

Troppo o troppo poco? Web of science, Scopus, Google scholar: tre database a confronto

di Ezio Tarantino

Il contesto

Nel 1999 venne istituito presso l'Università "La Sapienza" di Roma il progetto-servizio BIDS, Biblioteca digitale de "La Sapienza"¹. Nel corso del suo primo anno di vita vennero acquistate 10 banche dati e stipulato il primo contratto consortile con un grande editore di periodici elettronici (Elsevier).

Il progetto è via via cresciuto fino a raggiungere, oggi, circa 90 banche dati e oltre 15.000 periodici elettronici. Fra quei 10 iniziali repertori elettronici c'era anche Web of science, la banca dati prodotta dall'ISI, Institute for Scientific Information², acquistata in un "pacchetto" comprendente anche i Current contents e il Journal of citation reports (JCR).

All'epoca Web of science (WoS) era la sola risorsa a fornire, oltre a un ricco *database* bibliografico, uno strumento di analisi citazionale, molto apprezzato dall'utenza accademica.

Il suo funzionamento è molto elementare, sebbene di non semplice realizzazione, specialmente in un'epoca in cui le riviste scientifiche venivano pubblicate esclusivamente su carta. Accanto a ogni articolo indicizzato all'interno della banca dati l'utente può immediatamente visualizzare da quanti altri articoli è stato citato. E con un *clic* WoS consente anche di visualizzare quali essi siano.

Tuttavia l'abbonamento a WoS, Current contents e JCR (che eravamo obbligati ad acquistare insieme), sulla base delle statistiche d'uso che dimostravano un utilizzo piuttosto limitato, costituiva un impegno economico poco sostenibile. Infatti, a fronte di un utilizzo nemmeno confrontabile, per numero di ricerche effettuate, con le banche dati più consultate - Psycinfo e Medline - il costo per ricerca era molto elevato: WoS e Current contents 63 centesimi di euro per ricerca, rispetto ai 4 centesimi e 3 centesimi di euro per ricerca effettuata rispettivamente di Psycinfo e Medline (figure 1 e 2).

EZIO TARANTINO, BIDS, Biblioteca digitale dell'Università degli studi di Roma "La Sapienza", e-mail ezio.tarantino@uniroma1.it. L'articolo riprende, con alcune modifiche, il testo di un intervento all'"ICOLC Autumn meeting", Pozna, 29 settembre-1° ottobre 2005. L'ultima consultazione dei siti Web citati risale al 10 aprile 2006.

1 <<http://bids.citicord.uniroma1.it>>.

2 <<http://www.isinet.com/>>.

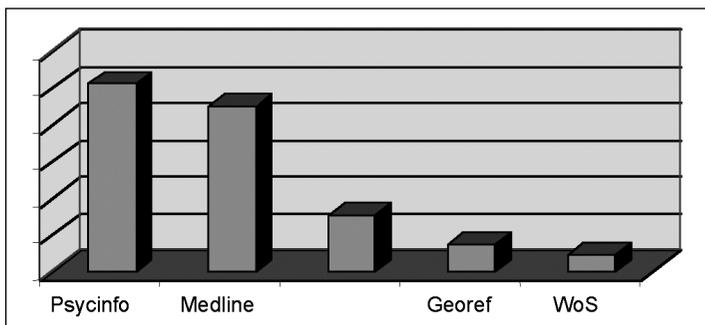


Fig. 1: Rappresentazione grafica del numero di ricerche per mese delle principali banche dati nell'anno 2002 (dati non dichiarati per motivi di riservatezza)

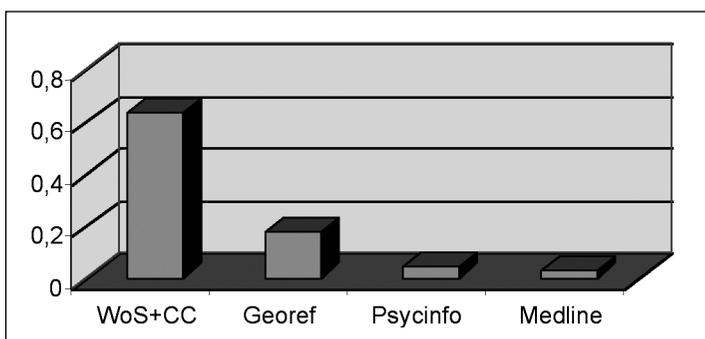


Fig. 2: Euro per ricerca (2002)

Nel 2003, allo scadere del contratto, la BIDS decise di non rinnovare l'abbonamento a WoS.

