná literatura

Family Member Cat Citation

By Examiner:

PC=, PN= US 5164667 A Valdmanis et al. "Subpicosecond Electrooptic

RF= Sampling: Principles and Applications", IEEE Journal
of Quantum Electronics, vol. QE-22, No. 1, Jan.
1986.

Citující patenty

Family Member Citing Patent Cat WPI Acc No Assignee/Inventor

By Examiner:

PC=, PN= EP 294815 A EP 616225 A3 A 94-287331/36 (HAMM) HAMAMATSU

IV= PHOTONICS KK/TAKAHASHI H; WAKAMORI K

CG= US 4982152 A US 5331468 A 94-169429/21 (EAST) EASTMAN KODAK
CO/

NOETHEN M L

RI= US 4982152 A WO 9701100 A1 X 97-087490/08 (ALLM) ASEA BROWN
BOVERI

CO= AB; (UNSY) UNIV SYDNEY/BJARME M; BASSETT I M

RX= US 5164667 A WO 9701100 A1 X 97-087490/08 (ALLM) ASEA BROWN
BOVERI

AB; (UNSY) UNIV SYDNEY/BJARME M; BASSETT I M

Citace obrázku viz poznámka ¹⁷⁶

Příklad citací patentů

c) 1996 Derwent Info Ltd. All rts. reserv.

01511747 WPI Acc No: 80-21930C/12

Water-based print paste compsn. for carpets and fabrics - contains hydroxyalkyl
carboxyalkyl cellulose as the thickening agent

Patent Assignee: (UNIC) UNION CARBIDE CORP

Author (Inventor): KURYLA W C; RAINES R H

Patent Family:

Search	Patent No	Kind	Date	Examiner Field of
	US 4192647	A	800311 (BASIC)	

Derwent Week (Basic): 8012

Number of Patents: 001

Number of Countries: 001

Number of Cited Patents: 002

Number of Cited Literature References: 001

Number of Citing Patents:003

CITED PATENTS

Family Member	Cited Patent	Cat	WPI Acc No	Assignee/Inventor
---------------	--------------	-----	------------	-------------------

By Examiner:

US 4192647 A	US 2361454	A		
--------------	------------	---	--	--

US 4192647 A	US 4010038	A	75-74382W/45	(KANZ) KANZAKI PAPER FG
--------------	------------	---	--------------	-----------------------------

CO LTD

¹⁷⁶ Získáno z www.dialog.com Bluesheets

CITED LITERATURE REFERENCES

Family Member	Cat	Citation
---------------	-----	----------

By Examiner:

US 4192647 A	A	Chemical Abstracts, vol. 84, 1976, pp. 99, 100, 519 & 1371cs, and vol. 73, 1970, p. 74.
--------------	---	---

CITING PATENTS

Family Member	Citing Patent	Cat	WPI Acc No	Assignee/Inventor
---------------	---------------	-----	------------	-------------------

By Examiner:

US 4192647 A	US 4286955	A	80-88362C/50 (ARMS)	ARMSTRONG CORK CO/LEWIS R D
--------------	------------	---	----------------------	-----------------------------

US 4192647 A	US 4426206	A	82-81583E/39 (DAII)	DAIICHI KOGYO SEIYAKU CO LTD/ HOSOKAWA T; KAKO S; HAGINO K
--------------	------------	---	----------------------	--

US 4192647 A	US 5385585	A	93-251039/32 (WOLF)	WOLFF WALSRODE
--------------	------------	---	----------------------	----------------

AG/KIESEWETTER R; REINHARDT E; KNIEWSKE R; SZABLIKOWSKI

18. Průzkum evropských lékařských časopisů

Ve svém průzkumu jsem se zaměřila na evropské lékařské časopisy. Položila jsem si otázku, jak si stojí v citačních rejstřících. Které jsou z nich nejlepší, jak si stojí jednotlivé země, jejich lékařské časopisy a jednotlivé oblasti medicíny. Použity jsou údaje z **Journal Citation Reports** vytvářené Institute for Scientific Information a polský citační rejstřík **Index Copernicus**. U obou rejstříků jsem vycházela z údajů za rok 2002, novější údaje nejsou z časových důvodů použity, jelikož jsem pro rigorózní práci tento průzkum dělala z celé práce jako první. Ovšem je možné porovnat údaje, které jsem použila již v diplomové práci a zjistit, jaké změny nastaly.

Index Copernicus a údaje z **Journal Citation Reports** nelze porovnávat. Spíše lze říci, že se jedná o databáze, které se doplňují. **Journal Citation Reports** sleduje mezinárodní literaturu, ale časopisy především z Polska, Maďarska a České republiky příliš neindexuje. **Index Copernicus** se spíše zaměřuje na časopisy východní a střední Evropy, i když jsou postupně přidávány časopisy z celého světa, ne jen Evropy a USA, a to především na medicínské, i když v poslední době se zde stále více objevují i další obory. Jsou zde země jako např. Jihoafrická republika, Irán, Saudská Arábie, Turecko, Nigérie, Japonsko, Nový Zéland a další.

V průzkumu jsem z obou databází vybrala lékařské evropské časopisy s jejich hodnotami – v JCR jsem sledovala hodnoty impakt faktoru, immediacy indexu.

U **Index Copernicus** jsem sledovala, zda daný časopis má impakt faktor a hodnotu indikátoru **Index Copernicus**. Hodnota indikátoru **Index Copernicus** je mimo jiné závislá právě na tom, zda časopis je excerpován ISI. Časopisy s impakt faktorem patří mezi nejlepší v **Index Copernicus**.

V databázi **Journal Citation Reports** je 1025 evropských lékařských časopisů, z toho 13 časopisů má impakt faktor větší než 10 a 598 časopisů má impakt faktor rovný nebo větší než 1.00. Nejvíce jsou to tedy Anglie a Nizozemí. Když jsem se podívala na 20 nejlepších časopisů dle impakt faktoru, tak 15 časopisů je vydáváno v Anglii a 5 v Nizozemí. Když jsem si seřadila 20 nejlepších časopisů dle immediacy Index, tak 14 časopisů je vydáváno v Anglii, 3 v Nizozemí, 2 v Německu a 1 v Dánsku. Což samozřejmě značí to, že nejenom v těchto zemích působí velcí vydavatelé, ale také v těchto zemích jsou nejproduktivnější vědci, kteří mají také i výhodu jazyka a také i podporu firem a vlád, které zdejší výzkum podporují. Potěšitelné může být, že z oboru medicíny a příbuzných oborů má Česká republika v JCR zastoupeno 5 časopisů.

V citačním rejstříku **Index Copernicus** z 484 titulů časopisů má 41 časopisů **Index Copernicus** větší nebo rovno 10.00, z toho je 3 Velká Británie, 20 Polsko, 1 Řecko, 2 Švédsko, 1 Dánsko, 8 Česká republika, 1 Francie, 2 Irsko, 1 Rumunsko, 2 Maďarsko.

Tabulky a grafy k průzkumu popisují:

Journal Citation Reports

- Počet časopisů evropských zemí v **Journal Citation Reports**
- Graf Počet časopisů evropských zemí v **Journal Citation Reports**
- 20 nejlepších časopisů dle impakt faktoru v **Journal Citation Reports**
- graf 20 nejlepších časopisů dle impakt faktoru v **Journal Citation Reports**
- Časopisy Polska, České republiky, Slovenska a Maďarska v JCR
- 10 nejlepších z dané země dle impakt faktoru

- Nejlepších 20 časopisů v JCR dle Immediacy Index
- graf Nejlepších 20 časopisů v JCR dle Immediacy Index

Index Copernicus

- Zastoupení časopisů jednotlivých zemí v Index Copernicus
- Graf Zastoupení časopisů jednotlivých zemí v Index Copernicus
- Procenta zastoupení časopisů v IC
- Nejlepší časopisy v Index Copernicus dle zemí, časopisy zařazené mezi 100 nejlepších dle IC (Index Copernicus)
- Nejlepších 100 časopisů v Index Copernicus (hodnoty za rok 2003)
- Nejlepší časopisy dle země a Index Copernicus
- 40 nejlepších evropských časopisů v Index Copernicus
- graf 40 nejlepších evropských časopisů v Index Copernicus

19. ZÁVĚR

Citační analýza je dnes již zavedou a používanou metodou hodnocení časopisů. Je také metodou, která nám umožňuje na základě citací neboli jakýchsi odkazů na jiné prameny nejen pomoci najít zdroje literatury, které mohou být potřebné pro náš výzkum a námi zkoumanou problematiku, ale umožňují nám takto vypátrat, kdo určitou problematiku zpracovanou někým ještě dále rozvíjí nebo kdo jí vyvrací. Zjistit, jaké práce spolu souvisejí, a tak nám pomáhá najít, jak se daná problematika nebo věda či její část rozvíjela, čím se daná věda zabývala dříve a jaké jsou dnes hlavní otázky, na které se daná věda snaží odpovědět.

