

Reference Manager

Un programma per gestire su personal computer database bibliografici, per ricercare con protocollo Z39.50 e formattare citazioni in un dattiloscritto

di Francesco Dell'Orso

I *bibliography formatting software* (BFS) sono programmi per bibliografie personali che non si limitano a far gestire database semplificati, a permettere di inserire a volontà parole chiave nelle schede e a stampare elenchi di titoli col corsivo al posto giusto. Piuttosto si tratta di programmi specializzati per trattare dati bibliografici, dotati di caratteristiche che i programmi per database generici, DBMS (da Access a FileMaker ecc.) – più potenti, versatili e per l'appunto non specializzati – non hanno e che possono arrivare ad offrire dopo molto lavoro di esperti.

Cosa caratterizza la categoria dei BFS?

1. La struttura dell'archivio è da considerare piatta, verticale: archivio -> record -> campi (sottocampi) -> valore multiplo e indici. Quindi thesaurus, rinvii, archivio delle intestazioni controllate, interrelazioni fra archivi diversi – normali nei programmi per la gestione integrata di biblioteche e realizzabili con i DBMS – qui, mediamente, non hanno cittadinanza. Arrivano pronti per l'uso e volutamente facili da usare: gli archivi sono già allestiti con tipi di record dotati di campi per i vari tipi di documenti: libri, articoli, video, e-mail, brevetti, tesi, pagine web...

2. I campi sono a lunghezza variabile e talora possono ospitare valori multipli (nomi, parole chiave...); spesso tutti i campi, anche note e abstract, vengono indicizzati; le varie procedure – ricerca, sort, stampa – onorano i peculiari attributi già stabiliti per: autori, titoli di riviste, pagine, date, parole chiave...

3. Sono previsti filtri per convertire e importare dati bibliografici prelevati da altre banche dati su cd-rom, cataloghi in Internet, programmi per database; una funzione ad hoc permette poi di intercettare i duplicati che varie importazioni possono affastellare.

4. Una procedura fa interagire con l'archivio dei dati un testo preparato in videoscrittura, in modo da creare nel testo – in nota e in fine – citazioni bibliografiche provenienti dai record dell'archivio.

5. Sono previsti stili di citazione bibliografica, basati sulle specifiche di editori, associazioni, riviste: ogni stile è commisurato ai vari tipi di documenti; la fase di stampa prevede l'ordinamento alfabetico a più livelli subordinati.

6. Si sta affermando la tendenza ad offrire un modulo di ricerca su cataloghi remoti con interfaccia Z39.50¹ potendo convertire ed im-

portare automaticamente i record bibliografici reperiti.

Sono tutti aspetti che mediamente non appartengono a DBMS generici. Con quelli un utente ha linguaggio e strumenti di sviluppo per creare nuovi oggetti e applicazioni. Può andare ben oltre quanto offrono i BFS, in realtà flessibili solo all'interno di quanto progettato dal produttore. Ma sviluppare da sé con un DBMS quello che i BFS offrono già pronto richiede esperienza, tempo ed energie considerevoli, soprattutto per la manutenzione nel tempo e l'aggiornamento legato all'evoluzione di bisogni, software operativo, software applicativo, formato dei dati da importare...

I BFS vennero concepiti per risolvere il problema di dare forma diversa alle citazioni bibliografiche scegliendo un altro stile e senza dovere cambiare nulla nei record. Oggi sono dei database testuali completi orientati ai dati bibliografici: le funzioni di ricerca, di scrittura (input ed edit), di output sono tutte cruciali, robuste e articolate, così che infine il nome più corretto per essi sarebbe *personal bibliography managers*.² Sono programmi per bibliografie "personali", non perché una biblioteca non possa usarli, ma perché essi si rivolgono eminentemente all'utente individuale quanto a tipo di hardware e software operativo, perché privilegiano speciali esigenze di output come quelle per i dattiloscritti e, d'altra parte, perché – privi come sono di struttura relazionale e procedure gestionali – non consentono di informatizzare procedure quali la circolazione e il controllo periodici; inoltre hanno multimedialità molto ridotta e non possono competere con le dimensioni di archivi comuni in cataloghi di biblioteche.

I BFS rappresentano uno degli anelli della catena comunicativa che vede individui e istituzioni come produttori e fruitori di documentazione: il ricercatore, il biblio-

tecario, può recuperare quello che biblioteche e servizi producono e distribuiscono, a sua volta organizza dati e li pubblica, su carta e sul web: sono un ponte fra il grande e il piccolo percorribile nei due sensi. Chi maneggia testi, scrive, legge, cita, prende appunti, utilizza bibliografie e fa a meno di questi programmi, compie comunque molto lavoro manuale in più, che il computer potrebbe fare al suo posto: ordinare diversamente un elenco, decidere di invertire la posizione di "città: editore", passare da virgolette a corsivo, ricercare velocemente un nome appuntato, inserire nel proprio file centinaia di record trovati nel cd di una bibliografia nazionale o nel catalogo della propria biblioteca, sono operazioni che si fanno male solo con un word processor.

I produttori di BFS, e conseguentemente i loro prodotti, si trovano a lavorare in una corrente di input → output su cui non hanno nessun controllo né influenza preventiva: i produttori di database usano i propri criteri per creare, formattare e rendere scaricabili i record (lo standard ANSI/NISO su un formato comune per il download dei dati non è ancora comparso). Editori, riviste, associazioni scientifiche hanno le loro norme e pratiche per la formattazione degli output. Database e riviste scompaiono e di nuovi ne compaiono. Ci si aggiungano produttori come Microsoft e Corel per quanto riguarda i word processor. L'uno non si preoccupa dell'altro e ancora meno di chi sta in mezzo. In mezzo ci sono i database gestiti con i BFS, chi li produce e chi li usa. I produttori di BFS rincorrono da una parte i fornitori di dati preparando centinaia di filtri di importazione e dall'altra rincorrono gli editori preparando centinaia di stili di citazione. C'è da credere che sia una fatica improba, sia perché le quantità non fanno che crescere, sia perché non c'è alcun controllo preventivo

sulle necessità di aggiornamento.