Questa decisione non fu indolore. Il nostro sospetto, che la consultazione di WoS si limitasse principalmente all'utilizzo del suo potente e ancora inimitato strumento di *citation index*, fu confermato. I nostri utenti infatti di questo si lamentarono, non di disporre di uno strumento bibliografico in meno per i loro progetti di ricerca.

Le domande, del tutto legittime, dei nostri ricercatori («Quante volte è stato citato il mio articolo?») esigevano una risposta.

Così, quando nel luglio del 2004 Elsevier ci propose di provare, senza costi né impegni da parte nostra, il nuovo *database* Scopus³, accettammo, perfezionando poi l'acquisto nel gennaio del 2005. Non perché la Biblioteca digitale avesse bisogno di un nuovo *database* multidisciplinare, ma perché avevamo bisogno di un nuovo strumento di *citation indexing* aggiornato.

Scopus si autodescrive come «il punto di accesso più completo nel campo scientifico, medico, tecnico e delle scienze sociali, comprendente tutta la letteratura più importante»⁴. La prima analisi comparativa che provammo a fare fu dunque quella

³ <<http://info.scopus.com/>>.

⁴ Per una descrizione analitica e critica delle numerose funzionalità di Scopus, della consistenza del suo archivio e della nuova *release* di Web of science si rimanda all'articolo di Goodman e Deis su «Charleston advisor» [5] e, fra gli altri, al più conciso ma esauriente articolo di Susan Fingerman [4].

sulla consistenza delle riviste spogliate da Scopus, per verificare se, sulla base di questa affermazione, non fosse possibile dismettere qualcuna delle banche dati disciplinari alle quali eravamo abbonati.

I risultati, come mostrati dalla tabella, non furono soddisfacenti. Le sovrapposizioni (con la parziale eccezione di Psycinfo) risultano infatti piuttosto limitate.

Tab. 1: Percentuale di sovrapposizione fra Scopus e alcune banche dati su piattaforma CSA⁵

Database	n. di fonti	n. di fonti in comune con Scopus	Percentuale di sovrapposizione
Psycinfo	1960	1270	64,75%
ASFA	5434	2005	36,89%
Econlit	973	377	38,74%
Biological sciences	6426	2810	43,72%
Sociological abstracts	1865	751	40,80%

Nel frattempo, annunciato come l'arrivo del *beaujolais nouveau*⁶, nel novembre del 2004 fece il suo ingresso sulla scena Google scholar⁷, il nuovo motore di Google per le ricerche accademiche. Non il primo esempio di questo tipo (da citare almeno l'esperienza, settoriale, di CiteSeer⁸ e quella di Scirus⁹, motore di ricerca specializzato nella documentazione scientifica, prodotto anche questo da Elsevier, e inserito all'interno di Scopus): ma la popolarità, l'autorevolezza e l'ambizione del progetto di Google ne hanno fatto immediatamente un reale, temibile concorrente. Così, in pochi mesi, WoS si trovò faccia a faccia non con uno, bensì due competitori, uno dei quali, per di più, gratuito.

Un'analisi quantitativa

Da un punto di vista strettamente quantitativo i requisiti più significativi per un database di queste dimensioni e caratteristiche sono senza dubbio il numero di riviste e di articoli indicizzati, l'arco temporale della copertura, la coerenza e la distribuzione delle fonti nelle diverse discipline.

Per quello che riguarda il numero di articoli indicizzati, WoS ne dichiara, più o meno, 35 milioni, Scopus 25 milioni (da 14.000 riviste *peer-reviewed*), mentre non conosciamo le dimensioni dell'archivio di Google scholar. Così come non ne conosciamo l'arco temporale (Google è tradizionalmente molto abbottonato nel rivelare le caratteristiche tecnologiche e di contenuto dei suoi archivi), mentre WoS (nella versione standard) fa cominciare l'indicizzazione completa dei documenti dal 1990 e Scopus dal 1996.

Ma mentre utilizzando WoS non è possibile in nessun caso poter rintracciare articoli precedenti il 1990 (a meno di acquistare a parte i *backfile*), in Scopus è possibile risalire ad articoli anche molto più vecchi (si parte dagli anni Sessanta), a condizione (parrebbe, perché la caratteristica non è dichiarata ufficialmente) che questi siano citati in articoli pubblicati successivamente al 1996.