Citační analýzy, především na základě vypočítání impakt faktoru se snaží porovnávat časopisy mezi sebou a přeneseně jsou také hodnoceni jednotliví vědci a instituce. Ovšem jak je v práci patrné, měli bychom se vyvarovat hodnocení vědců, vědy a institucí z jiných vědních oborů. Je nemožné porovnávat např. fyziku a psychologii, protože každá z těchto oblastí vědy má jiný vývoj, jinou rychlost šíření poznatků a jiné zákonitosti, kterými se řídí. Hlavně v některých oblastech by to mohlo být stejné jako porovnávat studenta střední školy s profesorem matematiky. Oba dva mají za sebou určitý vývoj a oba dva mají různý poměr poznatků, jako to může být u vědní problematiky, která je teprve v plenkách a vědou, která se vyvíjí po staletí. A leccos v ní již bylo objeveno.

I když i ono srovnání vědců v rámci stejné vědní disciplíny má řadu úskalí. Každý stát vytváří různé podmínky pro rozvoj vědy a taktéž instituce. Také v každé zemi může mít věda různou historii. Já osobně nejsem proti používání impakt faktoru nebo jiných prostředků pro hodnocení. Chápu, že určité porovnání musí být a je nutné postupovat podle nějakého stejného klíče. Ale nemyslím si, že naši lékaři jsou méně kvalifikovaní a kvalitní než ti např. ve Velké Británii nebo v USA jen pro to, že své poznatky publikují v českých a tedy národních časopisech s malými impakt faktorem. Také souhlasím s tím, že důležití jsou i ti, kteří propagují vědu, vyučují ji a učí novou generaci vědců a nemusí publikovat obsáhlé práce. Ono podle mého psát obsáhlé a velmi citované práce není všechno. Zvláště u lékařů je velmi důležité, jací jsou v ordinacích, na operačních sálech a ne, jestli umí dobře psát. Protože obzvláště v medicíně je potřebné jednak těch, kteří zkoumají možné způsoby léčení nemocí, nové postupy při operacích, účinky léků, ale také jsou velmi potřeba ti, kteří je pak v praxi využívají.

Podobně jako citace vědecké literatury nám umožňuje mapovat, jak se daný obor rozvíjel, jaké práce měly hlavní vliv na tomto rozvoji, nebo jaké jiné obory přímo nebo nepřímo daný obor ovlivnily, je zajímavé sledovat podobné vazby u patentů. V patentech je to o to ještě zajímavější, že je možné sledovat, jak se např. nějaký vynález dále rozvíjel, kdy se původní myšlenka dále rozvíjela, nebo byla zlepšována. Nebo patent v jednom oboru pomohl vyřešit určitý problém v oboru jiném. Ovšem v patentech nenalezneme jen citace patentů, ale také další vědecké literatury. Bylo by zajímavé vytvářet např. určitou mapu, něco jako myšlenkovou mapu, jak se určitý patent rozvíjel a jaké další patenty na něj navazovaly.

Jsem velmi potěšena, že vznikl polský **Index Copernicus**. Citační rejstřík nezaměřující se jen na polské, ale také na další východoevropské a středoevropské časopisy, který umožňuje porovnání časopisů z tohoto regionu mezi sebou a také, co je velmi důležité, informuje o jejich existenci. Kdyby v tomto rejstříku byly všechny, alespoň lékařské časopisy z této oblasti, mohlo by to přispět k povznesení této oblasti a rozšíření informovanosti o jejich existenci a možná také k větším nákladům těchto časopisů, které by pak začaly být požadovány i za hranicemi své země.

Slovník pojmů a zkratek

AHCI - viz **Arts and Humanities Citation Index**

Analýza shody slov (co-word analysis) - je aplikována na obsah článku. Poskytuje studium výzkumných témat. Identifikuje vazby mezi výzkumnými tématy podle paralelní přítomnosti klíčových slov v dokumentech.

Arts and Humanities Citation Index (AHCI) – citační rejstřík vytvářený ISI retrospektivou od r. 1980. Zaměřuje se na humanitní vědy.

Autocitace – citace, která cituje práci toho autora, který napsal dokument, jenž danou práci cituje.

Bibliometrické zákony – viz Lotkův zákon, Zipfův zákon a Bradfordův zákon.

Bibliometrie - statistický popis literatury – skupin dokumentů, které mají mezi sebou určitý vztah. To dovoluje, aby literatura byla popsána a monitorována. Bibliometrie se nejvíce zabývá studiem tištěné literatury.

Bibliometrické indikátory - popisují vědy a technologie v oblasti produkce a toku znalostí na mezinárodní, národní, regionální, subregionální a institucionální úrovni.

Bradfordův zákon - jeden ze základních bibliometrických zákonů - $1: n : n^2$, udává, že literatura k danému tématu se z 1/3 nachází v časopisech daného oboru, z 1/3 v časopisech příbuzných oborů a poslední 1/3 v ostatní literatuře.

Citační analýza – techniky využívané v bibliometrii. Citační analýza podává informace o některých aspektech publikací a citačních modelech dané vědní oblasti. Data poskytují informace o „nejvlivnějších“ jednotlivých představitelích disciplíny a o jejich pracích.

Cited half-life - "Poločas citovanosti" - udává po kolika letech (ve vztahu k aktuálnímu roku) se objeví 50% všech citací na články daného časopisu v citačních rejstřících. Napomáhá při hodnocení stáří citovaných článků. Uvádí se pouze u časopisů, které byly alespoň 100 a vícekrát citovány.

Citing half-life - "Poločas citování" - napomáhá při hodnocení stáří většiny článků, jež cituje daný časopis. Uvádí se pouze u časopisů, které citují 100 nebo více dokumentů.

Co-citation analysis - viz Kocitační analýza.

Co-word analysis - viz **Analýza shody slov**.

Current Contents – je databáze aktuálních obsahů časopisů, sborníků a jiných dokumentů vytvářená Institute for Scientific Information, založené Eugenem Garfieldem v roce 1957. Umožňuje na základě nastaveného osobního profilu snadno a rychle získávat informace o člancích a jiných dokumentech k žádanému tématu.

Český citační rejstřík lékařské literatury (CR-CZ) - citační rejstřík vytvořený v rámci grantového projektu Interní grantové agentury MZ ČR s názvem „**Informační charakteristiky vývoje lékařské vědy a výzkumu**“. Má dvě části: **Citační rejstřík (CR)** a **Citační rejstřík -faktory (CRF)**.

Eugene Garfield – zakladatel Institute for Scientific Information, zakladatel Current Contents a citačních rejstříků SCI, SSCI a AHCI.

faktor bezprostředního vlivu - viz **Immediacy Index**.

faktor vlivu - viz **Impakt faktor**.

IF - viz Impakt faktor.

II - viz Immediacy Index.

Immediacy Index (faktor bezprostředního vlivu; II; Garfieldův Index) - udává průměrný počet citací na články časopisu v roce, ve kterém byly publikovány.

$$II = \frac{\text{Počet citací v roce X}}{\text{Počet publikovaných článků v roce X}}$$

Impakt Faktor (IF; faktor vlivu) – vyjadřuje, kolikrát je článek uveřejněný v daném časopise v průměru citován.

$$IF = \frac{\text{Počet citací v roce X na články časopisu publikované v roce X-1 + X-2}}{\text{Počet článků publikovaných v časopisu v letech X-1 + X-2}}$$

Garfieldův Index - viz Immediacy Index.

Infometrie - obecnější termín, který pokrývá scientometrii a bibliometrii. Je to kvantitativní studie publikačních a citačních modelů.

Institute for Scientific Information (ISI) – instituce založená E. Garfieldem a vytvářející Current Contents, Science Citation Index, Social Science Citation Index a Arts and Humanities Citation Index.

ISI - viz Institute for Scientific Information.

JCR - viz Journal Citation Reports.

Journal Citation Reports (JCR) - samostatná část SCI (Science Citation Index), SSCI (Social Science Citation Index), ve které jsou obsaženy výsledky scientometrických analýz všech časopisů a publikací excerptovaných pro tyto rejstříky. Je vytvářena Institute for Scientific Information od roku 1973.

Kocitační analýza (co-citation analysis) - vymezuje vazby mezi články na základě citací. Tato vazba mezi dvěma články existuje v případě, kdy jsou současně citovány třetím článkem.

Lotkův zákon – jeden ze základních bibliometrických zákonů. Udává frekvenční distribuci vědecké produktivity

Matthewovy časopisy jádra - viz MCJ

Matthewův efekt - model špatného hodnocení vědecké práce popsáný **Robertem K. Mertonem** v roce 1968. Matthewův efekt spočívá v uznání za jednotlivé vědecké příspěvky vědcům s větší reputací a odepření ocenění vědcům, kteří ještě nejsou známí.