Ci si può allora abbandonare tranquillamente al flusso che porta dati in input e offre formati di output senza richiedere il nostro intervento? Per esperienza diretta, l'accuratezza riscontrata è lontana dall'essere impeccabile e dunque né uno stile né un filtro preconfezionati dovrebbero venire assunti senza verifica dei dettagli interni assieme con un campione di dati.

Nuova versione, nuova ditta per Reference Manager

Reference Manager (RM) è senza dubbio uno dei programmi più importanti per capacità, diffusione, bibliografia³ della famiglia dei BFS. Il 16 aprile di quest'anno la ditta californiana RIS-Research information systems annunciava l'uscita di Reference Manager in versione 9: non era trascorso un anno e mezzo dall'uscita della versione 8 già per Windows 95. Lo stesso giorno ISI Institute for Scientific Information annunciava di avere acquisito la ditta produttrice del programma concorrente EndNote della Niles, formando la nuova società ISI ResearchSoft. Questa si è venuta a trovare in posizione di quasi monopolio del settore, producendo, sviluppando e vendendo i tre programmi più affermati: ProCite, Reference Manager e EndNote.

Le reazioni insoddisfatte o almeno preoccupate si fanno sentire più di quelle soddisfatte e rassicurate. Già nel 1996 – quando RIS inglobò la PBS di V. Rosenberg col suo ProCite – si erano levate varie voci di

Tabella 1 - Carta d'identità

Reference Manager versione 9 Windows; (include modulo per l'importazione prima distinto come Capture); c'è compatibilità – nei due sensi – con database dell'edizione 8
 Altre versioni: sola lettura: no; in altre lingue: no; dimostrativa: sì (modulo in Internet); rete: sì, licenza o per numero di installazioni (\$US 1.400) o di utenti simultaneamente attivi
 Assistenza: e-mail: tech@risinc.com
 Prezzo: ca \$US 400 + spedizioni (sconti per studenti e insegnanti)
 Produttore: ISI ResearchSoft 800 Jones Street, 94710 Berkeley, CA (USA) (tel. 01-760-438-5526, fax: 01-760-438-5526); e-mail: sales@risinc.com
 Lista di discussione: Reference Manager: per abbonarsi <mailto:LISTSERV@RISINC.COM> messaggio: *SUBSCRIBE RIS-list [Nome Cognome]+

preoccupazione. Si è pervenuti, si perverrà, alla omologazione dei prodotti perdendo in varietà e originalità? Potrà la società unica assicurare lo sviluppo differenziato di tre programmi o la prospettiva verosimile configura la progressiva assimilazione e dunque una perdita di competitività? Come potrebbe essere diversamente? ProCite e Reference Manager non sono diventati uguali, però non si può negare che si assomiglino molto più di prima: disegno degli stili di output e funzione CWYW-Cite While You Write per corredare un dattiloscritto delle citazioni bibliografiche sono sostanzialmente identiche. È prevedibile che, ogniqualvolta possibile, il meglio di quanto già presente o sviluppato ex-novo in un prodotto venga portato anche nella successiva release di un altro della squadra. Ma ciò è meno facile e rapido di quanto si possa credere: l'ISI ormai possiede da oltre tre anni i due prodotti e tuttavia permangono sostanziose differenze fra essi.⁴

Un primo contraccolpo negativo per gli utenti c'è stato: non ci sarà più una versione Macintosh di RM; il presidente D. Kochalko, in una e-mail del 5 Agosto 1999 a rislist@ris.risinc.com, lo ha detto chiaramente: le opzioni sono le edizioni per Macintosh di EndNote e ProCite, appaiate al- ➤

la loro equivalente versione Windows. Quali le novità maggiori della nuova edizione 9 di RM? Ricerca in rete su cataloghi remoti attrezzati con Z39.50, ottenuta incorporando il programma BookWhere? 7; potenziamento delle funzioni di importazione ed esportazione e dunque della possibilità di far circolare i dati; sul versante dell'output la creazione di stili è assistita, e la formattazione delle citazioni in un dattiloscritto è più articolata e sofisticata. La ricerca rintraccia parole interne a descrittori composti. Per le caratteristiche generali, requisiti di sistema, dati amministrativi e dimensioni massime, rimando alle Tabelle 1 e 2.

Database, catalogo, record e liste

Si può creare, distruggere e risanare dall'interno un database, ma per salvataggi e copie occorre agire in Windows. L'opzione "proprietà" fornisce statistiche sul totale di: record - database o sottoinsieme - autori, parole-chiave, periodici e date, e sul criterio di sort in vigore. Attenzione a confidare nel comando di "Undo", anche qui, non funziona proprio per le operazioni più distruttive come: distruzione di record e global edit.

Varie operazioni possono venire simultaneamente compiute su più database: ricerca (potendo anche usare la lista degli autori e parole chiave di fino a 10 archivi), stampa dei risultati di ricerca, intercettazione dei duplicati, formattazione di un manoscritto.

La versione in rete - che non ho potuto provare direttamente - ha licenza per *n* installazioni o per utenti contemporanei e permette di autorizzare o negare diritti di accesso multiplo in scrittura, non è una novità, ma rimane una funzionalità molto apprezzabile, e che infatti ha il suo prezzo.

Ogni database rimane strutturato in due file (approssimativamente dati e indici); i tipi di documenti trattabili sono ancora 35, e sono inclusi nel database; purtroppo non se ne creano di nuovi, ma si modificano quanto a nome, input obbligatorio e posizione (molto utile); gli attributi dei campi (autori, date ecc.) sono fissi. Fondamentalmente solo autori e parole chiave possono avere valori multipli. I campi sono al massimo 33, 5 sono definibili a piacere.

