

M. AGOSTI - M. E. CRESCENTI - F. LESTUZZI - P. SCHIAVON

Sistemi informativi: DOC-4 e SCRIN

Estratto da: I Quaderni dell'Elaborazione Automatica

SISTEMI INFORMATIVI: DOC-4 E SCRIN

MARISTELLA AGOSTI

(Istituto di Statistica, Università di Padova)

MARIA EMANUELA CRESCENTI

(Centro di Calcolo, Sezione Scientifico-didattica, Università di Padova)

FRANCO LESTUZZI

(Centro Studi della Barbariga, Strà, Venezia)

PAOLO SCHIAVON

(Centro di Calcolo, Sezione Scientifico-didattica, Università di Padova)

Ricevuto in data 5 luglio 1977

SOMMARIO — Si descrivono due sistemi di documentazione automatica messi a punto su un minicalcolatore. Entrambi i sistemi sono colloquiali, richiedono poca memoria per il loro funzionamento ed hanno dei tempi di risposta estremamente bassi. Benché l'organizzazione dei dati e la logica interna di funzionamento siano molto diverse, entrambi i sistemi utilizzano lo stesso linguaggio di interrogazione per facilitare l'uso all'utente.

ABSTRACT — This paper describes two Information Retrieval systems developed on a minicomputer. Both systems operate on real-time basis, require size memory small and have a low response time. Though data organization and internal logic are very different, both systems have plain and easy to use query language.

1. - PREMESSA.

L'archiviazione ed il ritrovamento dei dati è attualmente uno dei più importanti problemi dell'elaborazione automatica delle informazioni.

Quando si pensa che esistono biblioteche con milioni di volumi risulta chiaro che nessuna tecnica manuale è in grado di garantire un aggiornamento ed un accesso in tempi ragionevoli.

Inizialmente il problema è stato studiato solo per gli archivi di grandi dimensioni (biblioteche nazionali, istituti bancari, ecc.).

Esistono però anche, e sono anzi più numerosi, archivi di dimensioni medie (fino a qualche decina di migliaia di documenti) per i quali una procedura automatica può essere estremamente vantaggiosa: si tratta degli archivi che definiremo medi omogenei, in quanto raggruppano documenti che hanno fra di loro una elevata affinità. Rientrano per esempio in questa classe le biblioteche di istituto universitario (i testi normalmente riguardano argomenti specifici) l'insieme delle cartelle cliniche di un ospedale, le fototeche, le cineteche, ecc.

Vogliamo ora descrivere due modelli sviluppati per le biblioteche di due istituti.

Si fa notare che anche se più avanti si farà riferimento esclusivamente alle applicazioni specifiche già sviluppate, gli stessi modelli possono essere facilmente utilizzati per altri archivi medi omogenei.

Le ricerche in una biblioteca di istituto sono essenzialmente di due tipi:

- 1) un testo del quale si conoscono già i dati bibliografici essenziali;
- 2) un insieme di testi che sono collegati fra loro da connessioni logiche particolari e difficilmente prevedibili a priori.

Il caso 2 è il più frequente durante la fase iniziale di documentazione di un nuovo lavoro. In questo caso infatti si ricercano per esempio i libri che trattano di alcuni specifici argomenti, o che sono stati scritti da un determinato autore eventualmente eliminando quelli pubblicati prima di una certa data. In base ai risultati di questa ricerca si passa poi al ritrovamento del testo che interessa.

Si può notare quindi che le esigenze dell'utente di queste biblioteche sono di tipo particolare: infatti molte volte il ricercatore porrà la sua domanda in termini piuttosto vaghi, salvo precisarla ulteriormente quando avrà assunto nuove informazioni.

Entrambi i modelli che verranno descritti più avanti sono stati studiati in particolare per rispondere a questa esigenza: si tratta cioè di sistemi colloquiali nei quali l'utente interroga il calcolatore ed ottiene in risposta un insieme di documenti.

Egli può successivamente modificare la propria domanda in modo da renderla più precisa, fino ad ottenere una risposta soddisfacente.

I sistemi consentono inoltre di ottenere informazioni più dettagliate su testi specifici.

Tutte queste operazioni avvengono in tempo reale. Il tempo di attesa per l'utente è quindi di pochi secondi.

Un ulteriore aspetto estremamente interessante di questi due modelli è che sono stati sviluppati utilizzando un minicalcolatore. Questo fatto è molto importante sia perché dimostra quali siano le possibilità attuali dei minicalcolatori nella gestione di banche di dati, sia per le applicazioni pratiche, in quanto le caratteristiche della macchina sono perfettamente compatibili con le possibilità e le esigenze di una biblioteca di medie dimensioni.

Il calcolatore utilizzato è un NOVA 840 con 80 K bytes di memoria centrale (quando i modelli sono stati sviluppati e messi a punto la memoria disponibile però era di 64 K bytes). La memoria di massa è costituita da due dischi da 2,5 M bytes. Le periferiche collegate sono un terminale video, un terminale stampante tipo teletype funzionante a 300 bauds, un lettore di schede, un lettore ed un perforatore veloci di banda.

2. - Lo SCRIN.

2.1. *Introduzione.*

Lo SCRIN (Sistema Colloquiale per la Ricerca dell'Informazione) è stato studiato e messo a punto per una biblioteca specializzata (di linguistica). Le sue caratteristiche però sono tali da renderlo adatto a trattare un qualunque archivio di dati(*) che debba essere consultato secondo schemi non prevedibili utilizzando particolari parole chiave dette descrittori.

I descrittori sono termini associati ad ogni testo e che ne descrivono in qualche modo il contenuto, o hanno col testo una attinenza logica.

A differenza dalle normali parole chiave un descrittore può anche non essere contenuto nel testo a cui è associato.