Matthewův efekt pro země - viz MEC - Matthewův efekt pro země

MCJ – Matthewovy časopisy jádra - jsou umístěny obvykle v regionech s velkým podílem publikování a počty citací. Ne vždy časopis s vysokým počtem prací a/nebo citací musí nezbytně být MCJ.

MEC - Matthewův efekt pro země - MEC spočívá v systematickém odchýlení počtu aktuálních citací od předpokládaného počtu citací.

Naukometrika - byla vytvořena nezávisle v SSSR v 30. letech 20. století. Bohužel v západních zemích tento pojem byl nahrazen třemi rozdílnými termíny: bibliometrie, scientometrie a infometrie.

Poločas citování - viz **Citing half-life**.

Poločas citovanosti - viz **Cited half-life**.

SCI - viz **Science Citation Index**.

Science Citation Index (SCI) – citační rejstřík vytvořený v roce 1963 společností **Institute for Scientific Information (ISI)** ve Philadelphii. Zahrnuje přírodní a technické obory. V elektronické podobě retrospektiva od r. 1974.

Scientometrie – se zabývá studiem rozvoje fyzikálních a biologických věd.

SFX (Special effects) – technologie vytvořená ve spolupráci odborníků Gentské univerzity (Belgie) a knihovny Národní laboratoře v Los Alamos (USA). U nás je tato technologie využívána v rámci projektu Národní knihovny a Karlovy Univerzity – **Jednotná informační brána**. **SFX** umožňuje pomocí hypertextového odkazu propojit citace nebo bibliografický záznam na plný text dokumentu.

Social Science Citation Index (SSCI) – citační rejstřík vytvořený **ISI** s retrospektivou v elektronické podobě od roku 1972, zaměřující se na sociální vědy

Special effects - viz **SFX**.

SSCI - viz **Social Science Citation Index**.

Total-Basic Score - celkové základní skóre používané v Index Copernicus pro hodnocení časopisů excerpovaných v databázi.

Web of Knowledge, Web of Science – webovské rozhraní citačních rejstříků Institute for Scientific Information (Science Citation Index, Social Science Citation Index a Arts and Humanities Citation Index a databáze časopisů excerpovaných **ISI Journal Citation Reports**)

Zipfův zákon - jeden ze základních bibliometrických zákonů. Popisuje distribuci slov v textu. Udává, že důvěrná slova bývají použita mnohokrát a mnoho slov bývá použito jen jednou.

Seznam použité literatury

- ADAM, David. The counting house. News feature. *Nature*. 2002, vol. 415, no. 14. february. S. 726-729.
- AMIN, M.; MABE, M. Impact factors: use and abuse. *Perspectives in Publishing*. 2000, no. 1.
- BJÖRNEBORN, Lennart. ; INGWERSEN, Peter. Perspectives of webometrics. *Scientometrics*. 2001, vol. 50, no. 1, s. 65-82.
- BJÖRNEBORN, Lennart; INGWERSEN, Peter. Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*. 2004, Vol. 55, no. 14, s. 1216-1227.
- BLANK, Lynne. A Survey of non-patent reference in Selected Patent Classes. [S.l.] : Sutesurfer Publishing LLC (Franklin Pierce Law Center), 1997-1998. Dostupné z www: <http://www.ipmall.info/hosted_resources/bp98/lbank.htm>.
- Bluesheets. Dostupné z www: <www.dialog.com>.
- BONITZ, Manfred; SCHARNHORST, Andrea. National Science Systems and the Matthew effect for countries. Prezentované na konferenci: Globalisierung und Wissensorganisation, Hamburg, Německo, 1999. A na konferenci Seventh International Conference on Scientometrics and Infometrics, Mexiko, Cilima, 5.-8. července 1999.
- BRATKOVÁ, Eva. *Bibliografické odkazy a citace tradičních i elektronických dokumentů podle mezinárodní normy ISO 690*. Příručka pro autory odborných publikací. Praha : Univerzita Karlova, Ústav informačních studií a knihovnictví FF UK, 2001. 27 s.
- BREITZMAN, Anthony ; THOMAS, Patric. Using patent citation analysis to target/value M & A candidates. *Research Technology Management*. 2002, vol. 45, no. 5, p. 28-36.
- BRIN, Beth; COCHRAN, Elissa. Access and ownership in the academic environment: one library's progress report. *Journal of Academic Librarianship*. 1994, vol. 20, no. 4, s. 207-212.
- BURRELL, Quentin L. Measuring similarity of concentration between different infometric distributions: two new approaches. *Journal of the American Society for Information Science and Technology*. 2005, vol. 56, no. 7, s. 704-714.
- Case study: patent citations*. STN Patent Forum: Patent Citations. Dostupné z WWW: http://www.stn-international.de/training_center/patents/PatentForum98/citat.pdf
- COTHNEY, Viv. Web-crawling reliability. *Journal of the American Society for Information Science*. 2004, roč. 55, č. 14, s. 1228-1238.
- Česká terminologická databáze knihovnictví a informační vědy (projekt TDKIV) [online]. [cit. 2003-08-05]. Dostupná z WWW: <<http://www.nkp.cz>>.

- DAVIS, Philip M. Effect of the Web on undergraduate citation behavior: guiding student scholarship in a networked age. *Portal: Libraries and the Academy*. 2003, vol. 3, no. 1, s. 41-51.
- DING, Ying; CHOWDHURY, Gobinda; FOO, Schubert. Incorporating the results of co-word analyses to increase search variety for information retrieval. *Journal of Information Science*. 2000, vol. 26, no. 6, s. 429-451.
- DOMINICK, Joseph R. Citation analysis of the Journal of Broadcasting & Electronic Media : another perspective. *Journal of Broadcasting & Electronic Media, The (Periodical)*. 1997, vol. 41, no. 3, 427-438.
- EGGHE, Leo. Type/token-taken informetrics. *Journal of the American Society for Information Science and Technology*. 2003, vol. 54, no. 7, s. 603-610.
- FAIRTHORNE, Robert A. Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction. *Journal of Documentation*. 2005, vol. 61, no. 2, s. 171-193.
- GALLAGHER, Hill. Misuse of citations is a dangerous game. *National Underwriter / Life & Health Financial Services*. 1998, vol. 102, no. 22, s. 11.
- GARFIELD, Eugene. Citation Analysis of Sports Medicine Research, 1981-1996 Productivity, Impact and Influence of Nations, Institutions and Researchers [online]. [Prezentované na American College of Sports Medicine, 44th Annual Meetin, 1997, Denver, USA. [cit. 2002-10-08]. Dostupné z WWW: <<http://www.garfield.library.upenn.edu/>>.
- GARFIELD, Eugene. Current Contents ninth anniversary. Essays of an information science. *Current Contents*. 1962-73, vol. 1, s. 12 (Current Contents, February 21, 1967).
- GARFIELD, Eugene. How it all began – with a loan from HFC. *Current Contents*. 1980, no. 3, s. 5-8. (Essays of an Information Scientist. 1979-1980, vol. 4, s. 359-362).
- GARFIELD, Eugene. Patent citation indexing and the notions of novelty, similarity, and relevance. *Essays of an Information Scientist*. 1984, Vol. 7, s. 536-542. Přetištěno z *Journal of Chemical Documentation*, 1966, no. 6, s. 63.
- GARFIELD, Eugene. *The application of citation indexing to the patent literature*. Citation Indexing. Chapter 4. Dostupné z www: <<http://www.garfield.library.upenn.edu/ci/chapter4.PDF> >.
- GELMAN, Sheldon R.; GIBELMAN, Margaret. A quest for citation? An analysis of and commentary on the trend toward multiple authorship. *Journal of Social Work Education*. 1999, vol. 35, no. 2, s. 201-212.
- GISVOLD, Sven-Erik. Citation analysis and journal impact factors – is the tail wagging the dog?. *Acta Anaesthesiologica Scandinavica*. 1999, vol. 43, s. 971-973.
- GITTELMAN, Michelle; KOGUT, Bruce. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. *Management Science*. 2003, vol. 49, no. 4 (April), s. 366-382.

