

Micro Cds/Isis[®]: tecnologia, funzioni e procedure

Presentazione ed analisi di un information retrieval system per personal computer (Parte Prima)

di Francesco Dell'Orso

Che in molti si continui a confondere il software con gli archivi costruiti mediante esso può spiegare questa banalità in esergo: Isis¹ è un programma software con cui si disegnano uno o più archivi. I database possono ospitare ogni genere di dati alfanumerici, ma tratti peculiari della struttura dei campi stimolano l'appetito di bibliotecari e documentalisti in particolare. Dal punto di vista strutturale Isis è annoverabile fra i programmi *file managers*, i dati memorizzati stanno in un database considerabile come un file, in pratica unico, vista la gregarietà e relativa invisibilità degli altri. Dal punto di vista funzionale fa parte dei sistemi di *Isr-information storage and retrieval* con netta vocazione per il *retrieval*, includente il recupero e l'output dei dati, a discapito dello *storage*. È anche stato classificato nella sottospecie dei *classical²* che presenta spiccate somiglianze con le caratteristiche peculiari dei sistemi di *information retrieval* da anni usati per interrogare le banche dati on-line. Queste caratteristiche, effettivamente condivise da Isis includono: struttura degli ar-

chivi interamente definibile dall'utente (*table driven* o 'parametrizzata'); campi variabili per numero e lunghezza; input generalmente arcaico, ma con ampia facoltà di disegnare le varie schede per l'immissione dei dati; ricerca incardinata su un indice detto *inverted file*, in cui viene trasposto il contenuto dei record, con libertà di indicizzare tutti o parte dei campi, in più di un modo ed anche manipolandoli; interrogazione mediante istruzioni composite, con uso di operatori booleani e relazionali, senza perlustrazione degli indici o navigazione. Si enunciano per questi software copie di caratteristiche inversamente proporzionali: quasi nulla di definito, e allora quasi tutto definibile; grande flessibilità e relativa potenza, e allora poco *maternage* dell'utente che si vuole prima volenteroso e poi esperto: il programma dà molto, e allora chiede molto; formalizzazione dei linguaggi, e allora scarsa colloquialità; mezzi efficaci per strutturare e manipolare analiticamente, e allora più laboriose documentazione, manutenzione, trasformazione. Insomma: nessuna *user friendliness* vie-

ne sventolata sulla soglia, ma chi entra e sa adattarsi trova l'amico e il tesoro, *si parva licet*.

Altre categorie di software servono a trattare: lunghi testi, biblioteche in modo integrato, archivi multimediali, note ai testi e bibliografie in vari stili; Isis non rientra in nessuna di esse, ma svolge in parte ognuna di tali funzioni.

Definizione degli archivi e organizzazione dei dati

Isis arriva nelle mani dell'utente quasi nudo, non c'è nessuna struttura di archivio neutra, ancorché semplice o minima, già preparata, da utilizzare come guscio o schema per sottrarre, aggiungere, mutare, né basta battezzarla perché prenda forma. Di strutture già pronte Isis offre solo vari menu che guidano, per scelte successive, allo svolgimento delle procedure. Per questo è *menu* e non *command driven*, ma una volta indicate le opzioni sui menu, il programma torna sostanzialmente a venire guidato con comandi: bisogna sapere cosa si vuole e cosa fare per ottenerlo. Non si usa mouse, si possono memorizzare sequenze fisse di comandi in tastifunzione, <F12> o <Shift-F6> ecc. Con <F6> si torna in Dos senza abbandonare Isis, ma è un ambiente limitato (*prompt* oscurato, perdita del *doskey*).

Al momento di disegnare l'archivio — ovviamente si possono usare archivi già definiti da altri — si è davanti alla tabula rasa, al gran foglio bianco, ad un piano deserto.

Per disegnare un archivio Isis reclama, in totale silenzio, che l'operatore compia quattro passi, imprescindibili, in sequenza fissa, senza avviso e senza guida: fattone uno si è sospinti inopinatamente nella stanza di quello seguente. Non c'è nulla di in- ➤

trinsecamente difficile e definitivo nel fare questi passi, si può perfino fingere di farli, ma questo lo intende chi già lo sa.

I quattro passi da compiere consistono nella definizione di quattro oggetti, fisicamente distinti come file con nome Dos controllato dal programma per l'estensione e per il resto identico al nome dato al nascituro archivio. Si tratta di definire:

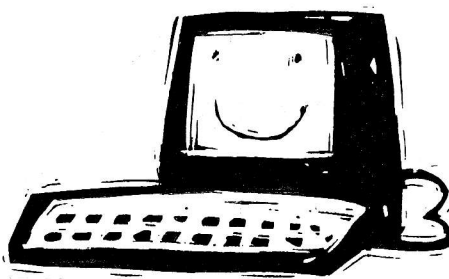
1) il tracciato del record — unico per ogni database — comprensivo dei campi di tutti gli eventuali diversi tipi di entità che si volessero registrare nel medesimo archivio (es.: prodotti e clienti, libri e utenti, musei / collezioni / oggetti / cataloghi);

2) un modulo — maschera — coi campi in cui inserire i dati;

3) la tabellina che elenca quali dati, contenuti, campi sono da indicizzare nell'*inverted file* ai fini di ricerca, con cruciali istruzioni su come farlo: anch'essa è unica per quanto attiene all'indice;

4) un formato di output per visualizzare, presentare i dati su schermo, carta, disco.