⁵ Il confronto è stato effettuato comparando i rispettivi file sulla base dell'ISSN della rivista elencata come fonte. Le banche dati utilizzate sono le uniche che, su CSA, elencano le fonti complete di ISSN.

⁶ «Nature», 432 (2004), n. 7016, p. 421.

⁷ <<http://scholar.google.com>>.

⁸ <<http://citeseer.ist.psu.edu/>>.

⁹ <<http://www.scirus.com>>

Si tratta di una scelta che certamente non garantisce uniformità e coerenza dell'archivio¹⁰, ma che in qualche misura sembrerebbe corrispondere a un criterio di qualità: infatti questo significa che nella complessa e di sicuro non breve attività di recupero della documentazione scientifica meno recente, viene data la precedenza a quegli articoli che, pur essendo vecchi di molti anni, sono ancora citati: si può sospettare che siano quindi quelli di rilevanza scientifica maggiore.

Tuttavia in questi casi la citazione non è quasi mai completa: è infatti limitata ai dati fondamentali (autore, titolo, titolo della rivista, data della pubblicazione). Quasi sempre mancano sia l'abstract sia la lista delle citazioni. Tuttavia è possibile, a partire da questi, risalire a tutti gli articoli (indicizzati da Scopus) che li citano.

In termini di coerenza dunque WoS è senza dubbio più accurato: tutti gli articoli sono trattati nello stesso modo. Tutti sono collegati agli articoli che li citano e tutti hanno una lista completa di *reference*. Non ci sono sorprese.

L'analisi che è stata condotta nella primavera del 2005 fra i tre *database*, non ha preso in considerazione aspetti quali l'usabilità o la presenza o meno di strumenti accessori nei tre prodotti (la cui utilità, o quantomeno il cui utilizzo presso un utente medio di un ateneo italiano, è tutta da verificare: parlo degli *alert*, della configurazione di un profilo personalizzato ecc.) per due fondamentali ragioni: primo, perché non sarebbe corretto giudicare Google scholar su questi parametri (l'interfaccia e gli strumenti "avanzati" mantengono l'efficacia spartana del marchio Google: sarebbe come voler comparare una utilitaria con una Ferrari – salvo scoprire che l'utilitaria va a 200 all'ora!); e perché non è stato possibile valutare la nuova interfaccia di WoS (Web of knowledge) poiché, come detto, l'Università "La Sapienza" aveva cancellato la sottoscrizione nel 2003, il che avrebbe chiaramente sbilanciato il confronto a vantaggio di quella di Scopus che anche il severo Goodman ha giudicato come particolarmente moderna ed efficace¹¹.

Stabilito che la richiesta della nostra utenza è focalizzata sulla necessità di disporre di un efficace e aggiornato strumento citazionale, il test comparativo è stato fondamentalmente condotto per valutare questa specifica caratteristica dei 3 prodotti.

Abbiamo cercato dieci articoli per ciascuna delle quattro fondamentali discipline presenti nella nostra università, compresa quella economica che non è fra quelle ufficialmente "coperte" dall'indicizzazione di Scopus. Per un totale, quindi, di quaranta interrogazioni. Gli articoli sono stati scelti in modo casuale in un arco di anni che va dal 1968 (un articolo) al 2003, dagli archivi di periodici elettronici disponibili presso la nostra università e dai più importanti repertori disciplinari.

Le risposte sono state positive (l'articolo è stato trovato) nell'85% dei casi in Scopus (34 articoli rintracciati), nel 52,5% dei casi in WoS (21 articoli) e nel 95% dei casi in Google scholar (38 articoli).

Per quanto riguarda il numero complessivo di citazioni associate a questi articoli, Scopus ne totalizza 292, WoS 362 mentre Google scholar arriva a 500.

È interessante notare come questi dati siano distribuiti fra le diverse discipline.

Google scholar raggiunge un ottimo risultato nel campo delle scienze economiche (187 citazioni, rispetto alle 37 di Scopus e alle 33 di WoS: e qui è significativo come Scopus, promosso come un database specializzato nei settori disciplinari STM – *science*

¹⁰ Su questo aspetto è stato particolarmente severo Goodman nel suo articolo su Charleston Advisor: «Thus we find it especially unprofessional that Elsevier has released this material without a comment or alert anywhere in the product or the online help» [5, p. 9].