Un campo permette di registrare indirizzi di file (oggetti) che grazie al meccanismo OLE (*object linking and embedding*: qui solo collegamento e non incorporamento) verranno aperti con l'applicazione correlata: vale per i browser e gli URL (anche più di uno nel campo 33) e per immagini, suoni, testi, tabelle. È il massimo di multimedialità che i BFS offrono per ora.

Fanno parte di ciascun database anche i tre indici per autori, parole-chiave e titoli di riviste. Filtri di importazione, stili, lista di frasi e di termini per periodici e correttori ortografici sono singoli file esterni e utilizzabili da più database. Ogni record ha un identificatore che deve essere unico e che può essere assegnato manualmente oppure dal sistema con controllo su possibili duplicati: vale come chiave di ricerca, sort e come campo di output.

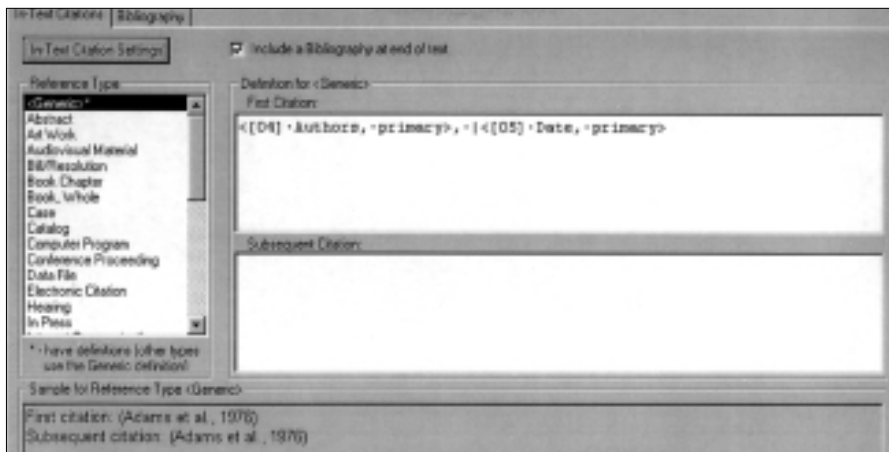
Le tre liste funzionano in fase di input e ricerca per riprendere voci in esse presenti: dal campo si apre

Tabella 2 - Requisiti di sistema e dimensioni

Processore: Pentium; Sistema operativo: Windows 95/98/NT4 o superiore; RAM: 16Mb; Hard-disk: ca 10Mb
Altro hardware: Mouse + Monitor (SVGA) 800x600 + cd-rom drive + stampante
Altro software: MS-Word per Windows 7, 97, 2000; Corel-WordPerfect 7, 8, 2000; un browser
Numero di database: illimitato
Numero di database simultaneamente aperti: 10
Numero di record per database: illimitato (100.000 raccomandato)
Numero di caratteri per record: illimitato
Numero di caratteri per campo: illimitato, i campi sono a lunghezza variabile (autori, parole chiave e titoli di riviste sono indicizzati fino a 255 caratteri e ammettono 255 voci per campo)
Numero di tipi di documento (schede): 35 (non se ne aggiungono)
Numero di campi: max 33, modificabili
Numero di stili di output: 460 + altri definibili, senza limite
Numero di filtri di importazione: centinaia + altri definibili
Lunghezza massima di una voce indicizzabile: 255 caratteri
Numero di liste utili a ricerca e catalogazione: 3 predefinite (autori, parole chiave, riviste) + 1 generica "Phrase list"
Dimensioni indice: indefinite
Numero di voci per lista: indefinito

- anche automaticamente - la finestra che fa scorrere l'elenco. Il contenuto dei campi corrispondenti non può non essere rispecchiato in esse, appena si salva un record la voce va nella lista. Ma esse sono dotate anche di autonomia: possono infatti ospitare voci non legate ai record, perché aggiunte, rimaste lì, importate. Inoltre possono avere "sinonimi", altri termini collegati come dei "vedi anche": 3 per i periodici (utilizzabili anche nell'output) e fino a 255 per nomi e parole chiave. La gestione dei sinonimi è automatica, generosa e rigida: se al termine X vengono attribuiti S e T come sinonimi, inevitabilmente nell'elenco compaiono anche S e T come termini, rispettivamente con T X, e S X come sinonimi, dopodiché ricercare per uno dei tre termini significherà utilizzare tutto il grappolo, ossia automaticamente anche gli altri due, che possono o no essere legati a dei record. I sinonimi di autori e parole chiave si

Fig. 1 - Catalogo in formato breve: nella parte superiore un record aperto in scrittura



possono fondere. Il “puoi anche vedere” è diventato un “vedrai anche comunque”. Non si va oltre a ciò, non c’è una gestione delle liste dei termini con rinvii esclusivi o con rapporti gerarchici. Una sola lista generica è riutilizzabile in catalogazione, in ogni campo, come serbatoio per stringhe ricorrenti che si vuole evitare di riscrivere: sfortunatamente non vi si possono importare file di testo.⁵ Due altri dizionari (liste) funzionano solo per l’importazione: per termini da non abbreviare e da non convertire in minuscolo. Per l’uso delle liste in input vedi oltre in Catalogazione.

Il database si presenta con la finestra standard della lista breve (record list) [Fig. 1 reclist], in alto il riquadro per il record sui cui si è posizionati e sotto quello per l’elenco dei record i campi sono incolonnati, nella posizione che si preferisce e teoricamente fino a 33 colonne quanti sono i campi. Il singolo record può venire presentato in due modi: formattato per l’output o come editabile direttamente. Per ordinare i record sulla base di un campo si clicca sulla barra della colonna. Si può anche provare a usare la lista per ritrovare un record scorrendo l’intero catalogo, ma qui non c’è ricerca rapida e dunque, con database appena

non piccoli, non è questa una credibile modalità di ricerca, quanto piuttosto di scorrimento di sottoinsiemi del database: risultato di ricerca, record importati o duplicati. L’aspetto della lista può essere diverso per il database e per i sottoinsiemi, ma purtroppo non per i diversi database: dunque, se in un database si usa un campo specifico e lo si fa comparire nella lista generale, lo si ritroverà, magari con una colonna vuota, anche nei database che non ne fanno uso. Questa lista è anche un formato standard di output che può venire stampato direttamente.