Compiuti questi quattro passi il programma annuncia che è nato un database e crea automaticamente altri dieci file, diciamo di servizio al funzionamento generale, non soggetti a manipolazione da parte dell'operatore. Come creare altri database? Chi non è né malizioso né masochista potrebbe stimare che, assolta una volta la corvée poi, per altri archivi, si possa procedere copiando la prima matrice e lavorando sulla copia. Lo si può fare se non si è affatto ingenui e se ne sa molto di più, anche di ciò che è esposto dal Reference Manual.³ Analogamente non è rapido né lineare cambiare nome ad un archivio. La cosa operativamente più facile da fare condividere agli archivi è poi la più preziosa, ossia i dati — anche mutandone il formato con so-



fisticate conversioni — lì l'operazione è prevista e spiegata, si arriva facilmente a fondere archivi, per il resto, molto più banale, bisogna ingegnarsi.⁴

Con questi quattro oggetti di base, Isis è pronto per consentirci di immettere dati, ricercarli, stamparli, esportarli. A questi quattro se ne possono aggiungere altri, strutturalmente simili e naturalmente diversi per scopi specifici: qui si mostra la generosità del programma, potenza e flessibilità arrivano a riscattare il burbero approccio iniziale. I limiti sono perlopiù numerico-teorici,⁵ non ci sono vincoli di campi obbligatori, tutto si può modificare a posteriori.

Non v'è limite numerico agli archivi disegnabili, essi comunque dimorano irrelati fra loro, se ne chiude uno per aprirne un altro, quanto appartiene ad uno non appartiene ad un altro. Ma all'interno di uno si possono collegare più record fra loro. L'operazione di giuntura funziona solo nell'output e nell'esportazione: grazie ad essa nel mostrare il contenuto di un record se ne può richiamare e visualizzare un altro, o più, in tutto o in parte. Questo embrione di relazionalità è retto da un legame che può venire stabilito o in base al numero che ogni record ha all'interno di un database (*mfn-master file number*), o grazie all'indice usato per la ricerca che funzionerà da ponte. Il legame viene comunemente usato per registrare forme brevi nelle descrizioni e le relative forme estese e controllate in altri record una sola volta; op-

pure è estendibile gerarchicamente: per cui da un record A si richiama il contenuto di un record D collegandosi a B che è connesso a C che è legato a D. Ma la varietà d'uso è tutta nelle mani di chi applica la funzione. La relazionalità a rigore non tocca neanche la ricerca né l'input in quanto tale: non c'è ombra di *authority file* nemmeno come liste esterne di riferimento, i record sono scatoline chiuse. Il collegamento dei dati è dunque parziale, ma serve comunque, e non poco.⁶

Memorizzazione e catalogazione

I record sono registrati nel database nell'ordine in cui vengono immessi, si può poi ordinare il database secondo ogni criterio, ma è un'operazione che occorre inventarsi, non propriamente lineare. Tutti i record sono necessariamente numerati, in un'unica serie sequenziale con progressione 1, senza duplicati. Il numero consente di richiamare i record in tutte le funzioni: ricerca, catalogazione, stampa, salvataggi e esportazioni. Che la ricerca preceda l'inserimento e modifica dei record, ossia la catalogazione, non è organicamente previsto — come invece tendono a offrire i sistemi di gestione catalografica per biblioteche — ma l'utente può farlo e passare da ricerca a scrittura dei dati con tre colpi, in alternativa si richiama un record tramite il suo numero.

I dati vengono registrati in campi identificati da un numero, detto "Tag" che è il solo nome tramite cui il programma li riconosce. L'immissione dei dati è una piana operazione di scrittura dei dati in una "scheda magnetica", ma la possibilità di ripartire il contenuto di un campo, al fine di manipolarlo successivamente, è articolata e

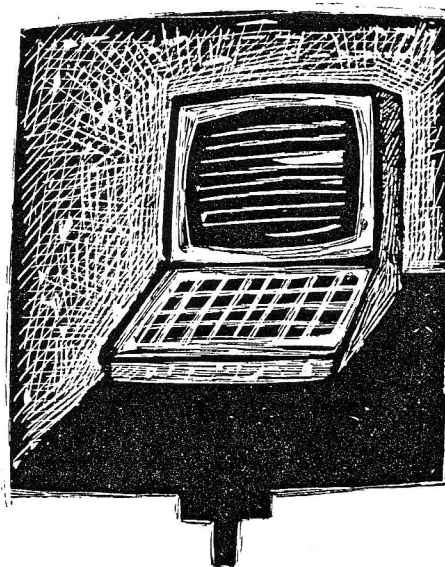
va incontro ai gusti dei catalogatori. Come nei formati Marc un campo può venire ripetuto — avere cioè più occorrenze — e frazionato in sottocampi, ripetibili però solo in distinte occorrenze. Potrò dunque registrare: `^1Finzi^2Claudio%^1Cimmino^2Luigi^cfilosofo` ecc., ma non `^1Correa^2Ismael^2Enrique`. Tutti i campi sono a lunghezza variabile, un database può ospitarne fino a 200; non c'è limite al numero di occorrenze né di sottocampi utilizzabili né dei campi a cui ciò sia applicabile. Campi e sottocampi sono distintamente intercettabili in: ricerca, stampa e ordinamento, esportazione/importazione e dunque possono venire individualmente richiamati, omessi, enfattizzati, dislocati, valorizzati. Tutte le occorrenze vengono interessate dai comandi, nell'ordine in cui sono registrate, come distinte unità. Per immettere dati si usano maschere disegnate interamente dall'utente, illimitate per numero, ripartibili su 20 schermi (22x80), ciascuno dei quali può contenere 19 campi, che non fluttuano tra gli schermi.