¹¹ «The results pages for Scopus are the best in both appearance and function that the reviewers have ever seen in a bibliographic database» [5, p. 15].

technology medicine – sopravanza WoS che pure ha una sezione dedicata al settore delle scienze sociali); WoS eccelle nelle scienze fisiche, matematiche e chimiche (156), mentre Scopus è più equilibrato e non si mette in luce in nessuna particolare disciplina.

Tab. 2: Numero di citazioni per disciplina

Disciplina	Scopus n. citazioni	WoS n. citazioni	Google scholar n. citazioni
Economia	37	33	187
Medicina e biologia	74	54	102
Psicologia e psichiatria	93	119	71
Scienze (fisica, chimica)	88	156	140
Totale	292	362	500

Nel rapporto fra il numero complessivo di citazioni e articoli trovati, WoS si rivela lo strumento più efficace, Scopus quello meno.

Tab. 3: Media delle citazioni per numero di articoli trovati

Database	Articoli	N. citazioni	Media
WoS	21	362	17,24
Google scholar	38	500	13,16
Scopus	34	292	8,59

Alla luce di quanto detto in precedenza, riguardo la capacità di Scopus di andare indietro nel tempo (a costo zero, senza acquistare a parte i *backfile*), è ugualmente importante verificare come le risposte positive si distribuiscano in diversi archi temporali.

Scopus ha trovato tutti e 5 gli articoli antecedenti al 1990 che abbiamo cercato.

Google scholar dal canto suo ha recuperato tutti e 28 gli articoli successivi al 1996, mentre WoS non ha ovviamente recuperato gli articoli antecedenti al 1990 e ne ha recuperati solo 2 sui 7 cercati pubblicati fra il 1990 e il 1996¹².

Tab. 4 : Articoli trovati in diversi archi temporali

Anno di pubblicazione	Scopus	WoS	Google scholar
<1990	5 (5)	0 (5)	4 (5)
1991-1996	4 (7)	2 (7)	6 (7)
>1996	25 (28)	19 (28)	28 (28)

Un'analisi qualitativa

Abbiamo quindi cercato di entrare nel merito dei risultati, selezionandone alcuni sulla base della curiosità generata da risposte estremamente incoerenti fra le tre risor-

¹² È da notare come anche l'interessante analisi di Bauer e Bakkalbasi [1] pubblicata nel settembre del 2005, quindi successivamente al nostro test, ed effettuata su un numero più elevato di dati, sottolinei la maggiore efficacia di Google scholar nel recupero di dati recenti (mentre WoS e Scopus più o meno si equivalgono). Questa è dovuta al più efficiente recupero di informazioni da pubblicazioni *open access* quali archivi disciplinari o istituzionali, o da atti di convegni, *dissertation* ecc. Come detto, Scopus permette l'interrogazione del Web, tramite il suo motore di ricerca Scirus, ma ai fini dell'analisi citazionale non tiene conto di questi risultati.

se. Il test è stato realizzato non per valutare la “bontà” dell’informazione, ma il funzionamento dei tre *database*, indagare le eventuali differenze delle risposte a medesima interrogazione e cercare di trovarne una spiegazione.

Qui di seguito, vediamo quattro esempi che non vogliono naturalmente esaurire una casistica numerosa e spesso contraddittoria, e non possono essere sufficienti per stilare graduatorie o giudizi, ma che riteniamo particolarmente significativi soprattutto nel valutare come meccanismi automatici di indicizzazione, quali quelli cui ormai tutti i produttori di banche dati fanno ricorso, possono portare abbastanza spesso a risultati non coerenti, approssimativi e, in definitiva, fuorvianti per l’utente.

Esempio 1.

Lois Wladis Hoffman. *The influence of the family environment on personality: accounting for sibling differences*. «Psychological Bulletin», 110 (1991), n. 2, p. 187-203.

Questo articolo, pubblicato nel 1991, risulta essere stato citato 51 volte secondo Scopus, 81 volte per WoS e 33 per Google scholar. È interessante analizzare il numero di citazioni uniche all’interno dei tre *database*. Delle 51 citazioni registrate da Scopus, 3 sono “uniche”, cioè non sono state elencate da nessuno degli altri due; in Google ci sono 4 citazioni “uniche” e 35 (ovviamente, dato l’alto numero di citazioni) in WoS.