Catalogazione

Nonostante la crescente importanza del recupero di dati disponibili in formato elettronico, l’input da tastiera, caricamento e correzione manuale dei dati è operazione ancora importante, specie per chi stima che l’input sia l’operazione fondamentale e che le altre funzioni siano un’elaborazione e sfruttamento di questo lavoro. In RM la funzione è arricchita da vari comandi specifici: dai campi autori, da parole-chiave e titoli di riviste si accede alle rispettive liste. Qui è ottimo lo scorrimento alfabetico: scrivere “Terry”

porta a “terry” e non a “y” e si può retrocedere a “ter”; si selezionano anche più voci non contigue. Nei record le voci nuove vengono evidenziate nel record: forma discreta di validazione; il separatore di più voci “;” viene inserito automaticamente. Il campo delle parole-chiave di ogni record può venire automaticamente integrato con le parole presenti nei campi titoli, note e abstract dello stesso record e che risultano già presenti nella lista delle parole-chiave del database: è un’integrazione all’operazione di importazione. È prevista una funzione di correzione ortografica con molte opzioni: aggiungere termini al dizionario, crearne di nuovi ecc. I duplicati si intercettano sulla base di un confronto dei campi che l’utente può personalizzare scegliendo i termini del paragone. I record si possono duplicare in più di una maniera all’interno e verso l’esterno: trascinalo col mouse, deposito nella cartella appunti, copia e incolla speciale con scelta di quali campi duplicare (utile proprio per fondere i duplicati dopo averli intercettati). Di buon livello ordinario la batteria degli altri comandi di scrittura: attingere alla tabella di caratteri e simboli, apporre stili a stringhe, traslitterare in greco, effettuare aggiunte e correzioni globali. L’uso del mouse comprende funzione di trascinalo e tasto destro. Mancano altre facilitazioni come: macroistruzioni, conversione maiuscolo in minuscolo, caricamento di valori default.

Importazione ed esportazione

Vista la disponibilità e spesso gratuità di dati bibliografici in formato elettronico, è sempre più importante potere recuperare dati da altri archivi e trasferirli all’interno del proprio sistema. In generale non è un’operazione facile, lo sembra quando va a buon fine utilizzando filtri ➤

pronti, ma intendere da sé il formato dei dati in ingresso, farlo corrispondere a quello di arrivo, leggere – i record ancor prima dei campi – interpretare e tradurre bene non è semplice, reclama tempo, attenzione, energie, e non c'è molto di garantito a priori: bisogna vedere caso per caso. Punti delicati sono la possibilità di riconoscere più tipi di documento nello stesso file in arrivo, fondere, omettere, sostituire, aggiungere e soprattutto frammentare campi. In RM9 la procedura di importazione è stata potenziata con l'IFE-Import file editor che rimpiazza il precedente modulo separato *Capture*. I filtri già pronti sono oltre 450 e di continuo nuovi ne vengono offerti nel sito ISI ResearchSoft. RM converte direttamente database ProCite e EndNote.⁶ Il programma Book-Where? può scaricare dati direttamente in database RM senza conversione; vari OPAC di biblioteche inseriscono fra le opzioni di cessione di dati i formati per RM, ProCite e EndNote, database ISI come Web of Science, Discovery Agent, Reference Update e Current Contents consentono di portare dati in database RM con plug-in gratuito <<http://www.risinc.com/webcap/iscap.html>>. L'IFE è documentato solamente nell'help.

I filtri già allestiti non possono venire modificati stabilmente, vanno salvati con altro nome: se ne possono tuttavia creare di nuovi anche clonando quelli esistenti. Si dispone di funzioni di aggiunta (fino a 5 parole chiave e data odierna), sostituzione e frammentazione di stringhe. La frammentazione è un'operazione importante e le opzioni a disposizione sono davvero parecchie: quello che è insolito e non confortevole è che per un'importazione di routine, come quella di un file delimitato, non c'è nulla di pronto e occorre più che ingenuamente utilizzare proprio la più articolata delle opzioni di fram-

mentazione. I file in entrata possono contenere record di vario tipo.⁷ Quando si progetta di mettere *dentro* a un sistema dati non solo di prova, la cosa più importante è assicurarsi di come fare per portarli *fuori*: vista l'evoluzione del software, presto o tardi si potrebbe trovare un altro programma più soddisfacente, e comunque, comunicare con l'esterno è sempre più ricorrente: pubblicare su Internet, scambiare dati con i colleghi ... è imperativo non dovere ridigitare nulla, ma vedere l'esportazione come una forma di output e dunque come una delle molte elaborazioni che un software deve consentire sull'unico input effettuato. L'export in RM è un'operazione automatica che replica su file comunque tutti i campi, in formato delimitato (virgola o tabulatore) o a etichette (RIS o MEDLARS, ben documentate nel manuale), un diverso formato a etichette si può ottenere senza difficoltà creando uno stile di output e scrivendo il file su disco.