- GRANT, Jonathan; COTTRELL, Robert CLUZEAU, Françoise; GAIL, Fawcett. Evaluating „payback“ on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *BMJ*. 2000, vol. 320, no. april, s. 1107-1011.
- HALL, Bronwyn H.; JAFFE, Adam; TRAJTENBERG, Manuel. *Market value and patent citations: a first look*. Working Paper. No. E01-304. 2001. UC Berkeley Working Papers. 65 s. Department of Economics. Dostupné z WWW: <<http://www.haas.berkeley.edu/groups/iber/wps/econwp.html>>.
- HARTER, Stephen P. The Impact of electronic journals on scholarly communication: a citation analysis. *The Public-Access Computer Systems Review*, 1996, vol. 7, no. 5, s. 5-34.
- HERCOVÁ, Jana. Dr. Eugen Garfield - 75. Narozneniny. *Lékařská knihovna*. 2000, roč. 5, č. 5/6, s. 11-14.
- HOAAS, David J.; MADIGAN, Lauren J. Citation analysis of economists in principles of economics textbooks. *Social Science Journal*. 1999, vol. 36.
- HYLAND, Ken. Self-citation and self-reference: credibility and promotion in academic publication. *Journal of the American Society for Information Science and Technology*. 2003, vol. 54, no. 3, s. 251-259.
- CHEN, Kuang-hua. The construction of the Taiwan Humanities Citation Index. *Online Information Review*. 2004, vol. 28, no. 6, s. 410-419. Dostupné z WWW: <www.emeraldinsight.com/1468-4527.htm>.
- Index Copernicus [online]. [cit. 2003-07-18]. Dostupný z WWW: <<http://www.cisi.org>>.
- Infometrics [online]. [cit. 2002-01-24]. Dostupné z: <<http://kmi.open.ac.uk/people/victoria/Infometrics.htm>>
- Institute for Scientific Information – webovská stránka [online]. [2003-07-29]. Dostupná z WWW: <www.isinet.com>
- ISI. Timeline [online]. [cit. 2003-07-23]. Dostupné z WWW: <<http://www.isinet.com/isi/about/timeline.html>>.
- JACSO, Peter. The future of citation indexing: an interview with Eugene Garfield. *Online*. 2004, vol. 28, no. 1, s. 38.
- JAROLÍMKOVÁ, Adéla. Index Copernicus. *Ikaros* [online]. 2002, č.4 [cit. 2002-04-01]. Dostupný z WWW: <<http://www.ikaros.cz/Clanek.asp?ID=200208477>>.
- Journal Citation Reports [online]. [cit. 2003-07-16]. Dostupný např. přes WWW: <<http://www.isinet.com/isi/products/citation/jcr/>>.
- KARLSSON, Fred Linguistics in the light of citation analysis. *Publications of Department of General Linguistics, University of Helsinki*. 1994, no. 23.
- KATZ, J. Sylvan; HICKS, Diana. How much is a collaboration worth? A calibrated bibliometric model. In *Proceedings on the Sixth Conference of The International Society for Scientometric and Infometric*, 1997, Jerusalem, Izrael, . S. 163-175.
- KEELY, Louise C. *Using Patents in Growth Models*. Working papers 30. Madison : University of Wisconsin. Department of Economics, 2000. Dostupné z www: < <http://www.ssc.wisc.edu/~lkeely/surv1000hs.pdf> >

- KOEHLER, Wallace; COLIMNIST, Guest. Nightmares in citation analysis. *Reference & User Services Quarterly*. 2002, vol. 42, no. 1, s. 41-42.
- LEYDESDORFF, Loet. Similarity measures, author cocitation analysis, and information theory. *Journal of the American Society for Information Science and Technology*. 2005, vol. 56, no. 7, s. 769-772.
- LUKACH, Ruslan; PLASMANS, Joseph. Knowledge spillovers in Belgium: evidence from the firm's patent citation behavior. Antverpy: Faculty of Applied Economics UFSIA-RUCA, University of Antverp, 2001. Dostupné z www: <<http://www.ufsia.ac.be/~plasmans/papers/patcit.pdf>>.
- MICHEL, Jacques; BETTELS, Bernd. Patent citation analysis : a closer look at the basic input data from patent search reports. *Scientometrics*. 2001, vol. 51, no. 1, s. 185-201.
- Lotka, Alfred James (1880-1949), USA [online]. [cit. 2003-06-17]. Dostupné z WWW: <<http://users-pandora.be/ronald.rousseau/html/loka.html>>.
- MATTHIESSEN, Christian Wichmann; SCHWARZ, Annette Winkel ; FIND, Søren. The top-level global research system, 1997-99 : centres, networks and nodality. An analysis based on bibliometric indicators. *Urban Studies*. 2002, vol. 39, no. 5-6, s. 903-927.
- MCMILLAN, G. Steven; HAMILTON, Robert D. Using bibliometrics to measure firm knowledge : an analysis of the pharmaceutical industry. *Technology Analysis & Strategic Management*. 2000, vol. 12, no. 4, s. 465-476.
- MENCZER, Filippo. Lexical and semantic clustering by web links. *Journal of the American Society for Information Science and Technology*. 2004, vol. 55, no. 14, s. 1261-1269.
- MERTON, Rober K. The Matthew effect in science II. *ISIS*. 1988, vol. 79, s. 606-623.
- MERTON, Robert K. The Matthew effect in science. *Science*. 1968, vol. 159, no. 3810, s. 56-63.
- NAJMAN, Jake M.; HEWITT, Belina. The validity of publication and citation counts for sociology and other selected disciplines. *Journal of Sociology*. 2003, vol. 39, no. March, s. 62-80.
- OPTHOF, Tobias. Sense and nonsense about the impact factor. *Cardiovascular Research*. 1997, vol. 33, s. 1-7.
- PITTEROVÁ, Květa. Využití citačních rejstříků a impact factor. *Praktický lékař*. 1999, roč. 79, č. 12 , s. 714-716.
- REDMAN, Arnold L.; MANAKYAN, Herman; TANNER, John R. The ranking of real estate journals: a citation analysis approach. *Financial Practice & Education*. 1998, vol. 8, no. 2, s. 59-69.
- ROMANO, Claudio; RATNATUNGE, Janek. Citation analysis of the impact of journals on contemporary small enterprise research. *Entrepreneurship : Theory & Practice*. 1996, vol. 20, no. 3, s. 7-21.

- ROSSEAU, Ronald. Journal Evaluation: technical and practical Issues. *Library Trends*. 2002, vol. 50, no. 3, s. 418-440.
- ROUSSEAU, Ronald. Robert Fairthorne and the empirical power laws. *Journal of Documentation*. 2005, vol. 61, no. 2, s. 194-202.
- ROY, D. [et al.]. Citation analysis of otorhinolaryngology journals. *The Journal of Laryngology and Otology*. 2002, vol. 116, no. 5, s. 363.
- Scopus. Dostupné z www:< <http://www.info.scopus.com>> nebo <www.scopus.com>
- SHERRI, Edwards. Citation analysis as a collection development tool : a bibliometric study of polymer science theses and dissertation. *Serials Review*. 1999, vol. 25, č. 1, s. 11-20.
- STEMBRIDGE, Bob; CORISH, Breda. Patent data mining and effective patent portfolio management. *Intellectual Asset Management*. 2004, no. October/November, s. 30-35.
- THELWALL, Mike; VAUGHAN, Liwen. Webometrics : an introduction to the special issue. *Journal of the American Society for Information Science and Technology*. 2004, vol. 55, no. 14, s. 1213-1215.
- THIESSEN, Richard. A bibliometric study of Direction. *Direction*. 1992, vol. 21, no. 1, s. 83-93.
- VAUGHAN, Liwen; THELWALL, Mike. Scholarly use of the web: what are the key inducers of links to journal web sites? *Journal of the American Society for Information Science and Technology*. 2003, vol. 54, no. 1, s. 29-38.
- Web of Knowledge [online]. [cit. 2005-10-05]. Dostupné z WWW: <<http://www.isiwebofknowledge.com/>>.
- Web of Science [online]. [cit. 2003-08-05]. Dostupné z WWW: <<http://www.isinet.com/isi/products/citation/wos/index.html>>.
- WRIGHT, Richard; MILLER, J. Mitchell. The most-cited scholars and works in corrections. *Prison Journal*. 1999, vol. 79, no. 1, s. 5-23.