Un descrittore può essere costituito da più parole (al limite da una frase intera).

Per esempio per quest'articolo alcuni possibili descrittori sono:

(*) Più avanti si parlerà sempre di testi implicitamente facendo riferimento al sistema attualmente implementato, fermo restando che le stesse considerazioni si possono fare per archivi di dati di tipo diverso (es. leggi e decreti, cartelle cliniche, ecc.).

Information retrieval, minicomputer (applications), hash-code(**).

Estendendo questo concetto anche gli autori od il titolo possono essere considerati descrittori ed in effetti lo SCRIN li tratta come tali.

Questo, come si vedrà più avanti, consente una notevole libertà nella domanda dell'utente.

Si vuol far notare infine che i descrittori così come sono stati definiti risolvono immediatamente situazioni che sarebbero altrimenti ambigue con le normali parole chiave: per esempio esistono nell'attuale implementazione i due descrittori « language of philosophy » e « philosophy of language ». Questi pur esprimendo concetti totalmente diversi risulterebbero indistinguibili in un normale sistema a parole chiave.

L'associazione testo-descrittore non è automatica ma deve essere fatta da una persona preparata. Si è constatato che questo metodo porta ad una maggiore precisione nell'identificazione dei testi.

L'utente dello SCRIN esprime la propria domanda come una sequenza di descrittori e/o di nomi di autori collegati fra loro dagli operatori logici AND, OR, AND NOT.

Il sistema provvede ad individuare l'insieme di testi associati ad ogni descrittore ed ad effettuare le operazioni richieste fornendo alla fine un insieme di testi che soddisfano alla domanda.

Verranno ora descritti i files del sistema, i programmi di aggiornamento e consultazione ed infine le attuali possibilità dello SCRIN e gli sviluppi previsti.

2.2. *Organizzazione dei dati.*

I dati che devono essere ritrovati vengono conservati tutti in un file (file BASE), mentre le chiavi di accesso (descrittori, nomi di autori) si trovano in files invertiti. Gli agganci tra questi files (si tratta di alcuni files di indirizzi) sono totalmente inaccessibili e trasparenti per l'utente.

I records del file BASE sono di lunghezza variabile al fine di ottenere il massimo sfruttamento dello spazio disponibile. Ogni record è diviso in più campi (numero di magazzino, autore, titolo, sottotitolo...) ognuno dei quali è pure di lunghezza variabile ed indirizzabile separatamente tramite opportuni puntatori. Alcuni di questi campi sono obbligatori

(**) Nel caso di un archivio di cartelle cliniche p. es. alcuni descrittori potrebbero essere la malattia, gli interventi, le medicine, ecc.

(cioè sono sempre presenti p. es. il titolo del testo) altri (p. es. il sottotitolo, il riassunto) possono anche mancare. Quando un campo manca, per esso non viene riservato spazio nel record.

Al file BASE è associato un file di indirizzi (PUNTBASE) con records di lunghezza fissa ognuno dei quali contiene il punto di partenza del corrispondente record del file BASE.

Per la gestione del sistema tutto avviene come se questi due files costituissero un solo file di tipo sequenziale con indice.

Il file dei descrittori (*) ha records di lunghezza fissa, essi contengono il nome del descrittore, il numero di testi associati ed un puntatore ad un file dove si trova l'elenco degli indici ai testi associati. All'interno del file i descrittori vengono mantenuti ordinati alfabeticamente in quanto la ricerca viene effettuata con un metodo dicotomico. L'accesso ai dati del file BASE avviene dunque secondo il seguente schema:

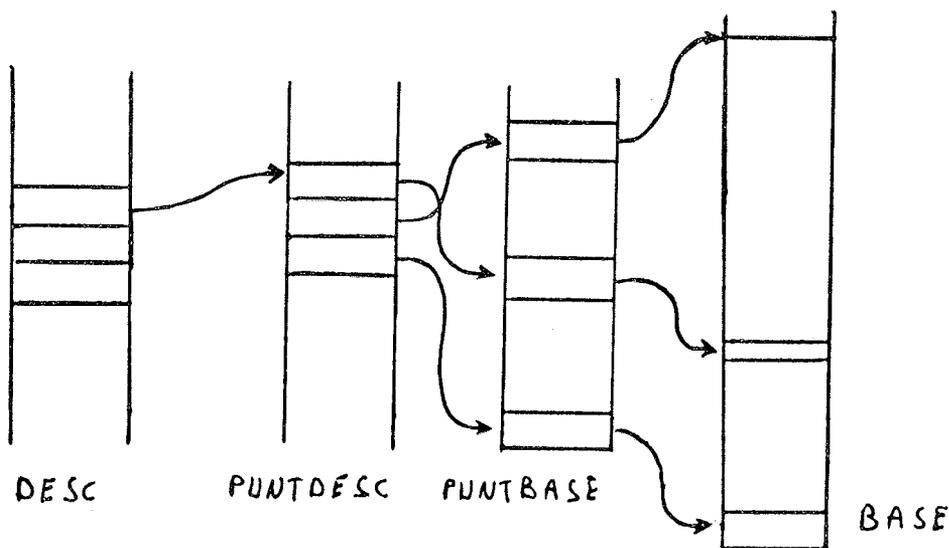


Fig. 1.

(*) Si parlerà genericamente dei descrittori, però tutte le considerazioni restano valide anche per gli autori. Nel sistema attualmente implementato il file degli autori è separato da quello dei descrittori propriamente detti. Sarebbe possibile però riunire i due files oppure aggiungerne altri per avere altri tipi di ricerca (esempio per titolo).

Cioè trovato un descrittore il sistema ricava l'insieme degli indici associati e da questi tramite il file PUNTBASE trova gli indirizzi nella base dei dati ai testi richiesti.