In catalogazione non si selezionano record rispetto al resto del database. Non si duplicano record né si intercettano i duplicati.

Un record è logicamente distrutto con un solo comando, fisicamente è tale solo quando l'archivio viene ricompattato.⁷

Lo strumento di scrittura cambia a seconda dell'operazione che si sta svolgendo. Isis prevede un *editor* di riga, di campo, di maschera e di schermo intero. Quello ovunque ricorrente è un *editor* di campo che fra l'altro usa i tasti standard della tastiera per il movimento del cursore, l'<Esc> come comando di <Undo> dentro al campo, e può fare ricorso a una zona di memoria temporanea (*buffer*) in cui trasferire un blocco di caratteri alla volta. Il *buffer* rimane attivo — ma non visibile — quan-

do ci si sposta fra campi dello stesso o di diversi record, e fra funzioni diverse, dimodoché si possono copiare, distruggere, spostare blocchi dentro ad un record, fra record, e trasformare una stringa selezionata in catalogazione in un termine di ricerca. Non c'è comando di scrittura per conversioni maiuscolo/minuscolo. <F1> evoca — come di consuetudine — messaggi di aiuto legati al campo in cui si trova (ma niente indica che lo si possa fare). La validazione del contenuto dei campi ricorre



solo per quelli che sono stati definiti nel tracciato record come a schema prestabilito.

Le convenzioni specifiche per la memorizzazione dei dati sono tre: separare le occorrenze di uno stesso campo (con il segno %, Ascii 37 oppure con altro, comunque unico per tutti i campi), segnare l'inizio di un sottocampo (col circonflesso ^, Ascii 94, seguito dal carattere voluto: numero o lettera, in maiuscolo o minuscolo), marcare quanto si vuole rendere ricercabile come tale, oppure si vuole ordinare alfabeticamente in modo diverso da come scritto e mostrato, (con <...>, Ascii 60 e 62, ad es.: <S.=santa> e <784.3=Lie-

der> faranno valere per l'ordinamento *santa* in luogo di S., e *Lieder* invece di 784.3.

Si possono fissare valori di *default* a due livelli: 1) generale e permanente, modificabile insieme agli altri parametri strutturali dell'archivio; 2) per l'operazione in corso, potendoli modificare *n* volte nel corso della medesima sessione e con prevalenza sul livello 1.

La registrazione di un record nel file dei dati è automatica ed immediata, quella nell'indice va deliberata, è in ogni caso effettivamente differita e relativamente fra le operazioni più lente.

Di norma l'aspetto tipografico dei caratteri è controllato nella fase di output, non è quindi in catalogazione che ci si fa carico di tabulazioni, rientri, grassetto corsivo etc. Mancano comandi di scrittura trasversale, cioè di correzione, aggiunta o distruzione cosiddette globali, tali da riguardare simultaneamente più record. Non c'è alcuna interazione organica con i word processor, non v'è ombra di procedure come quella offerta dai Bfs e nota come "manoscritto" per cui i richiami contenuti in un dattiloscritto vengono incrociati coi dati contenuti in archivio e da questo vengono estratti i record, ad es. citazioni bibliografiche, che risultano dal felice incontro.

Si possono inserire dati in un archivio Isis senza immetterli con la funzione di catalogazione ad esso propria. Se si dispone di un file di formato quasi identico allo Iso 2709⁸ lo si importa fisiologicamente in Isis con una funzione già prevista. Durante l'operazione si possono manipolare profondamente i record con una tabellina di filtro. Questo scambio di dati funziona come prima via fra utenti Isis. È anche comune disporre di file con formato dei campi ad etichette (*tagged*) o *comma delimited* (tipo AU: Landolfi, Tommaso, e, sotto: TI: La pietra ►

lunare, oppure, di seguito: "Landolfi, Tommaso", "La pietra lunare"). Molti programmi di tipo Irs, Dbms, Bfs, così come i software che gestiscono banche dati su cd-rom li producono come formato di esportazione. In questo caso si possono usare vari programmi di conversione ad Iso 2709, fra essi spicca Fangorn.⁹ L'operazione si svolge fuori da Isis, il file risultante viene poi importato, potendosi di nuovo avvalere della tabellina di filtro per un'ulteriore manipolazione.

Con Fangorn si trattano, distintamente in ogni singolo campo, le occorrenze plurime, i sottocampi (parzialmente), si cancella o rimpiazza testo, si accolgono, fondono, omettono, deviano campi senza ordine prestabilito, senza campi o contrassegni obbligatori. I campi si devono trovare su righe diverse, salvo che per il formato Cdf. Mancano: una funzione per estrapolare parti del contenuto di un campo (*parsing*); istruzioni condizionali; possibilità di aggiungere testo o campi, il che è invece, in diversa misura, possibile durante la successiva importazione in Isis, grazie al linguaggio di formattazione (cfr. oltre).