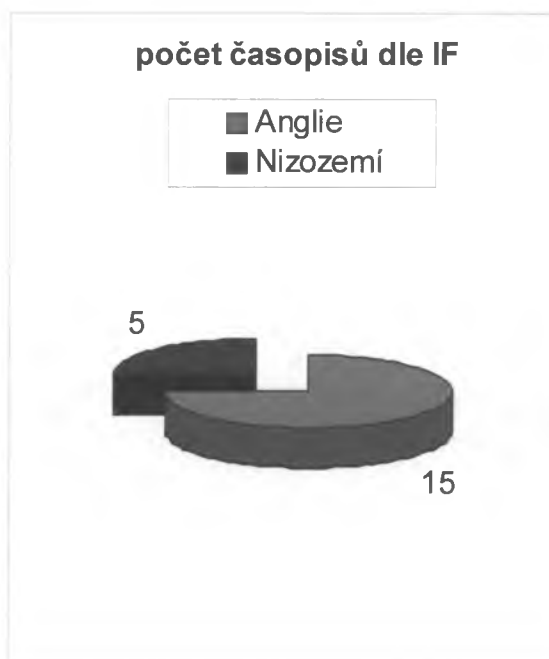
Příloha ke kapitole 12: Průzkum - Údaje z Journal Citation Reports

Počet časopisů evropských zemí v Journal Citation Reports

země	počet časopisů
Anglie	407
Belgie	7
Česká rep.	5
Dánsko	37
Finsko	3
Francie	55
Chorvatsko	1
Irsko	15
Itálie	37
Jugoslávie	1
Maďarsko	2
Německo	150
Nizozemí	131
Norsko	28
Polsko	8
Rakousko	6
Rusko	4
Řecko	4
Skotsko	21
Slovensko	2
Španělsko	14
Švédsko	4
Švýcarsko	82
Ukrajina	1
Celkem	1025

Progress in neurobiology	10.672	Anglie
Biochimica et biophysica acta - reviews on cancer	9.351	Nizozemí
Molecular medicine today	8.736	Anglie
Human molecular genetics	8.726	Anglie
Seminars in immunology	8.705	Anglie
Atherosclerosis supplements	8.300	Anglie
Trends in cognitive sciences	8.129	Anglie
Progress in lipid research	8.000	Anglie

Země	počet časopisů
Anglie	15
Nizozemí	5



Časopisy Polska, České republiky, Slovenska a Maďarska v JCR

Procenta znamenají kolísání impakt faktoru oproti předešlému roku

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Folia biologica	0.615	2002		Česká rep.
	0.519	2001	16%	
	0.351	2000	53%	
	0.493	1999	40%	
	0.632	1998	28%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Česká a slovenská neurologie a neurochirurgie	0.052	2002		Česká rep.
	0.041	2001	22%	
	0.059	2000	44%	
	0.065	1999	10%	
	0.029	1998	45%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Folia parasitologica	0.515	2002		Česká rep.
	0.557	2001	4%	
	0.844	2000	52%	
	0.796	1999	6%	
	0.706	1998	11%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Physiological research	0.984	2002		Česká rep.
	1.027	2001	4%	
	1.366	2000	33%	
	0.521	1999	62%	
	0.616	1998	35%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Acta virologica	0.660	2002		Česká rep.
	0.644	2001	2%	
	0.558	2000	13%	
	0.476	1999	15%	
	0.500	1998	5%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Neoplasma	0.679	2002		Slovensko
	0.637	2001	7%	
	0.579	2000	9%	
	0.448	1999	23%	
	0.657	1998	46%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
General physiology and biophysics	0.719	2002		Slovensko
	0.932	2001	30%	
	0.417	2000	45%	
	0.400	1999	4%	
	0.714	1998	79%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Haematologia	0.293	2002		Maďarsko
	0.400	2001	37%	
	0.405	2000	1%	
	0.315	1999	22%	
	0.104	1998	33%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Acta alimentaria	0.284	2002		Maďarsko
	0.275	2001	3%	
	0.246	2000	11%	
	0.177	1999	28%	
	0.400	1998	126%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Archivum Immunologiae et therapiae experimentalis	0.793	2002		Polsko

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Acta neurobiologiae experimentalis	0.910	2002		Polsko
	0.767	2001	16%	
	0.631	2000	18%	
	0.500	1999	21%	
	0.766	1998	35%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Folia neuropathologica	0.394	2002		Polsko
	0.471	2001	27%	
	0.258	2000	27%	
	0.338	1999	31%	
	0.211	1998	38%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Acta parasitologica	0.732	2002		Polsko
	0.732	2001	0%	
	0.433	2000	41%	
	0.434	1999	0%	
	0.410	1998	6%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Polish journal of pharmacology	0.684	2002		Polsko
	0.711	2001	4%	
	0.456	2000	36%	
	0.520	1999	14%	
	0.329	1998	37%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Journal of physiology and pharmacology	1.406	2002		Polsko
	1.459	2001	4%	
	1.025	2000	30%	
	1.270	1999	14%	
	1.114	1998	37%	

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Annals of agricultural and environmental medicine	0.851	2002		Polsko

název časopisu	impakt faktor	rok v JCR	procenta IF	země
Biology of sport	0.157	2002		Polsko
	0.255	2001	62%	
	0.164	2000	36%	
	0.157	1999	4%	
	0.194	1998	24%	

10 nejlepších z dané země dle impakt faktoru

název časopisu	impakt faktor	země
Nature reviews neuroscience	24.047	Anglie
Nature reviews genetics	21.762	Anglie
Lancet	15.397	Anglie
Nature reviews immunology	14.059	Anglie
Nature reviews cancer	13.625	Anglie
Current opinion in immunology	12.918	Anglie
Trends in ecology & evolution	11.929	Anglie
Current opinion in neurobiology	10.718	Anglie
Progress in neurobiology	10.672	Anglie
Molecular medicine today	8.736	Anglie

název časopisu	impakt faktor	země
Journal of the European academy of dermatology and venereology	1.021	Belgie
Acta cardiologica	0.673	Belgie
Acta gastro-enterologica belgica	0.636	Belgie
Acta neurologica belgica	0.577	Belgie
Neural processing letters	0.317	Belgie
Acta clinica belgica	0.287	Belgie
Acta chirurgica belgica	0.269	Belgie
název časopisu	impakt faktor	země
Physiological research	0.984	Česká rep.
Acta virologica	0.660	
Folia biologica	0.615	
Folia parasitologica	0.515	
Česká a slovenská neurologie a neurochirurgie	0.052	

název časopisu	impakt faktor	země
Immunological reviews	7.409	Dánsko
Allergy	3.666	Dánsko
Bipolar disorders	3.097	Dánsko
Xenotransplantation	2.581	Dánsko
Xenotransplation	2.581	Dánsko
Periodontology 2000	2.493	Dánsko
Liver	2.403	Dánsko
Experimental dermatology	2.303	Dánsko
Acta psychiatrica scandinavica	2.259	Dánsko
Clinical genetics	2.237	Dánsko

název časopisu	impakt faktor	země
Annals of medicine	3.422	Finsko
Scandinavian journal of work environment & health	1.848	Finsko
Annales chirurgiae et gynaecologiae	0.259	Finsko

název časopisu	impakt faktor	země
Chemical senses	2.526	Francie
European cytokine network	2.153	Francie
International journal of tuberculosis and lung disease	1.888	Francie
European journal of medicinal chemistry	1.705	Francie
Neurophysiologie clinique - clinical neurophysiology	1.524	Francie
Neuropsychologie clinique-clinical neurophysiology	1.524	Francie
European psychiatry	1.327	Francie
Diabetes & metabolism	1.244	Francie
Journal of physiology - Paris	1.070	Francie
Biomedicine & pharmacotherapy	1.035	Francie

název časopisu	impakt faktor	země
Neurologica croatica	0.143	Chorvatsko

název časopisu	impakt faktor	země
Toxicology letters	2.242	Irsko
Clinical neurophysiology	2.120	Irsko
Gait & posture	1.753	Irsko
Resuscitation	1.551	Irsko
Journal of dermatological science	1.279	Irsko
Journal of ethnopharmacology	1.188	Irsko
European journal of radiology	1.118	Irsko

Forensic science international	1.023	Irsko
International journal of medical informatics	1.000	Irsko
Patient education and counseling	0.995	Irsko

název časopisu	impakt faktor	země
International journal of immunopathology and pharmacology	4.091	Itálie
Haematologica	3.226	Itálie
Sarcoidosis vasculitis and diffuse lung diseases	2.831	Itálie
Archives Italiennes de biologie	2.021	Itálie
Quarterly journal of nuclear medicine	1.903	Itálie
Nutrition metabolism and cardiovascular diseases	1.679	Itálie
Aging clinical and experimental research	1.571	Itálie
International journal of biological markers	1.571	Itálie
Journal of endocrinological investigation	1.476	Itálie
Digestive and liver disease	1.303	Itálie

název časopisu	impakt faktor	země
Croatian medical journal	0.710	Jugoslávie

název časopisu	impakt faktor	země
Haematologia	0.293	Maďarsko
Acta alimentaria	0.284	Maďarsko

název časopisu	impakt faktor	země
Reviews of physiology biochemistry and pharmacology	5.615	Německo
Diabetologia	5.136	Německo
Thrombosis and haemostasis	4.357	Německo
European journal of nuclear medicine and molecular imaging	3.568	Německo