I programmi di gestione dello SCRIN sono stati studiati in modo da renderne l'uso il più semplice possibile per l'operatore e per l'utente, ed allo stesso tempo in modo da proteggere il più possibile la base dei dati da eventuali errori.

2.3. *Aggiornamento e consultazione.*

Il funzionamento dello SCRIN può essere diviso in due fasi: una fase di aggiornamento ed una di consultazione (un programma speciale da utilizzarsi una volta sola, provvede alla generazione iniziale del sistema creando tutti i files necessari, ma vuoti).

La fase di aggiornamento a sua volta è divisa in due parti:

- aggiornamento del file BASE
- aggiornamento del file descrittori.

L'operatore incaricato di queste operazioni deve limitarsi a preparare i nuovi dati in ingresso (nuovi testi da aggiungere) ed a richiamare nell'ordine i due programmi.

Il primo programma esamina i dati nuovi, verifica che questi rispettino una struttura formale predefinita (nel nostro caso una particolare scheda bibliografica), li aggiunge al file BASE e crea gli agganci necessari all'interno del file PUNTBASE.

Il secondo programma accede direttamente alla base dei dati, esamina i nuovi testi introdotti e provvede ad aggiornare il file DESCRITTORI ad il file degli indici associati.

L'esecuzione di questi programmi è totalmente automatica e non richiede alcun intervento da parte dell'operatore. Eventuali errori vengono segnalati con opportuni messaggi, e non comportano mai la perdita dei dati preesistenti all'operazione di aggiornamento.

È stato messo a punto un programma speciale per effettuare lo spoglio delle riviste; per evitare di dover ripetere, per ogni articolo, un certo numero di informazioni costanti (titolo della rivista, collocazione, luogo di pubblicazione, ecc.), questi dati vengono conservati su un file a parte. Facendo lo spoglio di una rivista è sufficiente dare il numero di codice della rivista stessa, il numero e la data di pubblicazione e i dati essenziali di ogni articolo (autore, titolo, descrittori e riassunto). Il sistema provvede a completare questi dati con le informazioni

fisse e prepara un file che viene passato al normale programma di aggiornamento.

La fase di consultazione come si è già detto è di tipo colloquiale. La domanda che l'utente pone allo SCRIN si presenta come una sequenza di descrittori collegati con operatori logici.

Non vi sono restrizioni nel numero di descrittori che possono essere specificati né sull'ordine (autori e descrittori propriamente detti possono essere mescolati).

Quando non è nota la grafia corretta di un descrittore questo può essere indicato in forma incompleta (naturalmente la richiesta diventa meno precisa; p. es. la domanda A = porta alla ricerca di tutti i descrittori che cominciano con la lettera A).

Lo SCRIN esamina la domanda da sinistra verso destra, individua per ogni descrittore l'insieme di indici associati ed effettua le operazioni logiche fra questi insiemi.

Non è stata stabilita alcuna precedenza fra gli operatori logici; per modificare l'ordine di esecuzione si possono usare le parentesi che possono essere inserite senza limite di numero e di livello.

Eventuali errori (es. descrittore non valido) vengono segnalati all'utente che può provvedere a riformulare la domanda.

Alla fine lo SCRIN sottopone all'utente un insieme di testi che soddisfano la richiesta. Questo insieme viene presentato sotto la forma di file sequenziale con indice: l'utente cioè può esaminare i testi ritrovati sia sequenzialmente (in forma diretta o inversa) sia casualmente. (Questo file non è fisicamente esistente, ma dal punto di vista dell'utente tutto avviene come se esso fosse stato costruito).

Nell'esaminare i testi l'utente può farsi stampare i dati bibliografici essenziali, l'elenco dei descrittori oppure il riassunto. Il sistema inoltre è collegato con un selettore automatico di microfiches.

La selezione del fotogramma corrisponde al testo desiderato avviene automaticamente ed è controllata dal calcolatore.

Questo consente di fornire all'utente un'informazione più completa (testi più lunghi o disegni).

Il programma di aggiornamento del file BASE può caricare circa 50 parte dell'utente che in nessun caso può alterare la base dei dati o altri files del sistema.

2.4. Prestazioni e limiti.

L'attuale implementazione è risultata del tutto soddisfacente sia dal punto di vista dei tempi di esecuzione sia per l'impiego di memoria

richiesto. Tutti i programmi richiedono solo 10 K parole (*).

Il programma di aggiornamento del file BASE può caricare circa 50 testi al minuto, mentre l'aggiornamento del file descrittori richiede tempi più lunghi che dipendono sia dal numero di dati nuovi sia dalle quantità di dati già presenti (il programma comprende una fase di SORT ed una di MERGE). Indicativamente si può dire che i tempi sono circa 2-3 volte i tempi necessari per aggiornare il file BASE.

Per quanto riguarda la fase di consultazione i tempi di risposta vanno da meno di 1 s. per domande semplici (un solo descrittore senza operazioni logiche) fino a 20-30 s. per domande complesse con una decina di descrittori ed operazioni logiche su insiemi di qualche centinaio di elementi.

Nel considerare questi tempi bisogna notare però che tutte le operazioni di I/O su disco vengono effettuate tramite buffers in memoria che vengono trasferiti fisicamente solo quando questo è necessario (il sistema conserva una traccia di tutte le porzioni di file già caricate in memoria). Questi buffers vengono allocati dinamicamente in modo da occupare tutta la memoria disponibile. I programmi risultano quindi tanto più veloci quanto maggiore è la partizione di memoria che viene loro assegnata (p. es. per la fase di SORT passando da 10 a 20 K si ha un guadagno di tempo di circa il 25 %).

I limiti dell'attuale implementazione dello SCRIN sono dati essenzialmente dal tipo di disco a disposizione e dalla necessità di limitare l'occupazione di memoria durante la fase di consultazione.