Aggregando questi dati abbiamo che Scopus e WoS condividono 43 record (su un massimo possibile di 51, cioè il numero di risposte di Scopus), WoS e Google ne condividono 25 (su un massimo possibile di 33); infine Scopus e Google ne condividono 26 (su 33).

Deduplicando i risultati ne consegue che WoS e Scopus insieme forniscono complessivamente 89 citazioni (le 81 di WoS più le 51 di Scopus, meno le 43 in comune).

Lo stesso risultato, 89, lo otteniamo sommando i record di WoS e Google ed eliminando i record duplicati [(81+33)-25], malgrado Google avesse indicizzato un numero di gran lunga inferiore di articoli, rispetto a Scopus (33 contro 51).

Apparentemente dunque l’accoppiata WoS + Google scholar in questo caso fornisce un maggior numero di risultati.

Ma questo è in realtà poco significativo. Infatti la domanda da farsi è: questa disparità di risultati è un vantaggio o uno svantaggio per l’utente?

In linea puramente teorica l’*optimum* non dovrebbe essere costituito dal 100% di sovrapposizione fra i diversi strumenti bibliografici? Un determinato articolo non dovrebbe teoricamente essere citato lo stesso numero di volte, indipendentemente dall’agenzia che effettua lo spoglio? Come mai questa differenza fra i risultati? Chi decide quale sia l’informazione corretta? L’utente? Il bibliotecario? O il fornitore del servizio?

Se poi si ordinano i risultati per numero di citazioni (vale a dire in base ai più citati tra gli articoli che a loro volta citavano quello da cui eravamo partiti¹³) è possibile osservare altri risultati interessanti. In Google scholar 8 dei 26 articoli condivisi con Scopus appaiono nella prima pagina (com’è noto, la più letta dagli utenti). D’altra parte in Scopus solo 4 di questi 26 articoli appaiono in prima pagina (e per di più, a parte uno, non nelle prime posizioni: in seconda, ottava, nona e decima). Questo significa che non solo Google scholar ha trovato, in questo caso, meno articoli di Scopus, ma non ha trovato quelli che, secondo Scopus, sarebbero i più citati. Come mai?

Non siamo in grado di rispondere¹⁴.

¹³ In Google scholar questo ordinamento è di default.

¹⁴ Non è stato possibile replicare il confronto anche con WoS perché, almeno nella versione in uso presso l’università “La Sapienza” non è possibile ordinare le citazioni a loro volta in base al numero di citazioni ricevute.

Esempio 2.

Marco Frittelli. *The minimal entropy martingale measure and the valuation problem in incomplete markets*. «Mathematical finance», 10 (2000), n. 1, p. 39-52.

In questo secondo esempio la stranezza sta nel fatto che a giudizio di Google, l'articolo sarebbe stato citato 116 volte, rispetto alle 36 di Scopus e alle 32 di WoS.

Incuriosito dall'alto numero di citazioni registrate da Google ho voluto vederci chiaro.

In questo caso Google scholar e Scopus condividono 30 articoli. Scopus cita 9 dei primi 10 articoli trovati da Google scholar.

Su Google, invece, solo 4 fra i primi 10 sono in comune. Sembra essere il caso opposto a quello precedente. Scopus in questo caso non ha rintracciato gli articoli che citano l'articolo di partenza e che Google propone in terza, quinta e settima posizione (quindi a loro volta molto citati).

Come mai?

Limitiamoci alla pagina uno. Per quanto riguarda il primo di questi, la risposta è semplice quanto poco rassicurante: Scopus non lo cita perché questo articolo, andando a leggerlo, non cita affatto *Minimal entropy martingale measure*. Non c'è nella bibliografia finale e neppure nel testo. Come mai Google lo ha inserito invece fra gli articoli che lo referenziano?

Il secondo caso è della stessa specie: facile da capire, meno da giustificare. L'articolo è infatti pubblicato dalla rivista «Finance and stochastics» ed è del 2001, quindi in un anno ampiamente coperto dalla banca dati di Elsevier.

Eppure, andando a cercare le proprietà del titolo, scopriamo che, inopinatamente, lo spoglio di questa rivista inizia a partire dal 2004 e non dal 1996, come ci si aspetterebbe.