Output

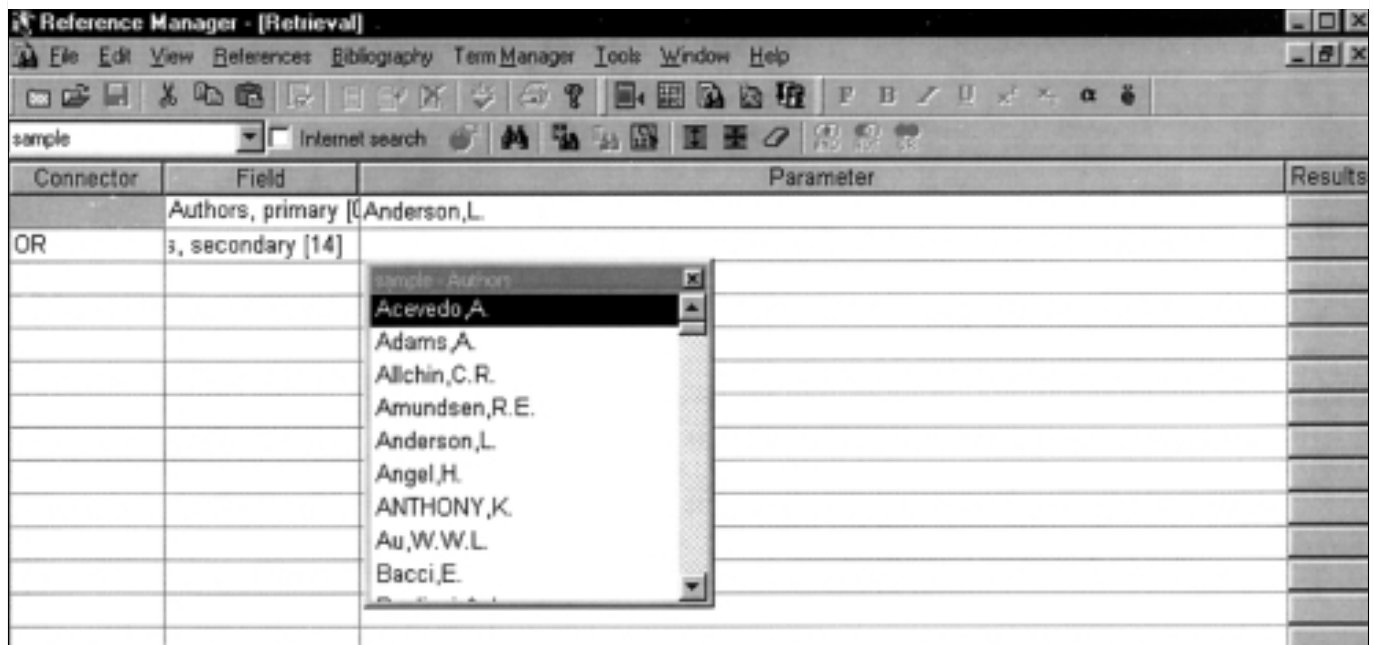
Per disegnare stili di citazione e dunque i formati di output dei record si dispone di un linguaggio essenziale, elementare e potente: praticamente identico e derivato da quanto offre ProCite. Solo tre istruzioni formali, fra queste utilissima è quella di gruppo, per cui un solo campo di una serie indicata basta a soddisfare la condizione posta a tutta la serie. A parte ciò non ci sono istruzioni condizionali se non quelle incapsulate dentro le moltissime opzioni previste per formattare nomi, titoli, date, pagine. Il disegno degli stili può essere guidato dalla tutela di un wizard e c'è verifica della correttezza sintattica. C'è una stampa diretta a partire dalla lista dei record, uno per riga. Manca invece completamente una funzione di stampa con intestazioni (tra-

dizionalmente e limitativamente detta *subject bibliography*), dove si usano dati dei record come chiavi di ordinamento e come esponenti premessi ai record che li condividono: è una mancanza consistente e significativa. Rivela infatti che le stampe vengono viste nell'orizzonte individuale di chi lavora soprattutto con dattiloscritti e con la loro bibliografia finale e non con cataloghi e indici, per una biblioteca o un lavoro bibliografico tradizionale questa mancanza sarebbe forse inaccettabile. L'output può venire salvato come file di formato testo, RTF per Word o WordPerfect, HTML. Sono previsti diversi criteri di ordinamento delle lettere: "a carattere", dove le lettere accentate seguono le altre: *liberti* precede *libertà*, "a parola", dove i diacritici non hanno peso, qui si possono usare le tabelle di Windows o quelle di Reference Manager (opzione consigliata) con piena portabilità del database. Le operazioni su più database non funzionano se questo criterio di sort non è il medesimo. Mancano stopword (dunque si ordinano male i titoli), e criteri compositi già pronti (ad esempio Autore/Titolo/Data o Autore-Titolo/Data); sono disponibili solo tre livelli di sort annidato.

Ricerca

È sempre più importante che sistemi che consentono di creare archivi di dimensioni ragguardevoli (fino a 100.000 record) offrano funzioni di ricerca articolate e potenti, per rispondere a richieste d'informazione e per preparare la via dell'output. Tradizionalmente i Text storage and retrieval systems hanno sempre sopravanzato i BFS per questo aspetto, ma oggi il numero di operatori è aumentato e in più si diffonde l'uso di indici da scorrere. In RM la ricerca propriamente detta si compie su uno schermo ad

Fig. 2 - Finestra della ricerca con accesso all'elenco dei nomi



hoc [Fig. 2 ricerca] che appare come un modulo di semplice struttura: una tabella con quindici righe per le espressioni e tre colonne: la prima, a sinistra, per collegare le linee con operatori booleani, poi una per dire "dove" ossia in quali campi cercare e infine la terza per dire "cosa", il contenuto da cercare. Bottoni a icona forniscono la scorciatoia per: comandi, operatori, campi e liste dei termini contenuti in autori, parole-chiave, titoli di rivista e data. Non vengono additati simboli, comandi in codice né sintassi da onorare. Alla fine, <F12> e il risultato viene affisso e mostrato. I termini usati in ricerca vengono evidenziati all'interno dei record, ammesso che lo si stia visualizzando formattato con uno stile e non come scheda editabile. Si possono salvare sia risultati in insiemi virtuali sia le espressioni di ricerca per rieseguirle senza riscriverle; si rivedono fino a 10 espressioni formulate nella sessione. È proprio tutto molto semplice. È indubbiamente meritevole cercare di semplificare la ricerca in modo che chiunque, senza addestra-