Un disco da 2.5 M bytes può contenere circa 5000 testi, tutti i files invertiti e tutti i programmi necessari.

Avendo a disposizione dischi di dimensioni maggiori è possibile arrivare a circa 20000 testi, oltre i quali è necessario modificare leggermente i puntatori per consentire l'indirizzamento di files più grandi.

Per limiti nella fase di consultazione la lunghezza globale di un singolo testo non può superare i 1200 caratteri mentre una richiesta non può superare i 252 caratteri (resta libero il numero di descrittori specificato).

Inoltre non ci possono essere più di 255 testi associati ad ogni

(*) I programmi sono tutti scritti in ALGOL al fine di assicurare una buona mantenibilità e trasportabilità. Questo ha portato un maggior impegno di memoria. Si è stimato che lo SCRIN fosse stato scritto in ASSEMBLER potrebbe funzionare in meno di 5 K parole.

descrittore e non è possibile superare i 511 testi durante le fasi intermedie delle operazioni logiche.

Aumentando però di solo 1 K l'impegno di memoria è possibile raddoppiare tutti questi numeri.

2.5. *Considerazioni finali.*

Lo SCRIN nella sua versione attuale è già soddisfacente per l'applicazione per la quale è stato studiato.

Sono allo studio parzialmente in fase di sviluppo alcuni perfezionamenti. Fra i più importanti si segnalano i seguenti:

- 1) introduzione di altri operatori logici (OR esclusivo);
- 2) collegamento con un selettore automatico di microfiches per la visualizzazione di dati non conservabili in disco (già in fase sperimentale);
- 3) un programma di parametrizzazione per poter descrivere al sistema dati di tipo diverso;
- 4) passaggio ad un sistema multi-user.

3. - DOC-4 SISTEMA INTERATTIVO DI DOCUMENTAZIONE AUTOMATICA.

3.1. *Introduzione.*

Il sistema DOC-4 realizza l'archiviazione e il reperimento di informazioni bibliografiche in tempo reale.

Nel sistema possono essere individuate tre fasi fondamentali:

- 1) fase di installazione del sistema;
- 2) fase di archiviazione di informazioni bibliografiche;
- 3) fase di recupero delle informazioni.

Queste tre fasi vengono realizzate da un package di tre programmi di tipo colloquiale; i primi due programmi possono essere utilizzati esclusivamente dai responsabili delle rispettive fasi; il terzo programma, invece, è specificatamente orientato all'utente di un sistema di documentazione automatica. L'utente che non conosca le modalità di interrogazione, o che desideri avere delle delucidazioni sulle stesse, oltre

ad avere la possibilità di consultare un manuale di utilizzo può avere delle delucidazioni dal sistema stesso. Infatti il sistema elenca, su richiesta dell'utente, quali sono le diverse possibilità di recupero delle informazioni, cioè la possibilità di eseguire la ricerca per titolo, per autore, per casa editrice o per parole chiave, ognuna corredata dalle relative modalità per formulare la richiesta.

Nella fase di installazione del sistema vengono stabilite le dimensioni dell'archivio delle informazioni in relazione alla biblioteca considerata.

La fase di archiviazione delle informazioni attua la costruzione dell'archivio delle informazioni, cioè l'inserimento delle informazioni relative all'insieme iniziale dei documenti considerati, ma attraverso questa fase vengono attuati anche tutti gli inserimenti relativi all'aggiornamento dell'archivio in relazione a nuovi documenti che accrescono il patrimonio documentario disponibile inizialmente. Se sono state introdotte delle informazioni errate, in questa fase si può procedere direttamente alla modifica di queste informazioni, sia che riguardino un documento nella sua completezza o siano relative solo a una sua parte; il sistema creerà automaticamente il collegamento con la parte delle informazioni corrette sostitutive di quelle precedentemente introdotte o effettuerà la cancellazione delle precedenti informazioni, se completamente errate, e archiverà quelle corrette. L'introduzione delle informazioni può avvenire indifferentemente da teletype o da scheda; l'introduzione di una grossa massa di informazioni può essere più agevole facendo eseguire la lettura da scheda, mentre l'aggiornamento di routine può essere più pratico ed economico eseguirlo da teletype.

3.2. *Metodo di classificazione.*

L'insieme dei documenti considerati è quello raccolto in una biblioteca specializzata. Quando si è affrontato il problema della classificazione di questi documenti si sono volute tenere presente due considerevoli esigenze degli utenti di una biblioteca; la prima è quella della possibilità di utilizzare, come linguaggio di ricerca automatica, un linguaggio il più vicino possibile al linguaggio naturale, la seconda è che la risposta sia mediata da un sistema di classificazione di tipo tradizionale in maniera tale che i documenti ritrovati possano anche essere ripartiti, come avviene usualmente, fra classi diverse. Il sistema DOC-4 realizza questi scopi utilizzando due diversi metodi di classificazione dei documenti collegandoli poi fra di loro come verrà esposto in seguito.

Il primo è il metodo dell'individuazione coordinata del contenuto semantico fondamentale di ogni documento mediante una lista di parole chiave associate al documento stesso. La scelta delle parole chiave dal documento deve essere fatta in maniera tale che ogni parola chiave individui un argomento fondamentale trattato nel documento stesso. Inoltre argomenti analoghi trattati in documenti diversi devono essere sintetizzati mediante la stessa parola chiave; in questo modo il linguaggio di ricerca è costituito da parole della lingua naturale, ma la ricerca risulta semplificata con l'eliminazione della sinonimia tra le parole chiave [1-2-3]. Le relazioni che intercorrono fra una parola chiave e il contenuto semantico da essa sintetizzato possono essere formalizzate come di seguito.