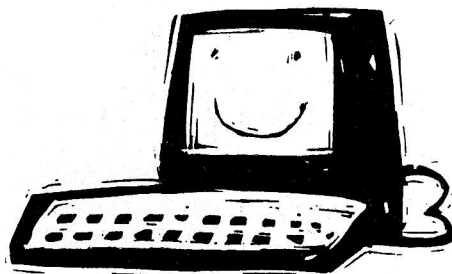
Altri programmi consentono di pervenire a file Iso 2709: Txt2Iso dell'istituto Ciat (Colombia) tratta file di testo ad etichette; una utility converte file .Dbf del *dbIII*; *Data Magician* — questo non gratuito, della Folland Software — può trattare come input file Ascii in formato ad etichette, Marc, *dbase*, *Inmagic* e *Star*. Piccoli programmi — come *Dlmimp* e *Comma* — vengono usati per scrivere direttamente file Cdf nel master file di Isis, senza passare per l'importazione.

Ha dunque senso scrivere i dati da immettere in un database Isis su un file del tutto esterno, con un wordprocessor avvalendosi delle sue capacità — macroistruzioni,

correzioni globali, liste di voci di autorità cui appoggiarsi (come in *Pro-Cite*), attività sotto *Windows*... — e poi trasferire questi dati in Isis con Fangorn, o altro.

Il linguaggio di formattazione

Isis dispone di un gruppo di istruzioni formali che compongono un linguaggio detto di formattazione (*formatting language*) di cui è data anche presentazione formale secondo lo schema Backus-Naur. Non è un linguaggio di programmazione o di elaborazione di stringhe alfanumeriche, ma un mezzo per selezionare il contenuto dei campi, e manipolarlo insieme a stringhe di caratteri e a co-



mandi di posizionamento.

Questo linguaggio è eminentemente connesso con le operazioni di output, ma è un'errata riduzione ritenere di doverlo usare solo per disegnare i formati di presentazione dei dati, ovvero per la stampa, su carta, disco, schermo che sia. Infatti il suo uso ricorre trasversalmente in altri tre ambiti funzionali di considerevole importanza: in ricerca — sia nella preparazione dell'indice analitico che nelle espressioni di ricerca sequenziale —; nell'estrazione di voci dai record e nel loro ordinamento alfabetico (*sort*); nella riformattazione dei dati durante l'import/export dei record. In tutti questi tre ambiti i comandi di formattazione espletano la propria

funzione incrociandosi con un altro parametro cardinale, che è detto "tecnica di indicizzazione" e consente di estrarre parole o più parole o porzioni di esse.

All'osso, le istruzioni del linguaggio servono ad indicare un campo tramite il suo nome (il "Tag" numerico), a selezionarne così il contenuto per poi svolgere varie azioni, la fondamentale delle quali è mostrare quel contenuto selezionato.

Sempre volgarizzando: "Prendi il campo 39 che qui contiene il nome dell'autore e mostralo sullo schermo"; oppure "Estraine solo il primo sottocampo e riponilo nell'*inverted file* affinché sia ricercabile"; oppure "Trattalo come entrata di un elenco alfabetico di tutti i nomi contenuti nell'archivio"; oppure ancora "Estrailo dai record ed esportalo in un file i cui record conterranno solo quello".

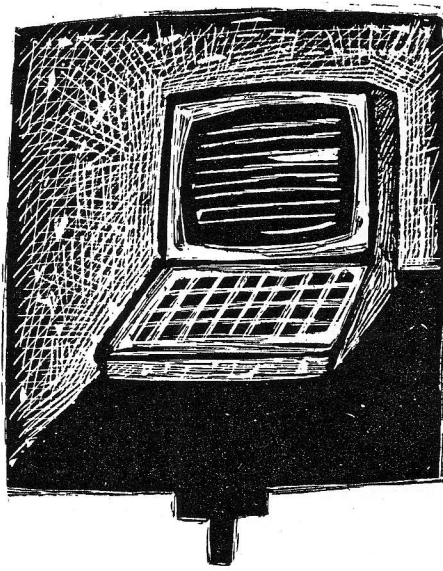
Dalla possibilità basilare di selezionare un campo discendono le possibilità di selezionare: le sue occorrenze (non selettivamente, ma tutte come un gruppo organico di sottounità), i suoi vari sottocampi partitamente, porzioni dei dati in ragione delle loro coordinate di ubicazione.

I comandi consentono di determinare le posizioni in output (verticale/riga, orizzontale/colonna, capoversi, rientri, nuova pagina), pre — o suffissare i dati con stringhe di caratteri (punteggiatura, etichette, *literals*). Si possono apporre banalmente dei segnalini strumentali al recupero del file con programmi di scrittura (da Word a Ventura) dimodoché questo con delle macroistruzioni li rintracci e rimpiazzi convenientemente: così si ottengono grassetto, corsivo, maiuscoletto e quant'altro, in alternativa alla possibilità di ottenerli con comandi ad hoc del linguaggio stesso, detti sequenze di *escape* e che hanno il difetto di essere vincolati alla stampante in uso.