mento, senza dovere leggere manuali né schermi di aiuto, ma solo grazie a un'interfaccia elementare e intuitiva, sappia compiere la ricerca. Però non è automatico riuscirci. In realtà, operazioni rapide, eseguibili con pochi comandi, offerte qui come altrove, all'utente che può non essere affatto coinvolto nell'alimentazione e organizzazione della banca dati, sono molto più complesse di quanto sembrano, e difficili, insicure negli esiti quando non si ha piena consapevolezza di quello che accade. La ricerca solo in parte consiste nel porre domande e ottenere risposte. Ci si può accontentare delle risposte, positive o negative, che si ottengono; ma per avere un vero – anche se non definitivo – controllo sulla ricerca non basta conoscere la morfologia e la sintassi dei comandi, occorre soprattutto sapere *che cosa* nell'archivio è stato indicizzato e *come*, ossia cosa giace sotto, non si vede, e motiva il risultato, che si vede. Ciò vale per tutti gli archivi in cui si fa ricerca automatizzata: dai motori di ricerca in Internet ai cataloghi OPAC, dai da-

tabase su cd-rom a quelli gestiti con DBMS per personal computer. È un argomento classico, vecchio di 40 anni almeno. Ha a che fare prima di tutto con il fatto che non si cercano direttamente i dati, ma i loro indici e che questi perlopiù non si vedono o si vedono solo in parte. Ha a che fare con analisi e decisioni umane di chi configura gli indici e con gli algoritmi che vengono applicati in conseguenza di ciò. Per una persona alfabetizzata scorrere un elenco scritto nel proprio alfabeto è una routine quasi istintiva quando le voci sono organizzate secondo un criterio subito evidente, come è quello alfabetico. Tutti sappiamo aprire e consultare un elenco telefonico, pure sappiamo che non è così ovvio trattare preposizioni, articoli, segni di punteggiatura, prefissi, nomi composti con e senza trattino... La ricerca in un database è un affare meno semplice ancora. Che significa sapere cosa e come è stato reso ricercabile in una banca dati? Lo spiegherò con un esempio e otto domande con le relative risposte di RM. Sono solo otto domande ➤

– altre se ne potrebbero porre circa diacritici, punteggiatura, lunghezza minima/massima, prossimità, wildcard interne... – che mi posso (devo) porre anche quando uso motori di ricerca apparentemente essenziali e lineari, inclusi quelli per Internet. È da notare che, nella fattispecie, RM non dice molto, anzi non informa a sufficienza, né in termini di completezza né di precisione, né nell'help in linea né nel manuale, ma naturalmente fa varie cose, anche quando non documentate e non ovvie.

Prendiamo una stringa come "Inquinamento acustico urbano".

a) È ricercabile comunque, ossia tutti i campi dei record sono ricercabili? Sì, tutti i campi sono ricercabili, ma alcuni sono definiti "indexed" e altri "non-indexed", il che non significa affatto che alcuni siano "indicizzati" e che altri siano "non-indicizzati" perché tutti i campi sono indicizzati e ricercabili anche se in modo diverso⁸ (vedi oltre).

b) La stringa è trattata alla stessa maniera se è un descrittore a soggetto, o un titolo o parte di un abstract? No, RM non tratta tutti i campi allo stesso modo: fa differenza fra indicizzati e non-indicizzati: numero di record, date, parole chiave, autori, periodici hanno ricerca e sort più veloci, e in più cambiano le modalità di ricerca dei termini interni.

c) Per trovare i record che contengono la stringa, posso scriverla per intero (con o senza delimitatori)? Oppure devo cercare le singole parole collegandole con operatori: *inquinamento E acustico E urbano*? Posso scrivere di seguito e per intero; più parole adiacenti non vanno racchiuse fra delimitatori e non richiedono l'operatore E.

d) Posso cercare solo una parte della stringa, anche interna come: *acustico*? Sì, si possono cercare parti, anche interne.

e) Posso cercare una radice di parola come *urban* (quanto corta al minimo, anche *urb?*), e magari co-

me *stico*? Sì, si possono cercare radici in tutti i campi, troncate a destra e a sinistra.

f) Devo contrassegnare le radici con un simbolo, per esempio *urb**? Radici di parole, o più parole interne a campi non-indicizzati non richiedono simbolo, mentre la stessa ricerca nei campi indicizzati reclama il simbolo "":⁹ "**acustico urbano*" "*inquinamento**" o "*inquinamento acustico*" sono tutte forme ammesse in ragione dei campi a cui si rivolgono.

g) Quali sono sintassi e ordine di esecuzione in un'espressione composta con più elementi e operatori logici: *acustico O atmosfer** E *inquinamento*? Non ci sono segni – come le parentesi – per espressioni composite e non c'è priorità fra operatori booleani: l'ordine di esecuzione è da sinistra a destra e dall'alto in basso, dunque: "*acustico O atmosferico E inquinamento*" non equivale a: "*inquinamento E acustico O atmosferico*".

h) Si può cercare la stringa simultaneamente su tutto il record come un unico testo intero oppure occorre sempre indicare almeno un campo? Non c'è ricerca su tutto il record, occorre specificare un campo o un gruppo di campi.

C'è dell'altro: l'operatore di fascia "da-a" è attivo solo per le date e gli identificatori dei record. C'è un solo operatore comparativo per un tipo speciale di data. Per applicare operatori assenti come un NOT unario e per ricercare i campi vuoti occorre andare oltre il manuale e essere creativi, e tuttavia il successo non è sicuro.¹⁰ Non ci sono opzioni di ricerca approssimativa né di classificazione dei risultati in base al peso dei termini. C'è invece una ricerca rapida e facile, ma è nascosta dentro la funzione di *Term Manager*; se si entra lì si può scorrere un indice, allargarlo a tutti i database, selezionare dei termini e lanciare una ricerca veloce scegliendo al volo l'operatore booleano con cui collegare i termini.