Si chiami $I = \{i_1, i_2, \dots, i_e\}$ l'insieme di tutte le informazioni elementari e $K = \{k_1, k_2, \dots, k_n\}$ l'universo delle parole chiave k_i . Dato il dominio K e il codominio $\mathcal{P}(I)$, insieme delle parti di I , si suppone l'esistenza di una funzione iniettiva f :

$$f: K \rightarrow \mathcal{P}(I)$$

$$f: k_i \rightarrow f(k_i).$$

Chiamiamo $e_1 = f(k_1), \dots, e_n = f(k_n)$, l'insieme $E = \{e_1, e_2, \dots, e_n\}$ è l'insieme universo degli argomenti; la generica k_i svolge, quindi, la funzione di designazione dell'argomento e_i . Si suppone inoltre che la funzione f sia tale che:

$$\text{se } i \neq r \quad e_i \cap e_r = \phi \quad \text{e} \quad \bigcup_{i=1}^n e_i = I.$$

Notiamo che supporre la funzione f con le suddette proprietà, è come dire che gli argomenti e_1, \dots, e_n sono una partizione dell'insieme I . Il supporre l'esistenza della funzione f è corretto, perché almeno una funzione f , che si può indicare con \bar{f} , esiste ed è quella che manda una parola chiave in un argomento e_i , costituito da una sola informazione elementare, cioè:

$$\bar{f}: k_i \rightarrow e_i \quad \text{dove} \quad e_i = \{i_i\}$$

e la numerosità $n = e$. La funzione \bar{f} è quella che determina la partizione più fine dell'insieme I ; sarà da ricercarsi, invece, una funzione f

che determina una partizione meno fine dell'insieme I , così che sarà necessario un numero inferiore di parole chiave per designare gli argomenti.

Si è adottato un sistema di ponderazione per spiegare quanto il contenuto semantico di una parola chiave viene esaurito in ogni documento al quale essa viene associata. Quindi nella fase di input delle informazioni bibliografiche per ogni documento si introducono, oltre alle informazioni usuali (titolo, autore, casa editrice, anno di pubblicazione), le parole chiave ad esso associate, ciascuna con un suo peso assoluto; il sistema automaticamente assumerà nulla l'associazione con tutte le altre parole chiave.

I dati, relativi al sistema di ponderazione, costituiscono una matrice $P = || p_{jk} ||$ di ordine $m \times n$, nella quale ad ogni riga corrisponde un documento e ad ogni colonna una parola chiave; il generico elemento p_{jk} è il peso relativo associato alla k -esima parola chiave associata al j -esimo documento. Infatti chiamato π_{jk} il peso assoluto che viene introdotto in fase di input $p_{jk} = \pi_{jk} / \sum_{j=1}^m \pi_{jk}$ e $\sum_{j=1}^m p_{jk} = 1$, quindi la matrice P è normalizzata per colonna. Il numero dei documenti, m , non è fissato a priori, se fosse fissato non si potrebbe tenere conto dell'aumento del fondo documentario della biblioteca; così pure il numero n delle parole chiave non è fissato a priori, questo per permettere l'introduzione di nuove parole chiave, via via che entrino nel lessico o che siano necessarie per individuare il contenuto di nuovi documenti (*).

Il secondo metodo di classificazione adottato viene denominato in letteratura [2] sistema di classificazione ad argomenti principali con codici associati. Poiché l'utilizzazione sperimentale del sistema DOC-4 si è effettuata classificando documenti che inferiscono all'informatica, si è deciso di utilizzare il sistema di classificazione adottato dall'ACM, la cui ultima stesura è del maggio 1976 [4].

Una caratteristica di questo tipo di sistema è quella di suddividere gli argomenti in alberi di piccole dimensioni completamente distinti l'uno dall'altro; i termini d'indagine si fanno corrispondere agli argomenti che non hanno ulteriori suddivisioni (argomenti atomici come per la classificazione per parole chiave), che sono le foglie di questi piccoli alberi; quindi un libro può essere attribuito esclusivamente alle foglie o

(*) L'unica limitazione riguarda il valore massimo di m e di n ed essa è dovuta al quantitativo di memoria di massa disponibile.

sottoclassi di classificazione. Ciò è come dire che gli argomenti che corrispondono alle foglie o sottoclassi sono disgiunti fra di loro e l'unione di tutti gli argomenti costituisce un ricoprimento dell'insieme $I = \{i_1, i_2, \dots, i_c\}$ delle informazioni elementari inferenti alla scienza cui si riferiscono i documenti della biblioteca in esame. Chiamato $S = \{s_1, \dots, s_l\}$ l'universo delle sottoclassi, la generica sottoclasse s_j designa l'argomento e_j sottoinsieme di I . Chiamiamo $E = \{e'_1, e'_2, \dots, e'_l\}$ l'insieme universo degli argomenti. Considerando S come dominio e E' come codominio, esiste una funzione biettiva g tale che:

$$g: S \rightarrow E'$$

$$g: s_j \rightarrow g(s_j) = e'_j.$$

L'utilizzazione di una classificazione di questo tipo porterebbe a collocare ogni documento in un'unica sottoclasse, così facendo, però, si verrebbe a preferenziare l'argomento trattato in maniera preponderante nel documento e quindi a perdere le informazioni riguardanti gli altri argomenti trattati nel documento. Per conservare la caratteristica della interdisciplinarietà di trattazione degli argomenti all'interno dei documenti si è data la possibilità dell'associazione di un documento a più sottoclassi. Anche questa associazione viene quantificata mediante un sistema di ponderazione che spieghi quanto il documento è legato ad ogni sottoclasse alla quale viene associato.