Il linguaggio di formattazione comprende istruzioni di logica condizionale (Se... Allora, con almeno due comandi distinti: If e Nx Dx) per cui è possibile selezionare x e agire su y. Ossia, il contenuto può venire mostrato ecc., direttamente, ma anche sfruttato mediatamente, cioè usato come una condizione per cui, in sua assenza o presenza o in ragione del fatto che è uguale o diverso da un altro valore (una data, una serie di lettere) etc., ne scaturisce una diversa azione su questo o su altri campi; tali condizioni possono diventare complesse per numero e per livello logico se si decide di combinarle ed annidarle. Vi sono anche comandi di formattazione che non dipendono dal contenuto dei campi o dalla loro esistenza, sono cioè incondizionati. Le funzioni matematiche sono limitate a quanto contenuto in un record. La già citata funzione di collegamento (Ref) è capace — grazie all'indice — di richiamare da un documento quanto appartiene a un altro, o più, e che rappresenta l'embrione di relazione dei dati contenuto in questo *file manager*.

Con tali istruzioni a disposizione si possono costruire formati secondo gli schemi di citazione voluti, da Isbd/Ifla a Mla, da Harvard style a Turabian ecc., misurandosi con il modo in cui i dati sono stati registrati nei record. Si ottengono output su disco in formato *Cdf, tagged, Rtf (rich text format)* funzionali all'esportazione in *dbIII, Texto, Pro-Cite, Library Master*, database statistici, fogli elettronici, wordprocessor, ecc., potendo decidere numero, ordine ecc. dei campi da esportare e assicurando la circolazione dei dati e delle informazioni.

Mancano istruzioni per selezionare le occorrenze in base al numero (chi può lo fa in Isis-Pascal; cfr. oltre) e porzioni del contenuto



dei campi in base a stringhe di caratteri e non solo alla posizione. Il linguaggio di formattazione è composto di non molte istruzioni, che in un formato di visualizzazione vengono iterate molte volte, finendo col costituire una sequenza di comandi che spaventa e discaccia chi ignora gli elementi di base della sua natura. Effettivamente, al di là della ripetizione e dell'aspetto, che i formati si debbano scrivere, vedere e correggere, a tutto schermo e non su righe distinte, come in genere accade ai codici sorgente dei programmi, non aiuta la lettura e l'intervento. In realtà difficilmente si arriva ad espressioni complesse bensì ci si accontenta di sequenze composite.

Output

Il contenuto di ogni output è preparato col linguaggio di formattazione. L'output può venire scritto come file su disco o carta quando è il risultato di un'azione di stampa o di un'esportazione, mostrato su schermo quando è l'esito di una ricerca. La procedura di stampa comprende istruzioni generiche di impaginazione e di ordinamento dei record. Le istruzioni ge-

neriche di stampa impartibili in Isis danno impostazioni circa: formato di presentazione dei dati, dimensione del foglio, numero di colonne e di righe, caratteri per riga, numerazione delle pagine, salto a nuova pagina, titolo corrente in alto... etc. Non vi sono comandi, oltre quelli di escape del linguaggio di formattazione, per specifici tipi di stampanti. Le righe sono chiuse da *Cr/Lf* Ascii 13/10, non v'è divisione, neanche meccanica, delle parole che restano scritte per intero su una riga. Di routine si stampano schedine catalografiche, indici, bibliografie a soggetto, cataloghi a volume e a colonne. Non si manda una stampa su schermo in *preview* e non si può interrompere una stampa senza brutalità.

L'ordinamento alfabetico è sempre una fase di un'operazione di "stampa", anche se usato per esportare, e magari poi reimportare i dati. Esso pone in gioco altri elementi: quattro livelli gerarchicamente annidati di chiavi di ordinamento (1. chiave, a parità o in assenza la 2. ecc.) che formano la fascia in cui vengono pesati i valori, i caratteri che governano l'ordinamento — ad es. il mese o un codice Dewey espresso in numero —, e dove ogni livello può contenere varie chiavi diverse. A ciò si aggiunge la fascia in cui una stringa di caratteri può venire mostrata come intestazione in rapporto alle chiavi citate. La stringa può coincidere con la chiave che la origina: ordino un indice dei nomi in base ai cognomi e stampo questi, oppure può essere diversa, del tutto (ad es. mostro il nome del mese a fronte del numero usato come chiave) o in parte (stampo quell'indice dei nomi antepo- nendo prenome a cognome). Opzioni già preparate guidano al controllo della stampa delle intestazioni, per schedine catalografiche o per indici, anche relativi a più campi, che riman- ➤

dino solo al numero di record, o a nulla. Passando per l'ordinamento di tutti i record di un archivio — la loro esportazione ed importazione, nonché la rigenerazione dell'*inverted file* — si perviene a riordinare alfanumericamente tutto un database. Si può usare un file di stopwords ad hoc per l'ordinamento, si può basare il sort su tre livelli di chiavi e mostrarne solo due. Si può estrarre solo la prima occorrenza di un campo o tutte: è l'unico ambito di Isis in cui esiste tale opzione. La lunghezza massima della chiave di sort, non dell'intestazione, è di 4.000 caratteri. Manca un comando per l'ordinamento discendente (da Z ad A) ma è imponibile con un intervento manipolatorio ben più profondo. Con un po' di fatica e di cammino si arriva anche a produrre un indice come quello dei nomi dei tipografi, contestualizzati dai luoghi in cui hanno operato, quando in uno stesso record possono esserci più tipografi e più luoghi. Si produce un indice Kwoc (in differita anche Kwac), ma non uno Kxic. Si stampano intestazioni nella forma cara a bibliografie anglosassoni come *dashbed entry*.