Ricerca tramite il protocollo Z39.50

La stessa interfaccia di ricerca, con gli stessi comandi, viene usata da RM per le ricerche su cataloghi per cui è stato implementato il protocollo Z39.50 e su PubMed <<http://www.ncbi.nlm.nih.gov>>: RM infatti incorpora BookWhere? della canadese Sea Change Corp.¹¹ È un BookWhere? con alcune opzioni in meno,¹² ma che mantiene il grosso delle opzioni della versione originale: consente ri-configurazione, creazione, test di host e database (ne offre in partenza rispettivamente ca 350 e 600), ricerca simultanea su fino a 255 database diversi, vedendo subito come record RM il risultato, poi importabile definitivamente nel proprio archivio selettivamente o in massa.

Formattazione del dattiloscritto

La formattazione del dattiloscritto è tuttora una specificità dei BFS che li distingue dagli altri DBMS e li caratterizza come strumenti personali. Qui l'omologazione fra RM e ProCite è quasi totale. Anche in RM la procedura si chiama *Cite While You Write* e funziona come tale: il menu Strumenti del word processor (MS-Word o Corel-WordPerfect) offre le opzioni per il programma incuneatosi dall'esterno. Restando dentro il testo dattiloscritto si inseriscono i rimandi ai record, scrivendoli o andando a cercare i record nel database, proprio con una funzione di ricerca che può utilizzare praticamente ogni parte dei record, anche inseriti in database diversi. I segnalini si possono mettere nel testo o in nota a piè di pagina, si possono nascondere del tutto o in parte, pre/suffissare con altre parole. Si possono creare richiami multipli, anche con un solo comando. La li-

sta finale avrà criterio di ordinamento e stile variabili. Ora RM9 sopravanza ProCite perché aggiunge dei dettagli, sia in termini di comandi operativi (vedere il record intero e non solo una riga, cercare nel testo legami non annodati al database, rivedere fino a 10 ricerche fatte e riprendere automaticamente l'ultima) che di prodotti: la citazione nel corpo del testo può essere diversa per articoli, libri, video, la prima diversa dalle successive e da quella della lista finale, che può anche non venire prodotta [Fig. 3 Output]. Quest'ultime sono migliorie richieste per molto tempo soprattutto da chi usa stili diffusi nell'ambito umanistico (Chicago, MLA, Turabian).

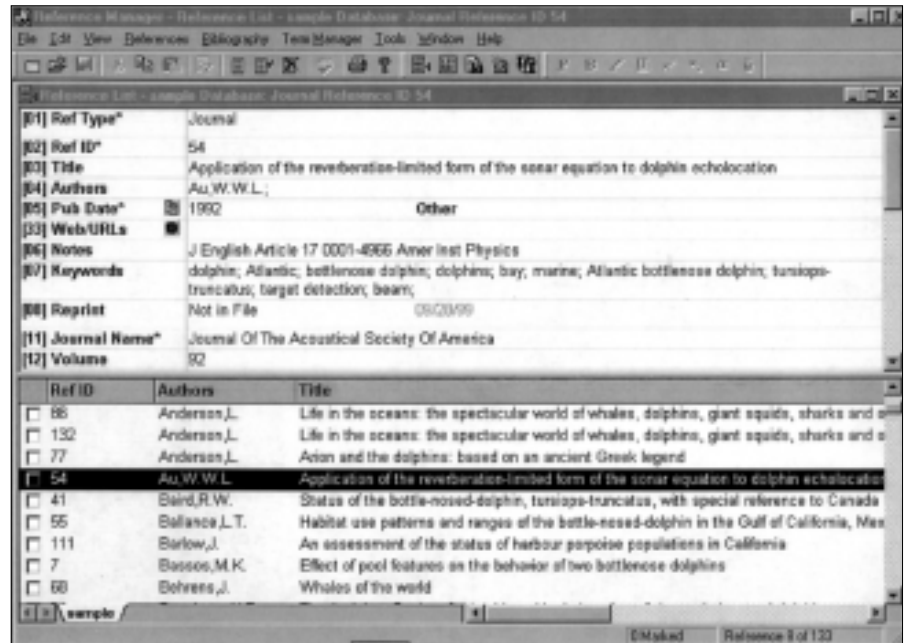
Internet

Come interagisce RM con il mondo Internet? Ha il client Z39.50 incorporato. Formatta i record anche in HTML (ma non si preparano link attivi). Grazie al prodotto di squadra Reference Web Poster permette di porre un database (RM o ProCite o EndNote) in rete, dinamicamente interrogabile. URL registrati nel campo 33 dei record possono venire cliccati attivando automaticamente il browser (quando lì c'è il numero PM di PubMed si va immediatamente alla scheda dell'articolo in questione). Prevede schede e stili per due tipi di documenti elettronici: Electronic citation e Internet communication.

Conclusioni

La documentazione offerta dal manuale e dall'help contestuale denota che si è privilegiata la semplificazione. Il manuale è dignitoso, comprensibile, con utili appendici, ma non eccellente, talora incompleto e impreciso. Manca ogni documentazione sulla modifica e

Fig. 3 - Specifiche di stile per le citazioni da inserire in un dattiloscritto



creazione dei filtri di importazione, è scarsa e fuorviante quella sulla ricerca, quanto ai sinonimi anche l'help è incompleto.

In sintesi quali i tratti più pregevoli di RM, anche in quest'ultima versione 9? Il client di ricerca Z39.50 BookWhere? integrato. Varie operazioni – ricerca inclusa – compiute simultaneamente su fino a 10 database diversi. Vera multiutenza nella versione in rete. Liste di termini buone per alimentazione, sinonimi, consultazione, validazione dei nuovi termini, uso in input e in ricerca anche fondendo indici di più archivi e cercando in tutti. Disegno degli stili di output semplice e potente. Formattazione del dattiloscritto sempre più completa. Correttore ortografico. Importazione con buone possibilità di riformattazione. La ricerca è veloce. Poi, di minore rilievo: evidenziazione dei termini usati per la ricerca e reperiti, record editabile nello stesso schermo in cui si scorre il catalogo, evidenziazione dei termini ricercati, facile modifica delle schede di immissione dati. Si può

scommettere che la prossima versione di ProCite incorporerà parte dei punti forti di RM che ancora gli mancano.