Questi valori vengono organizzati nella matrice $R = ||r_{ij}||$ d'ordine $l \times m$; dove l è il numero delle sottoclassi e m il numero dei documenti, come sopra. Chiamato ρ_{ij} il peso assoluto che viene introdotto in fase di input, fra le informazioni riguardanti il j -esimo documento, il generico elemento r_{ij} della matrice R è il valore relativo, che rispecchia il legame fra il documento j -esimo con la sottoclasse i -esima, che viene calcolato $r_{ij} = \rho_{ij} / \sum_{i=1}^l \rho_{ij}$, $\sum_{i=1}^l r_{ij} = 1$, e quindi anche la matrice R ha la particolarità di essere normalizzata per colonna.

Le informazioni raccolte con i due sistemi di classificazione esposti vengono riunite nella matrice $C = ||c_{ik}||$ di ordine $l \times n$ prodotto della matrice R per la matrice P . Con passaggi elementari si dimostra che anche la matrice C è normalizzata per colonna ($\sum_{i=1}^l c_{ik} = 1$), cioè dà la distribuzione delle informazioni individuate dalle parole chiave rispetto alle sottoclassi. Il generico elemento c_{ik} della matrice C , che è una ma-

trice di classificazione, indica quantitativamente il legame che intercorre fra la sottoclasse i -esima e la parola chiave k -esima. Per limitare notevolmente l'occupazione di memoria di massa si è elaborato un metodo che permette di aggiornare la matrice di classificazione C (quando vengono introdotte le informazioni relative ad un nuovo documento o quando i valori della C devono essere modificati perché sono state introdotte delle informazioni errate), senza dover memorizzare interamente la matrice $P = ||p_{ij}||$. Si è dimostrato che è sufficiente la memorizzazione del vettore ${}^mV = ||{}^mV_k||$ di ordine n , dove ${}^mV_k = \sum_{j=1}^m \pi_{jk}$. Se si indica con $C' = ||c'_{ik}||$ la matrice che si deve ottenere dalla $C = ||c_{ik}||$ dopo la introduzione delle informazioni relative al $(m+1)$ -esimo documento, si dimostra che il generico valore c'_{ik} è uguale al valore calcolato mediante la formula seguente:

$$c'_{ik} = c_{ik} \cdot \frac{{}^mV_k}{{}^mV_k + \pi_{m+1;k}} + \frac{\pi_{m+1;k}}{{}^mV_k + \pi_{m+1;k}} \cdot r_{i;m+1}$$

Se si indica con $C'' = ||c''_{ik}||$ la matrice che si deve ottenere dalla C tenendo conto dell'estrazione dal sistema di informazioni errate introdotte in relazione al documento r -esimo (qualsiasi), il generico valore c''_{ik} si dimostra che è uguale al valore calcolato con la formula seguente:

$$c''_{ik} = c_{ik} \cdot \frac{{}^mV_k}{{}^mV_k - \pi_{rk}} - \frac{\pi_{rk}}{{}^mV_k - \pi_{rk}} \cdot r_{ir}$$

3.3. Il linguaggio di ricerca.

La descrizione del linguaggio per la ricerca per parole chiave, utilizzando la forma normale di Backus, è la seguente:

$\langle \text{domanda} \rangle ::= \langle \text{frase} \rangle | \langle \text{frase} \rangle \langle \text{continuazioni} \rangle$
 $\langle \text{continuazioni} \rangle ::= \langle \text{continuazione} \rangle | \langle \text{continuazione} \rangle \langle \text{continuazioni} \rangle$
 $\langle \text{continuazioni} \rangle ::= \langle \text{operatore} \rangle \langle \text{frase} \rangle$
 $\langle \text{frase} \rangle ::= \langle \text{operando} \rangle | (\langle \text{frase} \rangle) | \langle \text{frase} \rangle \langle \text{operatore} \rangle$
 $\langle \text{operando} \rangle ::= \langle \text{lettera} \rangle | \langle \text{lettera} \rangle \langle \text{operando} \rangle$

$$\langle \text{lettera} \rangle ::= A|B|C|D|E|F|G|H|I|J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X| \\ Y|Z|-|'|/|0|1|2|3|4|5|6|7|8|9 \quad (1)$$

$$\langle \text{operatore} \rangle ::= \langle \text{spazio} \rangle O \langle \text{spazio} \rangle | \\ \langle \text{spazio} \rangle E \langle \text{spazio} \rangle | \\ \langle \text{spazio} \rangle \text{NON} \langle \text{spazio} \rangle$$

$$\langle \text{spazio} \rangle ::= b$$

Una caratteristica del linguaggio che non può apparire dalla (1) riguarda l'operando. $\langle \text{operando} \rangle$ è una qualsiasi parola di senso compiuto appartenente ad una delle seguenti quattro lingue: italiano, inglese, francese o tedesco; la parola chiave è formata con le ventisei lettere dell'alfabeto inglese, quando è una parola composta come spaziatore viene usato il trattino, se si deve accentare viene utilizzato l'apice. L'utente può decidere di effettuare la sua ricerca utilizzando parole appartenenti a una sola di queste quattro lingue, in questo caso specificherà al sistema la lingua nella quale vuole effettuare la ricerca; oppure specificherà al sistema l'intenzione di utilizzare nella stessa stringa di ricerca parole appartenenti alle diverse lingue. Nel caso della ricerca multilingue, il sistema è in grado di segnalare, se esistono fra le parole chiave, omonimi omografi ma in lingue diverse (il sistema in input di informazioni non li accetta nella stessa lingua); in questa eventualità il sistema farà una specifica domanda all'utente per sapere in quale lingua deve considerare quella parola chiave. La frase di richiesta dell'utente consiste, come risulta dalle (1), in una espressione formata da una successione di parole chiave legate fra loro da operatori logici. Si è preferito adottare la scrittura usuale degli operatori logici, per non introdurre un simbolismo che poteva essere ostico all'utente. La stringa viene risolta da sinistra verso destra. È possibile modificare la sequenza che risolve l'espressione utilizzando le parentesi.