Neurogenetics	3.361	Německo
Journal of neurology	3.345	Německo
Journal of molecular medicine - JMM	3.303	Německo
Psychofarmacology	3.275	Německo
Psychopharmacology	3.275	Německo
Chembiochem	3.233	Německo
Journal of molecular evolution	3.041	Německo

název časopisu	impakt faktor	země
Trends in neurosciences	14.474	Nizozemí
Trends in pharmacological sciences	13.276	Nizozemí
Trends in genetics	13.216	Nizozemí
Immunology today	12.856	Nizozemí
Biochimica et biophysica acta - reviews on cancer	9.351	Nizozemí
Parasitology today	7.253	Nizozemí
Mutation research-reviews in mutation research	7.085	Nizozemí
Brain research reviews	6.178	Nizozemí
Journal of hepatology	4.974	Nizozemí
Current medical chemistry	4.966	Nizozemí

název časopisu	impakt faktor	země
Cephalalgia	3.775	Norsko
Scandinavian journal of rheumatology	2.000	Norsko
Acta oncologica	1.909	Norsko
Scandinavian journal of gastroenterology	1.847	Norsko
Scandinavian journal of immunology	1.782	Norsko
Amyothropic lateral sclerosis and other motor neuron disorders	1.639	Norsko
Acta dermato-venerologica	1.580	Norsko
Cytokines cellular & molecular therapy	1.560	Norsko
Blood pressure	1.344	Norsko
Scandinavian journal of rehabilitation	1.333	Norsko

medicine		
----------	--	--

název časopisu	impakt faktor	země
Journal of physiology and pharmacology	1.406	Polsko
Acta neurobiologiae experimentalis	0.910	Polsko
Annals of agricultural and environmental medicine	0.851	Polsko
Archivum Immunologiae et therapiae experimentalis	0.793	Polsko
Acta parasitologica	0.732	Polsko
Polish journal of pharmacology	0.684	Polsko
Folia neuropathologica	0.394	Polsko
Biology of sport	0.157	Polsko

název časopisu	impakt faktor	země
Journal of neural transmission - supplement	3.312	Rakousko
Journal of neural transmission	2.289	Rakousko
Archives of virology	1.967	Rakousko
Acta neurochirurgica	0.779	Rakousko
Wiener klinische Wochenschrift	0.591	Rakousko
Neuro-orthopedics	0.000	Rakousko

název časopisu	impakt faktor	země
Bulletin of experimental biology and medicine	0.314	Rusko
Kardiologiya	0.277	Rusko
Journal of evolutionary biochemistry and physiology	0.219	Rusko
Gematologiya i transfuziologiya	0.091	Rusko

název časopisu	impakt faktor	země
International journal of oncology	2.931	Řecko
Anticancer research	1.447	Řecko
Oncology reports	1.171	Řecko
In vivo	1.115	Řecko

název časopisu	impakt faktor	země
Blood reviews	2.243	Skotsko
Clinical nutrition	1.551	Skotsko
Platelets	1.477	Skotsko
Neuropeptides	1.477	Skotsko
Fibrinolysis & protelysis	1.435	Skotsko
Prostaglandins leukotrienes and essential fatty acids	0.958	Skotsko
Tissue & cell	0.808	Skotsko
British journal of plastic surgery	0.796	Skotsko
International journal of oral and maxillofacial surgery	0.754	Skotsko
Journal of cranio-maxillofacial surgery	0.750	Skotsko

název časopisu	impakt faktor	země
General physiology and biophysics	0.719	Slovensko
Neoplasma	0.679	Slovensko

název časopisu	impakt faktor	země
Histology and histopathology	1.881	Španělsko
Revista espanola de cardiologia	0.941	Španělsko
Medicina clinica	0.854	Španělsko
Journal of physiology and biochemistry	0.696	Španělsko
Revista espanola de enfermedades digestivas	0.594	Španělsko
Drug news & perspectives	0.524	Španělsko

Nefrologia	0.513	Španělsko
Methods and findings in experimental and clinical pharmacology	0.489	Španělsko
Drugs of the future	0.386	Španělsko
Revista clinica Espanola	0.340	Španělsko

název časopisu	impakt faktor	země
Neuroendocrinology letters	1.278	Švédsko
Journal of rehabilitation medicine	1.000	Švédsko
Acta radiologica	0.883	Švédsko
Swedish dental journal	0.711	Švédsko

název časopisu	impakt faktor	země
Who technical report series	7.900	Švýcarsko
Brain pathology	5.652	Švýcarsko
International journal of cancer	4.056	Švýcarsko
Psychotherapy and psychosomatics	3.188	Švýcarsko
Mechanisms of ageing and development	2.867	Švýcarsko
Dementia and geriatric cognitive disorders	2.804	Švýcarsko
Bulletin of the World Health Organization	2.694	Švýcarsko
Drug and alcohol dependence	2.528	Švýcarsko
Neuroendocrinology	2.511	Švýcarsko
Journal of vascular research	2.458	Švýcarsko

název časopisu	impakt faktor	země
Experimental oncology	0.196	Ukrajina

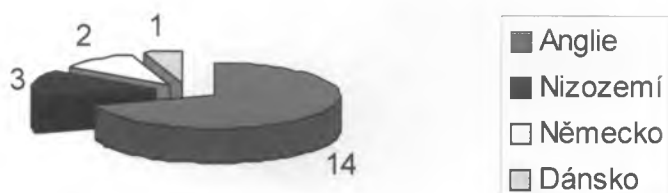
Nejlepších 20 časopisů v JCR dle Immediacy Index

název časopisu	immediacy index	země
Lancet	5.299	Anglie
Nature reviews immunology	4.656	Anglie
Nature reviews cancer	3.484	Anglie
Nature reviews neuroscience	3.116	Anglie
Nature reviews neuroscience	3.116	Anglie
British Medical Journal	2.893	Anglie
Current opinion in immunology	2.885	Anglie
Nature reviews genetics	2.675	Anglie
Nature reviews genetics	2.675	Anglie
Trends in genetics	2.305	Nizozemí
Reviews of physiology biochemistry and pharmacology	2.167	Německo
Neuropsychiatrie	2.071	Německo
Trends in pharmacological sciences	1.913	Nizozemí
Periodontology 2000	1.800	Dánsko
Trends in molecular medicine	1.798	Anglie
Trends in ecology & evolution	1.754	Anglie
European heart journal	1.715	Anglie
Nature reviews drug discovery	1.681	Anglie
Biochimica et biophysica acta - reviews on cancer	1.667	Nizozemí
International journal of neuropsychopharmacology	1.625	Anglie

20 nejlepších časopisů dle Immediacy Index

země	počet časopisů
Anglie	14
Nizozemí	3
Německo	2
Dánsko	1

**20 nejlepších časopisů dle Immediacy Index
dle počtu časopisů**



Index Copernicus

Zastoupení časopisů jednotlivých zemí v Index Copernicus

Země	počet časopisů
Velká Británie	3
Turecko	11
Švédsko	2
Španělsko	2
Řecko	8
Rumunsko	2
Polsko	356
Maďarsko	53
Litva	5
Itálie	1
Irsko	3
Chorvatsko	1
Francie	7
Estonsko	1
Dánsko	1
Česká republika	28
Celkem	484

Zastoupení časopisů v IC pro jednotlivé země počet časopisů



■ Velká Británie	■ Turecko	□ Švédsko
■ Španělsko	■ Řecko	■ Rumunsko
■ Polsko	■ Maďarsko	■ Litva
■ Itálie	□ Irsko	■ Chorvatsko
■ Francie	■ Estonsko	■ Dánsko
■ Česká republika		

Procenta zastoupení časopisů v Index Copernicus

Země	procenta
Velká Británie	0.620
Turecko	2.270
Švédsko	0.410
Španělsko	0.410
Řecko	1.650
Rumunsko	0.410
Polsko	73.550
Maďarsko	10.950
Litva	1.030
Itálie	0.210
Irsko	0.620
Chorvatsko	0.210
Francie	1.450
Estonsko	0.210
Dánsko	0.210
Česká republika	5.790

Nejlepší časopisy v Index Copernicus dle zemí, časopisy zařazené mezi 100 nejlepších dle IC (Index Copernicus)

název	index copernicus	země
Physiological research	16.18	Česká rep.
Collection of Czechoslovak chemical communications	13.61	Česká rep.
Folia microbiologica	12.910	Česká rep.
Folia biologica	11.360	Česká rep.
Homeostatis in health and disease	11.180	Česká rep.
Folia parasitologica (Česká republika)	11.080	Česká rep.
Československá psychologie	10.240	Česká rep.
Chemické listy	10.200	Česká rep.
Česká a slovenská neurologie a neurochirurgie	10.060	Česká rep.
Central european journal of public health	6.870	Česká rep.
Acta chirurgiae plasticae	6.670	Česká rep.
Časopis lékařů českých	6.020	Česká rep.
Otorinolaryngologie a foniatrie	5.820	Česká rep.
Vnitřní lékařství	5.820	Česká rep.