Ricerca

Esistono tre modi di eseguire la ricerca: 1) scorrere il catalogo, 2) formulare richieste e scandire sequenzialmente il database, 3) formulare richieste ed accedere ai dati tramite l'indice trasposto (*inverted file*), vuoi consultandolo vuoi interrogando direttamente. Mancano svolgimenti della ricerca per *query by example*, o per qualsivoglia tipo di navigazione o andatura ipertestuale.

Lo scorrimento del catalogo utilizza solo il numero di record come chiave di accesso e mostra i record uno dopo l'altro nel formato

di visualizzazione di default, o in altro che si trascoglie.

La ricerca sequenziale scandisce il database sin dal primo record, comunque, ma non necessariamente fino all'ultimo, e se ne può restringere il risultato a quanto reperito in un sottoinsieme del database. È operazione senz'altro più lenta rispetto all'interrogazione dell'indice, ma presenta vantaggi peculiari proprio perché non passa attraverso esso. L'indice può non esistere — non essere generato o aggiornato — e si può tuttavia ricercare; si possono cercare dati in campi non indicizzati; si usano comandi del linguaggio di formattazione che non funzionano quando si ricerca interrogando l'indice. Solo qui, infatti, si impiegano espressioni relazionali (con gli operatori \geq \leq $=$ $>$ $<$ $:$) che servono a confrontare i valori contenuti nei campi con delle costanti o col valore di altri campi, e funzioni booleane, verificanti la presenza/assenza dei campi.¹⁰ Più espressioni relazionali possono venire combinate in un'unica espressione complessa anche con varie funzioni booleane. Solo qui si possono cercare voci troncate a sinistra e contemporaneamente a sinistra e a destra senza carattere di troncamento. Non si usano operatori di prossimità. La sintassi dei comandi è complicata per un utente non esperto ed assolutamente inattuabile per chi vede il programma la prima volta. Ogni richiesta è rapportata ad uno o più campi che vanno indicati, tramite il loro Tag definito nel trac-

ciato record, di cui dunque occorre avere documentazione. Mancano istruzioni di mascheramento (*wildcards: Ma??a* per reperire: *Marta, Manna, Malva, Massa* ecc.), e un esplicito operatore di *range* per indicare una fascia continua di elementi: *da 1965 a 1975*, occorre costruirselo come *v7>='1965' and v7<='1975'*.

Quando la ricerca passa per l'indice può esordire in due modi: un modo del tutto *command driven* in cui si sottopongono subito espressioni alla maniera tradizionalmente in uso nell'interrogazione dei database on-line, oppure la previa perlustrazione dell'indice a video, la selezione dei termini ed il ritorno col bottino scritto nello schermo da cui comunque si sottopongono richieste.

La ricerca tramite l'*inverted file* non legge direttamente cosa si trova nei record, ma solo cosa è nell'indice. Ciò ha un legame con quanto presente nei record non di piatto rispecchiamento, ma di riflesso articolato in base alle istruzioni date da chi disegna l'indice. Costui ne ha totale controllo sia per cosa che per come viene indicizzato grazie all'uso congiunto del linguaggio di formattazione e delle nove tecniche di indicizzazione previste.

Che tipo di voci sono registrabili nell'indice, ed in genere sono indicizzabili in Isis? Sostanzialmente due: parole singole o "frasi" come sequenze di più parole.¹¹ La sequenza può corrispondere ad un campo, o a un sottocampo, o a quanto deliberatamente marcato a fini di indicizzazione fra $\langle \dots \rangle$. Tutte le voci, parole, frasi, porzioni di parole e di frasi, possono venire prefissate e in parte suffissate. Un file testo di *stopwords*, interamente definibile dall'utente per ciascun database, può registrare sino a 799 parole lunghe fino a 10 caratteri, che dunque non verranno comprese fra le parole ricerca-



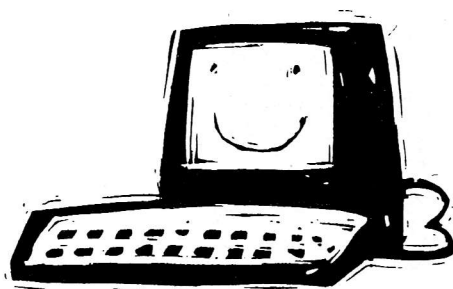
bili singolarmente: è normale usarlo per parole non significative in ricerca.

Tutti i campi di un record possono venire indicizzati, lo stesso campo può venire indicizzato più volte in modo diverso, entro il limite massimo delle 200 istruzioni ammesse nella tabella che governa l'*inverted file*.

L'*inverted file* mantiene precisa notizia della posizione delle voci in un analitico indirizzo (*posting*), e ciò è anche ragione delle dimensioni dell'*inverted file*, del tempo che ci vuole a crearlo e ad aggiornarlo e della possibilità di usare l'*E* logico con tutti i suoi raffinamenti. L'indirizzo registra: 1) tutti i numeri di record in cui un termine dell'indice compare; 2) l'identificatore numerico del campo (grosso modo il "Tag") in cui il termine ricorre; 3) il numero di occorrenza, se è un campo ripetibile; 4) il numero di sequenza che il termine occupa all'interno dei termini del campo o occorrenza.