L'ultimo caso è quello forse più semplice: la causa questa volta è un errore di stampa: l'articolo *Optimal investment in complete financial markets* che cita *Minimal entropy martingale measure* riporta in modo errato il nome del suo autore (Frittelli in luogo di Frittelli).

È lecito domandarsi: quanti errori ci si deve aspettare da sistemi di indicizzazione divenuti col tempo sempre più dipendenti da sofisticati meccanismi automatici¹⁵? È possibile ipotizzare una percentuale? E ancora: quanto è "buona" l'informazione che ci viene fornita (qualche volta a caro prezzo)? E quella gratuita chi la garantisce? Sta a noi bibliotecari, sta agli utenti controllare ogni singolo record in cerca di errori? Sta a noi verificare se la copertura di un determinato periodico parte dal 2001 anziché, come ci si aspetta, dal 1996? Per finire, l'utente chiederà forse leggermente irritato: «Perché non è citato il mio articolo?».

Esempio 3.

Erwin Bunzel – Julian M. Dust – François Terrier. *Rationalizing the regioselectivity in polynitroarene anionic sigma-adduct formation: relevance to nucleophilic aromatic substitution*. «Chemical reviews», 95 (1995), n. 7, p. 2261-2280.

L'articolo è del 1995, quindi precede l'anno di partenza ufficiale dell'indicizzazione di Scopus. In effetti presenta solo i dati essenziali, non riporta le *reference*. Sarebbe essere citato da 3 articoli. Ma per WoS gli articoli che lo citano sono 65.

Ho selezionato uno di questi 65 articoli elencati in WoS (fra i più recenti) e l'ho cercato in Scopus.

¹⁵ Meccanismi che hanno avuto una larghissima applicazione, grazie soprattutto al successo tributato dalla comunità editoriale a Crossref <<http://www.crossref.org/>>, un progetto nato con l'obiettivo di facilitare, sfruttando la tecnologia Openurl e il linguaggio di marcatura XML, i legami e i rimandi fra documenti e favorire l'interoperabilità. Per un'introduzione a Openurl e Crossref si veda [13].

Scopus lo trova e scopro che all'interno del record, fra le *reference*, è correttamente riportato l'articolo di partenza (da cui invece sembrava non essere citato), al quale vengono in questa sede attribuite 64 citazioni (non 3!).

Come è possibile? Forse in virtù di un evidente errore nella paginazione ("p. 2261", come si vede nella figura 3).

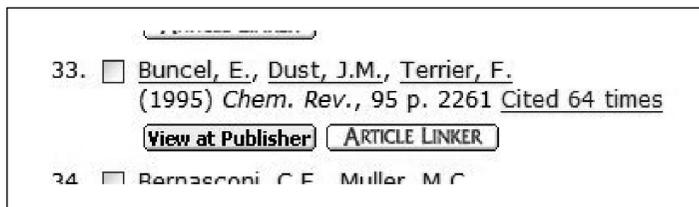


Fig. 3: La citazione dell'articolo in Scopus

No, la causa non può essere questa, perché anche i tre articoli che vengono indicati come gli unici che citino *Rationalizing the regioselectivity* riportano l'indicazione della pagina in modo corretto ("2261-2280").

La sola differenza è che in questi 3 casi la descrizione bibliografica è completa (figura 4), mentre in tutti gli altri casi si presenta in forma concisa. È questa la ragione? Forse. Forse no. Quello che è sicuro è che per Scopus questo articolo viene citato 67 volte, ma l'utente non lo saprà mai.

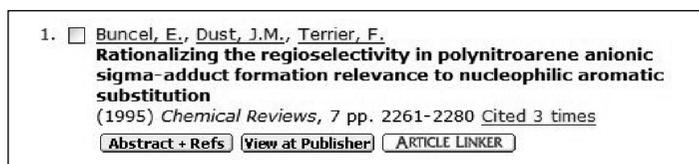


Fig. 4: La descrizione in forma completa dell'articolo in Scopus

Esempio 4.

Jean Paul Ryckaert – Giovanni Ciccotti – Herman J. C. Berendsen. *Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. «Journal of computational physics», 23 (1977), p. 327-341.

Trattandosi di un articolo del 1977 né Scopus né WoS lo rintracciano, e in questo non c'è nulla di anormale. Curiosamente Google scholar, al contrario, non solo lo trova (seppur sotto forma di citazione, senza link al *full text*: una eventualità non rarissima ma neppure molto frequente), ma gli attribuisce 1.276 citazioni. Appare a questo punto strana l'omissione in Scopus che, come abbiamo visto, degli articoli precedenti il 1996 indicizza quanto meno quelli citati in articoli posteriori a quella data. Oltretutto fra queste 1.276 citazioni ve ne sono anche di molto recenti.