L'utente, se non è soddisfatto della risposta avuta dal sistema alla sua prima richiesta, può o riformulare la domanda con $\langle \text{frase} \rangle$, oppure aggiungere una $\langle \text{continuazione} \rangle$ per modificare la risposta ottenuta; questa possibilità può essere utilizzata dall'utente in maniera iterativa, come risulta dalle (1).

Alla domanda dell'utente il sistema dà una prima risposta segnalando quali sottoclassi, in base al secondo tipo di classificazione utilizzato, sono state selezionate; ogni sottoclasse S_s è corredata di un peso

w_s che rispecchia l'attinenza di ogni sottoclasse selezionata con la richiesta dell'utente; quindi la risposta del sistema sarà del tipo:

$$S_1, w_1; S_2, w_2; \dots; S_s, w_s; \dots; S_k, w_k. \quad \text{Con } \sum_{s=1}^k w_s = 100.$$

La seconda risposta del sistema è l'elencazione dei dati essenziali relativi ai documenti selezionati in base alla richiesta dell'utente. Ogni documento D_t è corredato da un coefficiente α_t , chiamato coefficiente di attinenza, che spiega l'attinenza del documento con la richiesta dell'utente, infatti il coefficiente α_t viene calcolato in base alla formula:

$$\forall D_t : \alpha_t = \sum_{s=1}^k w_s \cdot r_{st}, \quad 0 \leq \alpha \leq 100.$$

3.4. *Struttura del sistema.*

L'organizzazione del sistema è stata progettata considerando specificamente le esigenze dell'utente, come è già stato detto, e la possibilità di assicurare qualsiasi tipo di modifica delle informazioni archiviate.

I file del sistema che vengono utilizzati più frequentemente durante le richieste dell'utente sono i file AUTORI, TITOLI, CASAED e KEYWORD, che si sono quindi costruiti come file random, indirizzabili mediante la tecnica hash. La funzione hash utilizzata è quella della divisione [5]; come legge di scansione è stata scelta la legge di scansione quadratica a coefficienti pesati [6] in quanto si riesce a eliminare il fenomeno dell'agglomerazione primaria e anche quello dell'agglomerazione secondaria. Gli altri file del sistema sono di supporto ai file precedentemente elencati e mediante un sistema di puntatori vengono indirizzati direttamente, i principali di questi sono il file MATCLAS, che contiene le informazioni della matrice di classificazione, il file SOTLIB che contiene le informazioni relative al sistema di ponderazione utilizzato per quantificare le informazioni relative all'attribuzione dei documenti nelle sottoclassi della classificazione adottata; l'occupazione su memoria di massa relativa a questi due file viene estremamente contenuta mediante particolari tecniche di compattazione. Altri file, come il file ORDER e il file CLASS, permettono una stampa più veloce delle liste alfabetizzate delle parole chiave e degli altri cataloghi di cui dispone la biblioteca.

Tutte le informazioni contenute nei file sono state organizzate in maniera tale da evitare ogni ridondanza.

Nel file MINIFIL vengono archiviati sistematicamente dal sistema i dati relativi al processo di archiviazione delle informazioni (ad es. il numero dei documenti, in relazione ai quali sono state introdotte le informazioni, il numero di parole chiave, il numero di case editrici diverse fra loro, etc.) e i dati relativi alle scansioni che si sono effettuate nei file random, per permettere delle valutazioni statistiche in relazione alla lunghezza di ricerca in questi file ai responsabili del mantenimento del sistema.

Il sistema è stato implementato nel linguaggio FORTRAN IV; l'occupazione del sistema su disco è di quasi 80 K bytes; l'occupazione massima di memoria è di 38 K bytes facilmente riducibili a 26 K bytes. Utilizzando un unico disco da 2,5 M Bytes come supporto per l'archivio delle informazioni e per il sistema, si possono archiviare informazioni relative a più di 7000 documenti bibliografici.

3.5. *Conclusioni.*

Si sono considerati i possibili sviluppi futuri del sistema DOC-4. Si sono considerate, in particolare, le possibilità di modificare la fase di installazione del sistema, in maniera tale che, oltre ad avere la possibilità attuale di stabilire inizialmente le dimensioni dell'archivio delle informazioni, si possa rendere il sistema in grado di gestire dati di tipo diverso e la possibilità di rendere il sistema multi-user.

Si considererà la possibilità dell'applicazione di metodi di cluster analysis per la realizzazione di una classificazione automatica delle parole chiave.

3.6. *Ringraziamenti.*

Gli autori desiderano ringraziare il Dr. Livio Colussi per le utili discussioni sulla definizione del metodo di classificazione.

4. - CONCLUSIONE.

Qui di seguito si riportano due esempi di interrogazione ai due sistemi presentati.

4.1. *Colloquio con lo SCRIN.*

L'utente sta cercando testi scritti da PIERCE, che trattino di *Grammatica trasformazionale*, oppure sulla *Teoria di Chomsky* (fig. 2).

Si nota la domanda posta dall'utente ed una serie di messaggi stampati dal sistema (questi sono immediatamente riconoscibili perché iniziano con ***).

Lo SCRIN ha trovato nella base dei dati 43 testi che trattano la *Teoria di Chomsky*, 39 testi sulla *Grammatica trasformazionale* e 3 testi scritti da PIERCE (si osservi che le iniziali del nome non erano state indicate nella domanda).

Alla fine di tutte le operazioni logiche rimane un solo testo.

Le informazioni archiviate vengono stampate su richiesta (***QUERY? ALL). Si può notare che il testo contiene entrambi i descrittori richiesti; lo SCRIN ha tenuto conto di questo fatto quindi il testo è stato contato una sola volta.