L'*inverted file* è anche il ponte che consente l'instaurarsi della relazione fra i record dettata con comandi del linguaggio formattazione ed il cui esito apparirà in output ed anche in esportazione dei dati.

In consultazione l'*inverted file* è mostrato come un unico indice/dizionario in cui tutte le voci sono lunghe al massimo 30 caratteri e di norma, ma non obbligatoriamente, convertite in maiuscolo con appiattimento delle lettere accentate (é = è = E = É ecc.). Tutti i termini, di tutti i campi per cui è stata apprestata indicizzazione finiscono in esso. Si può perlustrare l'indice e scegliere i termini già con l'operatore booleano voluto per la ricerca. Se il termine richiesto non esiste, si ha la visualizzazione dei termini alfabetici prossimi a quello. Non si vede il numero di documenti collegati alle voci (memento: l'indice non serve a



nulla in catalogazione, non è nemmeno visibile).

Selezionati i termini dall'indice si è poi riportati allo schermo in cui si formulano le espressioni e si avvia la ricerca. La creazione ed aggiornamento dell'*inverted file* sono sempre, comunque, differite. Chi opera può accettare l'aggiornamento proposto automaticamente alla fine di ogni sessione di lavoro o rinunciare posponendolo. La durata dell'operazione è senz'altro influenzata dal tipo di macchina in uso, dalle dimensioni del database e dei record da aggiornare, dalla struttura dell'indice che determina quante voci vanno indicizzate e come; l'aggiornamento è in assoluto una procedura lenta. Durante lo svolgimento delle operazioni sull'*inverted file*, Isis non è utilizzabile per altro; con database di migliaia di record esse vanno annoverata fra i servizi di manutenzione.

Quanto ai comandi: è ammesso il troncamento a destra, e solo quello, di ogni termine, con esplicito carattere. Sono disponibili gli operatori booleani *E*, *O*, *NON*. L'*OR* esclusivo si ottiene a mano, l'*AND* logico ha vari livelli più specifici come operatori di compresenza dei termini nel campo, nell'occorrenza, entro un certo numero di termini, ad una distanza esatta. Naturalmente i vari operatori e i termini sono combinabili, anche fra parentesi, in espressioni lunghe e complesse. Si possono apparentare dei termini, ad esempio i comuni della provincia di Cosenza, e cercarli tutti — connessi da *O* logico — mediante la sola

voce *Cosenza*, preceduta da *any*: deve essere stato scritto, o copiato per ogni database, un file testo, detto appunto *any file* che contiene questi apparentamenti, *cluster*, di voci, ciascuno su una linea distinta: usarlo per sinonimie è solo un modo fra i vari.

Per limitare la ricerca di un termine ad uno o più campi occorre indicarne esplicitamente l'identificatore numerico, che può, e quindi può non, coincidere col "Tag". Un errore di sintassi è immediatamente rilevato e codificato secondo un numero decifrabile solo per chi consulta il *Reference*.

Come già detto, la funzione di ricerca è preliminare alla catalogazione solo se così decide chi opera. Ai tre modi di accesso descritti (scorrimento dei record, ricerca con espressioni sequenziale o tramite l'indice) corrispondono tre distinte opzioni nel menu. L'esito finale è il successo o l'insuccesso, ma nella ricerca tramite indice viene sempre mostrato a video un riepilogo in cui per ogni elemento utilizzato nella richiesta è riportato il numero di occorrenze nell'indice e di record effettivamente reperiti. Non c'è classificazione del risultato in base a *best match and ranking* o a *weighting* dei termini. Le ricerche eseguite nel corso della sessione possono venire mostrate a video col totale numerico dei record reperiti, possono venire richiamate tramite il loro numero di sequenza e combinate con altri termini o altre espressioni, così — tra l'altro — superando il limite di lunghezza di 255 caratteri ammesso per qualsiasi espressione. Le espressioni di ricerca — strategie — non possono venire accantonate e richiamate oltre una sessione. Si vede l'esito di una ricerca con un colpo, si va in correzione di esso con tre, si cambia il formato con cui visualizzarlo a video e si vede l'effetto con quattro, più il nome del formato. I termini ►

cercati e trovati non sono evidenziati sullo schermo nel loro contesto (*highlighting*). I record reperiti formano un sottoinsieme accantonabile: si può salvare un solo esito di ricerca alla volta che è riutilizzabile come input in tutte le operazioni. Non si deseleggono record dal gruppo di quelli reperiti: si può arrivare a farlo a mano, ma è spesso più semplice intervenire con uno strumento di scrittura sull'output su disco. Si può cercare per numero di record quale che sia l'approccio trascelto.