Procedendo come nel caso di prima, cercando uno di questi articoli, e andando a visualizzare la sezione dedicata alle *reference*, effettivamente l'articolo c'è, e risulta essere citato non 1.276 volte, ma 2.292 volte.

E invece dell'articolo non v'è traccia, neppure cercando per nome dell'autore (sia "Ciccotti" che "Ryckaert", dei quali sono indicizzati numerosi articoli antecedenti il 1996), neppure raffinando la ricerca alle sole *reference* (quando invece abbiamo appena evidenziato come nelle *reference* la citazione compaia in migliaia di articoli).

Conclusioni

Il test non è stato realizzato per valutare la superiorità di un *database* rispetto a un altro (per questo occorre incrociare con attenzione tutta la bibliografia comparativa che sta uscendo con una certa regolarità negli ultimi mesi), quanto piuttosto per renderci conto di come questi strumenti, per quanto potenti e dai risultati comunque nel complesso più che soddisfacenti, non possano garantire uniformità, coerenza, affidabilità certa. Nel caso di Google scholar, senza che questo suoni come una bocciatura (l'analisi ne ha dimostrato anzi con ampiezza di documentazione il valore), è indispensabile tenere conto della impossibilità di determinare l'effettiva consistenza delle fonti, requisito fondamentale per la «validità di qualsiasi ricerca di documentazione in ambito accademico» [1]. Del resto non si può non sottolineare come la vastità del Web, sua unica incontrollabile fonte (con evidenti rischi di «rumore» di fondo e di duplicazioni, specialmente negli archivi aperti dei quali non è facile valutare né la qualità dell'informazione, né l'esclusività e l'originalità dei contenuti) e il ricorso a meccanismi di ricerca totalmente affidati ad automatismi in cui l'intervento umano è limitato alla (peraltro sofisticata e tuttora segreta) architettura logaritmica, rendano Google scholar un prodotto programmaticamente diverso dai suoi due concorrenti, ai quali si affianca, senza pretendere di sostituirsi.

Internet, e le tecnologie connesse (marcatura XML, uso «intelligente» di meta-dati) offre enormi possibilità, ma anche, proprio per la sua complessità, molte più occasioni di incappare in errori.

Anche la funzione tradizionale di intermediazione del bibliotecario è compromessa, poiché neppure a lui è data la possibilità di analizzare e interpretare i risultati in modo completo. Si può ipotizzare che la causa risieda nel nuovo tipo di approccio all'informazione: la biblioteca dell'accesso implica il mancato controllo diretto delle fonti e dei repertori, la cui attendibilità, almeno sul terreno del *citation indexing*, sembra messa in discussione con l'aumentare del loro numero.

D'altra parte l'ingresso nel mercato (o nel territorio libero della rete aperta a tutti, come nel caso di Google) di nuovi attori ha come conseguenza quella di aver messo in luce le lacune di un indice citazionale, a prescindere dall'agenzia che effettua la rilevazione, mettendo sull'avviso gli utenti sulla sua inevitabile imprecisione. E questo processo di relativizzazione di meccanismi automatici di indicizzazione (tanto importanti, specialmente nella definizione delle carriere accademiche) è da considerarsi, almeno a livello di presa di coscienza, come un risultato senz'altro positivo.

RIFERIMENTI BIBLIOGRAFICI

[1] Kathleen Bauer – Nisa Bakkalbasi. *An examination of citation counts in a new scholarly communication environment*. «D-Lib Magazine», 11 (2005), n. 9, <<http://www.dlib.org/dlib/september05/bauer/09bauer.html>>.

[2] Richard K. Belew. *Scientific impact quantity and quality: analysis of two sources of bibliographic data*. 2005. <<http://arxiv.org/abs/cs.IR/0504036>>.

[3] Declan Butler. *Science searches shift up a gear as Google starts scholar engine*. «Nature», 432 (2004), n. 7016, p. 423.

[4] Susan Fingerman. *Scopus: profusion and confusion*. «Online», 29 (2005), n. 2, p. 36-38.