Journal of physiology and pharmacology	26.83	Polsko
Acta neurobiologiae experimentalis	15.49	Polsko
Acta theriologica	14.20	Polsko
Archivum immunologiae et therapiae experimentalis	13.71	Polsko
Acta parasitologica	13.61	Polsko
Polish journal of pharmacology	12.680	Polsko
Polish journal of environmental studies	12.130	Polsko
Acta biochimica Polonica	11.990	Polsko
Cellular and molecular biology letters	11.820	Polsko
Folia histochemica et cytobiologica	11.770	Polsko
Acta palaeontologica Polonica	11.560	Polsko

Polish journal of chemistry	11.370	Polsko
Nukleonika	11.340	Polsko
Chemia analityczna	11.180	Polsko
Acta protozoologica	11.180	Polsko
Folie neuropathologica	11.170	Polsko
Folia biologica (Polsko)	11.090	Polsko
Bulletin of the Polish academy of Sciences - chemistry	10.580	Polsko
Biology of sport	10.260	Polsko

Human reproduction	40.80	Velká Británie
Molecular human reproduction	36.44	Velká Británie
European journal of clinical investigation	34.83	Velká Británie
Gynecological endocrinology	14.43	Velká Británie

International journal of molecular medicine	24.28	Řecko
Haema	6.070	Řecko

Acta paediatrica	24.14	Švédsko
Neuroendocrinology letters	22.94	Švédsko

Acta obstetrica et gynecologica scandinavica	20.82	Dánsko
--	-------	---------------

European journal of dermatology	14.69	Francie
---------------------------------	-------	----------------

European journal of obsterics, Gynecology and reproductive biology	14.46	Irsko
International journal of gynaecology and obstetrics	12.460	Irsko

Journal of cellular and molecular medicine	11.770	Rumunsko
--	--------	-----------------

Acta biologica hungarica	11.120	Maďarsko
Models in chemistry (dřive Acta chimica Hungarica)	10.000	Maďarsko
International urology and nephrology	6.930	Maďarsko
Pathology oncology research	6.720	Maďarsko
Acta physiologica hungarica	6.570	Maďarsko
Acta microbiologica et immunologica hungarica	5.970	Maďarsko
Acta chirurgica hungarica	5.870	Maďarsko

Nejlepřích 100 časopisů v Index Copernicus (hodnoty za rok 2003)

Název časopisu	Index Copernicus	země
Journal of clinical endocrinology and metabolism, the	56.07	USA
Human reproduction	40.80	Velká Británie
Fertility and sterility	39.82	USA
Journal of the Society for Gynecologic Investigation	37.92	USA
Molecular human reproduction	36.44	Velká Británie
European journal of clinical investigation	34.83	Velká Británie
Journal of physiology and pharmacology	26.83	Polsko
International journal of molecular medicine	24.28	Řecko
Acta paediatrica	24.14	Švédsko
Neuroendocrinology letters	22.94	Švédsko
Acta obstetrica et gynecologica scandinavica	20.82	Dánsko
Physiological research	16.18	Česká rep.
Acta neurobiologiae experimentalis	15.49	Polsko
Annals of agricultural and environmental medicine	14.72	Polsko
European journal of dermatology	14.69	Francie
European journal of obsterics, Gynecology and reproductive	14.46	Irsko

biology		
Gynecological endocrinology	14.43	Velká Británie
Acta theriologica	14.20	Polsko
Archivum immunologiae et therapiae experimentalis	13.71	Polsko
Acta parasitologica	13.61	Polsko
Collection of Czechoslovak chemical communications	13.61	Česká rep.
Folia microbiologica	12.910	Česká rep.
Western journal of medicine	12.890	USA
Polish journal of pharmacology	12.680	Polsko
International journal of gynaecology and obstetrics	12.460	Irsko
Polish journal of environmental studies	12.130	Polsko
Acta biochimica Polonica	11.990	Polsko
Cellular and molecular biology letters	11.820	Polsko
Folia histochemica et cytobiologica	11.770	Polsko
Journal of cellular and molecular medicine	11.770	Rumunsko
Acta palaeontologica Polonica	11.560	Polsko
Polimery	11.480	Polsko
Polish journal of chemistry	11.370	Polsko
Folia biologica	11.360	Česká rep.
Nukleonika	11.340	Polsko
Homeostatis in health and disease	11.180	Česká rep.
Chemia analityczna	11.180	Polsko
Acta protozoologica	11.180	Polsko
Folie neuropathologica	11.170	Polsko
Acta biologica hungarica	11.120	Maďarsko
Folia biologica (Polsko)	11.090	Polsko
Folia parasitologica (Česká republika)	11.080	Česká rep.
Journal of animal and feed sciences	10.630	Polsko

Bulletin of the Polish academy of Sciences - chemistry	10.580	Polsko
Medycyna weterynaryjna	10.530	Polsko
Acta physiologiae plantarum	10.490	Polsko
Bulletion of the Veterinary institute in Pulawy	10.450	Polsko
Acta biologica Cracoviensia series botanika	10.450	Polsko
Acta societatis botanicorum poloniae	10.400	Polsko
Biology of sport	10.260	Polsko
Československá psychologie	10.240	Česká rep.
Przemysl chemiczny	10.220	Polsko
Chemické listy	10.200	Česká rep.
Acta physica hungarica	10.190	Maďarsko
Fibres and textiles in Eastern Europe	10.110	Polsko
Inzynieria chemiczna i procesowa	10.070	Polsko
Česká a slovenská neurologie a neurochirurgie	10.060	Česká rep.
Acta chromatographia	10.000	Polsko
Models in chemistry (dříve Acta chimica Hungarica)	10.000	Maďarsko
Medical Science monitor	9.850	Polsko
Orvosi hetilap	8.660	Maďarsko
International journal of occupational safety and ergonomics	8.290	Polsko
Annales academiae medicae bialostocensis	7.960	Polsko
Nuclear medicine review	7.910	Polsko
Folia morphologica	7.660	Polsko
Postepy higieny i medycyny Doswiadczalnej	7.210	Polsko
Acta microbiologica polonica	7.010	Polsko
International urology and nephrology	6.930	Maďarsko
Psychiatria Polska	6.920	Polsko

Journal applied genetics	6.870	Polsko
Central european journal of public health	6.870	Česká rep.
Pathology oncology research	6.720	Maďarsko
Acta chirurgiae plasticae	6.670	Česká rep.
Acta poloniae pharmaceutica - drug research	6.620	Polsko
Polish journal of pathology	6.570	Polsko
Acta physiologica hungarica	6.570	Maďarsko
Annals of transplantation	6.490	Polsko
Postepy biochemii	6.420	Polsko
Klinika oczna	6.420	Polsko
Iranian journal of medical sciences	6.340	Irán
Polski merkuriusz lekarski	6.320	Polsko
Annals of hepatology	6.280	Mexiko
Polskie archiwum medycyny wewnetrznej	6.180	Polsko
Polish journal of radiology	6.120	Polsko
Medycyna pracy	6.120	Polsko
Wiadomosci lekarskie	6.090	Polsko
Otolaryngologia Polska	6.080	Polsko
Central european journal of immunology	6.070	Polsko
Haema	6.070	Řecko
Časopis lékařů českých	6.020	Česká rep.
Chirurgia narzadow ruchu i ortopedia Polska	6.020	Polsko
Folia cardiologica	6.020	Polsko
Acta microbiologica et immunologica hungarica	5.970	Maďarsko
Reproductive biology	5.970	Polsko
Medycyna doswiadczalna i mikrobiologia	5.930	Polsko
Postepy biologii komorki	5.920	Polsko
Acta chirurgica hungarica	5.870	Maďarsko
Polski przeglad chirurgiczny	5.840	Polsko

Otorinolaryngologie a foniatrie	5.820	Česká rep.
Vnitřní lékařství	5.820	Česká rep.