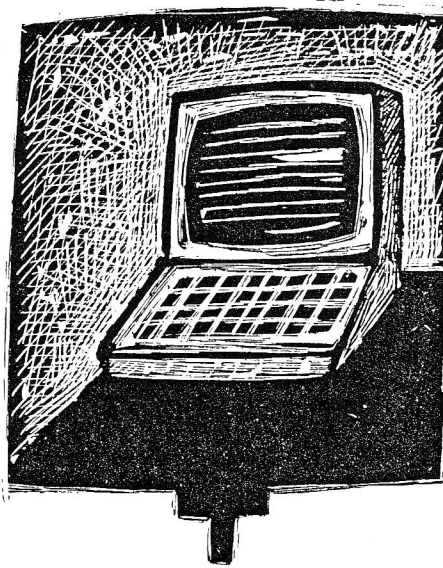
Non esiste propriamente una funzione per il thesaurus: quanto viene talvolta citato impropriamente come tale è il frutto di un esercizio di applicazione del linguaggio di programmazione Isis-Pascal che funziona solo in ricerca, con navigazione nella struttura ad albero, selezione dei termini, visualizzazione delle scope notes. È fisicamente un database distinto da quello dei dati il cui pregio certo non sta nella scorrevolezza della creazione e manutenzione. ■

Note

¹ Precisamente: *Micro Cds/Isis Computerized Documentation Service, Integrated Set of Information Systems*, Unesco®. La versione micro disponibile anche per Pc Wang, per Unix-Sco e per Vax/Vms; per altri componenti della famiglia Isis si veda la parte II che sarà pubblicata prossimamente in questa rubrica. Ho analizzato la versione Dos 3.07, novembre 1993. Manoscritto terminato il 12 settembre 1994.

² E.G. SIEVERTS [et al.], *Software for Information Storage and Retrieval Tested, Evaluated and Compared. Part 2, Classical Retrieval Systems*, "The Electronic Library", 9(1991), 6, p. 301-317.

³ UNESCO, Division of Software Development and Applications. Office of Information Programmes and Services, "Mini-micro Cds-Isis Reference Manual (version 2.3)", Paris: Unesco,



1989, xiv, 285, ISBN: 92-3-102-605-5; cui seguono vari aggiornamenti su dischetto; traduzione in italiano a cura della DbA di Firenze, Ed. Titivillus, nel 1992, aggiornata alla versione 3.0.

⁴ Per creare un database copiandone un altro, occorre copiarne con nuovo nome i 12 files in ambiente Dos, fuori da Isis; si va poi dentro il tracciato record, la cosiddetta *Fdt-field definition table*, con uno strumento di scrittura, si trovano i nomi dei file legati fisiologicamente all'archivio usato per clonare: vanno cambiati con nomi appropriati al nuovo database. Nè è semplice condividere a posteriori, fra diversi database, tabelle, formati, maschere: per farlo occorrerà creare dall'interno del database ricevente un finto nuovo esemplare, poniamo, di formato e poi copiarci sopra, da Dos, l'originale, nel caso rispettando convenzioni onomastiche. Maschere di stampa e di sort ed altri menu non sono considerati fisiologicamente indispensabili ad un database e dunque non vengono creati con esso, per se ne può equipaggiare un archivio, seguendo convenzioni onomastiche — che vi risparmio — e di indirizzo fisico su disco. Nulla di intuitivo e lineare, niente di difficile, molta libertà nella sostanza.

⁵ Qualcuno toccato dal limite di 16 milioni di record per database? Altre restrizioni sono di rilevanza pratica: i quasi 8.000 caratteri per record — o per campo — sono un'enormità per

un record bibliografico o per un indirizzo, ma un moncone per una banca dati full text. In output si può arrivare a >32.000 caratteri se si usa la memoria espansa Ems, ed ha rilievo visto che si possono correlare più record in fase di output.

⁶ Un equilibrio acrobatico: il numero di sistema di ogni record assolutamente soggetto a cambiamenti ad ogni più che probabile riorganizzazione di archivio. Se invece mi servo dell'indice, esso deve essere sempre aggiornato, le istruzioni per stabilire il legame sono più di una e non semplicissime. L'indice stesso, a rigore, non può esprimere il legame, giacché serve a crearlo: ma ci si arriva con dei marchingegni.

⁷ Con operazioni di *backup e restore* o import/export, e comunque, se si tratta di un errore, i dati del master file si recuperano — fuori da Isis — con un programma di *recovery* sviluppato dall'Unesco ad hoc, anche a prescindere da queste ricompattazioni.

⁸ Differenze: il file Isis viene scritto su righe da 80 caratteri, tutte chiuse da Cr/Lf, l'ultima pu essere più breve, Iso 2709 scrive di continuo a lunghezza variabile; Iso non prescrive un unico delimitatore di sottocampo mentre Isis usa il circonflesso ^ che è anche parte del contenuto del record; in Iso 2709 il separatore di record è "Is;" di Iso 646, Isis usa # e lo usa anche come separatore di campo, che in Iso 2709 è "Is;" sempre di Iso 646. Con il *Ccf Converter* di P. Simmons, si convertono — fra l'altro — i due Iso 2709 in entrambe le direzioni.

⁹ Il programma, gratuito, frutto di una collaborazione volontaria con l'Unesco da parte di H. Besemer e P. Nieuwenhuysen. Il lavoro fu motivato proprio dalla difficoltà di reperire sul mercato o nel *public domain* programmi di conversione verso Iso 2709.

¹⁰ Cercare i record solo in base al fatto che uno o più campi siano pieni o vuoti, del resto fattibile anche nella ricerca tramite l'indice, se chi lo definisce dà le istruzioni opportune.

¹¹ Le parole singole sono sequenze di caratteri riconosciuti in una tabella ad hoc, interrotte da un carattere ivi non riconosciuto (ad es. il *blank* Ascii 32); una frase è una sequenza di parole, lunga fino a 30 caratteri.