[5] David Goodman – Louise Deis. *Web of science (2004 version) and Scopus*. «The Charleston Advisor», 6 (2004), n.3, p. 5-21. <<http://www.charlestonco.com/comp.cfm?id=43>>.

- [6] *Google scholar's impact on libraries*. [Edited by] Joe Buenker. <<http://www.west.asu.edu/jbuenke/librarianship/google-scholar.html>>.
- [7] Cheryl LaGuardia. *Scopus vs Web of science*. «Library Journal», 130 (2005), n 1, p. 40-42. <<http://www.libraryjournal.com/article/CA491154.html%22>>.
- [8] Philipp Mayr. *Google scholar: how deep does this search engine dig?* 2005. <<http://eprints.rclis.org/archive/00004499/>>
- [9] Péter Jacsó. *Google scholar (redux)*. «Péter's digital reference shelf» (2005). <<http://www.galegroup.com/reference/archive/200506/google.html>>.
- [10] Péter Jacsó, *ISI Web of science, Scopus and Sportdiscus*. «Online», 28 (2004), n. 6, p. 51-54.
- [11] *On Google scholar*. <<http://schoogle.blogspot.com/>>.
- [12] Gargan Prathap. *Who's afraid of research assessment?* «Current Science», 88 (2005), n. 1, p. 14-17. <<http://www.ias.ac.in/currensci/feb102005/331.pdf>>.
- [13] Herbert Van de Sompel – Oren Beit-Arie. *Open linking in the scholarly information environment using the Openurl framework*. «D-Lib Magazine», 7 (2001), n. 3. <<http://www.dlib.org/dlib/march01/vandesompel/o3vandesompel.html>>.
- [14] Maurizio Zani. *La rivincita delle citazioni*. «Bibliotime», 8 (2005), n. 2. <<http://didattica.spbo.unibo.it/bibliotime/num-viii-2/zani.htm>>.
- [15] Marcus Zillmann. *Academic scholar search engines and sources: an Internet miniguide annotated link compilation*. <<http://virtualprivatelibrary.blogspot.com/Scholar.pdf>>.

Too much or too little? Web of science, Scopus, Google scholar: three databases compared

by Ezio Tarantino

The BIDS (Digital Library of “La Sapienza” project-service) was established in the “La Sapienza” University of Rome in 1999. Its first year of life saw the acquisition of 10 data bases and the stipulation of the first consortium contract with a large electronic periodicals’ publisher (Elsevier). The project has gradually grown to include today about 90 data bases and over 15,000 electronic periodicals. Among those first ten electronic data bases was the Web of science, the data base produced by the ISI, Institute for Scientific Information.

At the time the Web of science (WoS) was the only source that apart from providing a rich bibliographic data base, also functioned as a tool of citation indexing, very much appreciated by academic users.

In 2004 Elsevier proposed to try the new Scopus database, without any costs or commitments on our part. This was not because the digital Library needed a new multidisciplinary database, but because we needed a new updated citation indexing tool. The first comparative analysis that we did therefore was on the number of periodicals examined by Scopus, to see if, on the basis of this, it were not perhaps possible to eliminate some of the disciplinary data bases to which we were subscribing.

In the meantime, November 2004 saw the appearance of Google scholar, the new Google motor for academic research. It was not the first example of this kind (we must at least mention the sectorial experience of Citeseer and that of Scirus, the research engine specialized in scientific documentation, this too produced by Elsevier, and inserted into Scopus); but the popularity, authoritativeness and ambition of Google’s project immediately made it a truly fearful competitor. Thus, in the space of two months, WoS found itself face to face with not one, but two competitors, one of which moreover free of charge.

An investigation was therefore carried out, not to evaluate the superiority of one database over another (to do this one would have to carefully cross-examine all the comparative bibliography that is appearing with a certain regularity in recent months); but rather to help us see how these tools, even though powerful and with results that are overall more than satisfactory, cannot guarantee uniformity, consistency, real trustworthiness. The entrance in the market (or in the free territory of the net that is open to everyone, as in the case of Google) of new actors results in highlighting the gaps in a citation index, no matter what agency carries out the survey, so warning users of its inevitable lack of precision. And this process of relativization of automatic indexing mechanisms (so important especially in defining academic careers) should be considered, at least at awareness level, as a very positive result.

EZIO TARANTINO, BIDS, Biblioteca digitale dell’Università degli studi di Roma “La Sapienza”, e-mail ezio.tarantino@uniroma1.it.