Nejlepší časopisy dle země a Index Copernicus

název	index copernicus	země
Human reproduction (Oxford)	40.800	Velká Británie
European journal of clinical investigation	34.830	Velká Británie
Gynecological endocrinology	14.430	Velká Británie

název	index copernicus	země
Klinik psikofarmakoloji Bulteni (Bulletin of clinical psychopharmacology)	4.780	Turecko
Journal of pediatric neurology	4.540	Turecko
Türk otolarenoloji Arsivi	3.230	Turecko
Yeni symposium (New symposium)	2.560	Turecko
Firat Medical Journal (Firat Tıp Dergisi)	0.000	Turecko
Nörolojik bilimler dergisi (Journal of neurological sciences (Turkish))	0.000	Turecko
Neuroanatomy	0.000	Turecko
Türkie klinikleri tip etigi hukuku tarihi	0.000	Turecko
Türkie klinikleri Týp bilimleri dergisi (Türkie klinikleri Journal of medical Sciences)	0.000	Turecko
Bagimlilik dergisi (Journal of Dependence)		Turecko
Journal of sports science and medicine		Turecko

název	index copernicus	země
Acta paediatrica	24.140	Švédsko
Neuroendocrinology letters	22.940	Švédsko

název	index copernicus	země
European journal of anatomy	5.670	Španělsko
Trastornos adictivos	0.000	Španělsko

název	index copernicus	země
Acta stomatologica Naissi	0.000	Srbsko a Černá hora
Medicus	0.000	Srbsko a Černá hora

název	index copernicus	země
International journal of molecular medicine	24.280	Řecko
Haema	6.070	Řecko
Archives of hellenic medicine	4.880 předpoklad	Řecko
Hellenic journal of nuclear medicine	4.480	Řecko
Hepato-gastroentetology	0.000	Řecko
Iatriki	0.000	Řecko
Primary health care	0.000	Řecko
Stomatologia	0.000	Řecko

název	index copernicus	země
Journal of cellular and molecular medicine	11.770	Rumunsko
Romanian journal of internal medicine	0.000	Rumunsko

Polsko – 20 nejlepších časopisů

název	index copernicus	země
Journal of physiology and pharmacology	26.830	Polsko
Acta neurobiologiae experimentalis	15.490	Polsko

Annals of agricultural and environmental medicine	14.720	Polsko
Archivum immunologiae et therapiae experimentalis	13.710	Polsko
Acta parasitologica	13.610	Polsko
Polish journal of pharmacology	12.680	Polsko
Polish journal of environmental studies	12.130	Polsko
Acta biochimica Polonica	11.990	Polsko
Celluar and molecular biology letters	11.820	Polsko
Folia histochemica et cytobiologica	11.770	Polsko
Polish journal of chemistry	11.370	Polsko
Nukleonika	11.340	Polsko
Chemia analityczna	11.180	Polsko
Folia neuropathologica	11.170	Polsko
Folia biologica (Polsko)	11.090	Polsko
Acta biologica cracoviensia series Botanica	10.450	Polsko
Biology of sport©	10.260	Polsko
Przemysl chemiczny	10.220	Polsko
Inzynieria chemiczna i procesowa	10.070	Polsko

Maďarsko - 20 nejlepších časopisů

název	index copernicus	země
Acta biologica hungarica	11.120	Maďarsko
Models in chemistry (dříve Acta chimica Hungarica)	10.000	Maďarsko
Orvosi hetilap	8.660	Maďarsko
International urology and nephrology	6.930	Maďarsko
Acta physiologica hungarica	6.570	Maďarsko
Acta microbiologica et immunologica Hungarica	5.970	Maďarsko
Acta chirurgica hungarica	5.870	Maďarsko
Neurobiology	5.620	Maďarsko

Magyar onkologia (Hungarian oncology)	4.480	Maďarsko
Nogyogyaszati onkologia	4.230	Maďarsko
Central european journal of occupational and environmental medicine	3.880	Maďarsko
Clinical neuroscience (Ideggyogyaszati Szemle)	3.880	Maďarsko
Magyar Noorvosok Lapja (Hungarian journal of obstetrics and gynaecology)	3.730	Maďarsko
Osteologiai Kozlemenyek	3.690	Maďarsko
Medicina thoracalis	3.580	Maďarsko
Psychiatria hungarica	3.580	Maďarsko
Anthropologiai Kozlemenyek	3.330	Maďarsko
Magyar reumatologia (Hungarian rheumatology)	3.330	Maďarsko
Magyar radiológia	3.230	Maďarsko
Gyermeckgyogyaszat	3.180	Maďarsko

název	index copernicus	země
Medicina	5.170	Litva
Seminars in cardiology	4.580	Litva
Acta medica lithuanica	4.430	Litva
Medicina teorija ir Praktika	3.530	Litva
Lietuvos chirurgija (Lithuanian surgery)		Litva

název	index copernicus	země
POL. It psychiatry on line Italia	0.000	Itálie

název	index copernicus	země
European journal of obstetrics, gynecology and reproductive biology	14.460	Irsko
International journal of	12.460	Irsko

gynaecology and obstetrics		
Transplantation proceedings	0.000	Irsko

název	index copernicus	země
Medicina	0.000	Chorvatsko

název	index copernicus	země
European journal of dermatology	14.690	Francie
Epileptic disorders	5.120	Francie
Virology	4.330	Francie
Epilepsies	4.100	Francie
L'information psychiatrique	3.430	Francie
Psychologie & Neuropsychiatrie du Vieillessement	0.000	Francie
Sciences sociales et sante	0.000	Francie

název	index copernicus	země
Papers on anthropology	0.000	Estonsko

název	index copernicus	země
Acta obstetrica et gynecologica scandinavica	20.820	Dánsko

Physiological research (dříve Physiologia Bohemoslovaca)	16.180	Česká rep.
Collection of czechoslovak chemical communications	13.610	Česká rep.
Folia microbiologica	12.910	Česká rep.
Folia biologica (Czech Republic)	11.360	Česká rep.
Homeostasis in health and disease	11.180	Česká rep.
Folia parasitologica (Czech republic)	11.080	Česká rep.
Chemické listy	10.200	Česká rep.

Česká a slovenská neurologie a neurochirurgie	10.060	Česká rep.
Central European journal of public health	6.870	Česká rep.
Acta chirurgiae plasticae	6.670	Česká rep.
Časopis lékařů českých	6.020	Česká rep.
Otorinolaryngologie a foniatrie	5.820	Česká rep.
Vnitřní lékařství	5.820	Česká rep.
Česká gynekologie	5.420	Česká rep.
Česko-slovenská dermatologie	5.370	Česká rep.
Česká revmatologie	5.070	Česká rep.
Česká a slovenská gastroenterologie	4.530	Česká rep.
Czech and Slovak gastroenterologie	4.530	Česká rep.
Anesteziologie a nekladná péče	3.430	Česká rep.
Klinická mikrobiologie a infekční lékařství	3.380	Česká rep.
Pracovní lékařství	3.380	Česká rep.
Klinická onkologie	3.180	Česká rep.
Hygiena	3.130	Česká rep.
Praktická flebologie	2.840	Česká rep.
Cesko-slovenská pediatrie	2.390	Česká rep.
Neonatologické listy	1.740	Česká rep.
Gynekolog	1.640	Česká rep.
Lékař a technika	1.190	Česká rep.

název	index copernicus	země
Pathology oncology research	6.720	neurčeno

40 nejlepších evropských časopisů v Index Copernicus

název	index copernicus	země
Human reproduction (Oxford)	40.800	Velká Británie
European journal of clinical investigation	34.830	Velká Británie

Journal of physiology and pharmacology	26.830	Polsko
International journal of molecular medicine	24.280	Řecko
Acta paediatrica	24.140	Švédsko
Neuroendocrinology letters	22.940	Švédsko
Acta obstetrica et gynecologica scandinavica	20.820	Dánsko
Physiological research (dříve Physiologia Bohemoslovaca)	16.180	Česká rep.
Acta neurobiologiae experimentalis	15.490	Polsko
Annals of agricultural and environmental medicine	14.720	Polsko
European journal of dermatology	14.690	Francie
European journal of obstetrics, gynecology and reproductive biology	14.460	Irsko
Gynecological endocrinology	14.430	Velká Británie
Archivum immunologiae et therapiae experimentalis	13.710	Polsko
Acta parasitologica	13.610	Polsko
Collection of czechoslovak chemical communications	13.610	Česká rep.
Folia microbiologica	12.910	Česká rep.
Polish journal of pharmacology	12.680	Polsko
International journal of gynaecology and obstetrics	12.460	Irsko
Polish journal of environmental studies	12.130	Polsko
Acta biochimica Polonica	11.990	Polsko
Celluar and molecular biology letters	11.820	Polsko
Folia histochemica et cytobiologica	11.770	Polsko
Journal of celluar and molecular medicine	11.770	Rumunsko
Polish journal of chemistry	11.370	Polsko
Folia biologica (Czech Republic)	11.360	Česká rep.
Nukleonika	11.340	Polsko

Chemia analityczna	11.180	Polsko
Homeostasis in health and disease	11.180	Česká rep.
Folia neuropathologica	11.170	Polsko
Acta biologica hungarica	11.120	Maďarsko
Folia biologica (Polsko)	11.090	Polsko
Folia parasitologica (Czech republic)	11.080	Česká rep.
Acta biologica cracoviensia series Botanica	10.450	Polsko
Biology of sport	10.260	Polsko
Przemysl chemiczny	10.220	Polsko
Chemické listy	10.200	Česká rep.
Inzynieria chemiczna i procesowa	10.070	Polsko
Česká a slovenská neurologie a neurochirurgie	10.060	Česká rep.
Acta chromatographica	10.000	Polsko

Země	Počet
Polsko	20
Česká republika	8
Velká Británie	3
Svédsko	2
Irsko	2
Řecko	1
Dánsko	1
Francie	1
Rumunsko	1
Maďarsko	1