Per contro, quali i lati deboli e, dove possibile, i suggerimenti per miglioramenti? Non ha, e non avrà più, un'edizione Macintosh. Il prezzo è alto, ma ora che praticamente include la ricerca Z39.50 si giustifica di più. Se in input i campi – come accade per i nomi – prevedessero sottocampi, se ne avvantaggerebbero ricerca, sort e stampa. La gestione dei sinonimi è un po' piatta e rigida: legami gerarchici e veri rinvii la migliorerebbero. Si dovrebbero potere aggiungere nuovi tipi di scheda. Farebbero comodo altre liste di termini con possibilità di importare file di testo esterni. Lo schermo iniziale di scorrimento del catalogo è personalizzabile e dunque andrebbe legato ad ogni database, potrebbe recare più indicazioni, dovrebbe incorporare la ricerca rapida. La funzione di ricerca andrebbe razionalizzata, spiegata meglio e corredata con più operatori. Quanto all'output: ➤

aggiungere stampa con esponenti e indici; potenziare il sort con più livelli di chiavi e stopword. La documentazione è da migliorare soprattutto per ricerca e importazione. ■

Non intrattengo alcun rapporto promozionale o commerciale con la ISI Researchsoft. I nomi dei prodotti software citati sono marchi dei rispettivi produttori soggetti a copyright. Manoscritto terminato il 3 settembre 1999. Ho provato la versione 9 monoutente di Reference Manager con Windows 95 su un Pentium II 233 Mhz con 64 Mb RAM.

Note

¹Z39.50 è un protocollo che permette a un database (server) di rispondere a interrogazioni provenienti da stazioni di lavoro (client) equipaggiate con software diverso da quello del server quando entrambe hanno implementato lo stesso protocollo Z39.50. Esso è dunque il linguaggio di interfaccia che entrambi, client e server, parlano. Informazioni presso la Library of Congress a: <<http://lcweb.loc.gov/z3950/agency/>>.

²Perché continuare a usare la denominazione BFS allora? Per l'originaria, "storica", destinazione di questi programmi: la *formattazione* delle citazioni, la denominazione conseguentemente loro assegnata all'inizio è divenuta tradizionale ed è assestata, e comunque in essa rimane il riferimento ad una delle caratteristiche e funzioni tuttora di primo piano.

³Fra i più recenti contributi apparsi in Internet nel 1999: JUDY MATTHEWS (cur.), *Reference Managers*, <<http://www.nature.com/software/refman/intro.html>>: è una rassegna di otto pacchetti fra cui Reference Manager 9.0 (beta), è apparsa nella sezione "Software Reviews" di "Nature", v. 400, (29 July 1999) n. 6743; MAGGIE SHAPLAND, *Evaluation of Reference Management Software: Comparing Papyrus with ProCite, Reference Manager, Endnote, Citation, GetARef, Biblioscope, Library Master, Bibliographica, Scribe, Refs,*



M.C. ESCHER

<<http://www.cse.bris.ac.uk/~ccmjs/rm/eval99.htm>>: fa seguito all'analogo lavoro del '98, contiene analisi con tabelle e commenti aggiunta di programmi e aggiornamento; a cura delle Biomedical Libraries of Dartmouth College, *Downloading Bibliographic References*, <<http://www.dartmouth.edu/~biomed/download.html>>: è una breve presentazione di alcuni programmi fra cui Reference Manager con speciale attenzione allo scarico dei dati da una serie di fonti e alla loro conversione.

⁴Per un'analisi comparativa dettagliata:

F. DELL'ORSO, *Bibliography Formatting Software: An Evaluation Template: Head-to-head comparison between ProCite, EndNote and Reference Manager (ISI Researchsoft) via an evaluation grid*, 2nd ed., August 18 1999, presso: <<http://www.burioni.it/forum/ors-bfs2/ors-bfs.htm>>. Nei test RM si è mostrato molto meno veloce di ProCite per importare e distruggere record e per il sort del database ed anche meno stabile.

⁵per accedervi occorre usare il tasto destro del mouse: il manuale qui (p. 381) è sbagliato.

⁶Corrispondenze analitiche fra tipi di record e campi sono fornite in due utili file di testo. Però l'esperienza di lettura-conversione diretta di un database ProCite in RM, fra l'altro lentissima, è stata una delusione quanto a precisione.

⁷Due dizionari corredano l'importazione: uno per i termini - come DNA, US - che devono comunque conservare le maiuscole anche se si imposta la conversione generale a minuscolo e un altro per quei termini che non vanno scambiati per abbreviazioni anche se inseriti in titoli abbreviati, ma in cui non è usato il punto per le abbreviazioni.

⁸La distinzione riguardava precedenti edizioni di RM e la terminologia è stata conservata per comodità di riferimento.

⁹A meno che non si tratti di tutto un cognome, perché grazie alla virgola RM sa riconoscerlo! e vale anche per il curatore di una collana, inserito erroneamente fra i non-indicizzati e dunque da suffissare con l'asterisco di troncamento.

¹⁰Ad esempio per trovare i campi vuoti: si deve cercare un range di record (RefID 1-500) poi un NOT e "*" per "any character", ma ciò vale solo per gli *indexed fields* perché i non-indexed non vogliono l'asterisco e lo interpretano come un carattere da cercare *.

¹¹Per una descrizione di BookWhere? cfr. F. DELL'ORSO, *BookWhere?: la ricerca in cataloghi OPAC di sistemi bibliotecari presenti in Internet con un client Z39.50*, "Biblioteche oggi", 15 (1997), 3, p. 26-33.

¹²Ad esempio display in formato MARC e espansione della ricerca grazie all'albero dei soggetti.