4.2. *Colloquio col DOC-4.*

L'esempio, qui sotto riportato, è relativo a tre diversi tipi di ricerca di informazione che l'utente può effettuare al sistema DOC-4; DOMANDA (l'utente può interrogare il sistema mediante una frase); AUTORE (l'utente richiede i documenti scritti da un determinato autore); CASA EDITRICE (l'utente richiede la lista dei documenti editi da una determinata casa editrice).

I messaggi stampati dal sistema sono preceduti da " == >".

Nella prima ricerca (fig. 3a) l'utente desidera individuare i documenti legati alle parole chiave Data-sciences oppure Data-structures (si noti che la parola chiave Data-sciences si è utilizzata nella classificazione dei documenti introdotti, col significato che si raccomanda in [4]).

Il sistema ha selezionato le tre sottoclassi Data-structures con attinenza 42, Generale di Information Retrieval con attinenza 34, Data-Base con ottinenza 11; la somma dei tre pesi non raggiunge il 100; questo significa che il sistema ha selezionato altre sottoclassi che non vengono stampate per la loro scarsa significatività. Ora il sistema chiede all'utente se è interessato ad avere le informazioni relative ai documenti selezionati; alla risposta affermativa fornisce il numero totale di documenti

selezionati e inizia a listare le informazioni. Ogni documento è corredato dal coefficiente di attinenza alla domanda (si noti che il sistema li fornisce ordinati rispetto a questo coefficiente); la stampa viene fatta di cinque in cinque documenti.

Al termine dell'esame dei documenti il sistema chiede se si è interessati alla formulazione di una domanda continuazione (fig. 3b); alla risposta affermativa il sistema scrive: " == > CONTINUAZIONE". L'utente, ora, vuole escludere le informazioni individuate dalla parola chiave Data-base (quindi scrive NON DATA-BASE). Come si può vedere nelle due risposte del sistema, di conseguenza, vengono modificati i valori di attinenza e la sottoclasse Data-base non viene più selezionata. (Si noti che l'ultimo documento compare anche in questa lista perchè è associato alla parola chiave « data-structures »).

Infine si danno due esempi di ricerca non per parole chiave (fig. 3c). La prima è una ricerca per autore, si desidera l'elenco dei documenti che hanno per autore Knuth D. E. e il sistema ne effettua l'elencazione; la seconda è una ricerca per casa editrice, la Prentice-Hall, e il sistema effettua l'elencazione dei documenti editi dalla casa editrice data.

```

*** READY
PIERCE, = .AND [ TRANSFORMATIONAL GRAMMAR .OR CHOMSKY<THEORY>]
*** SEARCH FOR 'PIERCE,
*** FOUND: PIERCE, J.E.
*** 3 TEXTS FOUNDS
*** SEARCH FOR 'TRANSFORMATIONAL GRAMMAR
*** FOUND: TRANSFORMATIONAL GRAMMAR
*** 39 TEXTS FOUNDS
*** SEARCH FOR 'CHOMSKY<THEORY>
*** FOUND: CHOMSKY<THEORY>
*** 43 TEXTS FOUNDS
*** ONE TEXT FOUND
*** LIST OF TEXT FOUNDS ? <YES/NO> YES
*** TEXT 1 OF 1
00000E03520 X0010145 E,E
PIERCE, J.E.
FINITIVE VS INFINITE STATE GRAMMARS

*** QUERY ALL
00000E03520 X0010145 E,E
PIERCE, J.E.
FINITIVE VS INFINITE STATE GRAMMARS
IN: LINGUISTIOS
THE HAGUE O 1971, 65, 072-074
! TRANSFORMATIONAL GRAMMAR, GENERATIVE GRAMMAR, CHOMSKY<THEORY>, FINITE STATE
GRAMMAR, SENTENCE, DESCRIPTIVE ADEQUACY

```

Fig. 2.

```
==> TIPO DI RICERCA : DOMANDA
==> DOMANDA : DATA-SCIENCES O DATA-STRUCTURES
==> LA SOTTOCLASSE= DATA STRUCTURES
    E' STATA SELEZIONATA CON UN VALORE DI ATTINENZA= 42
==> LA SOTTOCLASSE= 6-INFORMATION RETRIEVAL
    E' STATA SELEZIONATA CON UN VALORE DI ATTINENZA= 34
==> LA SOTTOCLASSE= DATA BASE
    E' STATA SELEZIONATA CON UN VALORE DI ATTINENZA= 11
==> VOLETE LA STAMPA DELLE INFORMAZIONI DEI DOCUMENTI SELEZIONATI? <SI O NO>
? SI
==> SONO STATI SELEZIONATI 17 DOCUMENTI
==> LISTA DEI DOCUMENTI SELEZIONATI :

==> COEFFICIENTE DI ATTINENZA ALPHA= 42
    INFORMAZIONI RELATIVE AL DOCUMENTO:
COLLOCAZIONE: H123 ANNO DI STAMPA: 1968
TIT: THE ART OF COMPUTER PROGRAMMING - VOL.1
AUT: KNUTH,D.E.
CASA ED.: ADDISON-WESLEY

==> COEFFICIENTE DI ATTINENZA ALPHA= 34
    INFORMAZIONI RELATIVE AL DOCUMENTO:
COLLOCAZIONE: U3 ANNO DI STAMPA: 1973
TIT: INFORMATION-RETRIEVAL
AUT: DOUDNIKOFF,B.
CASA ED.: AUERBACH

==> COEFFICIENTE DI ATTINENZA ALPHA= 34
    INFORMAZIONI RELATIVE AL DOCUMENTO:
COLLOCAZIONE: U7 ANNO DI STAMPA: 1973
TIT: INFORMATION SYSTEMS
AUT: VICKERY,B.C.
CASA ED.: BUTTERWORTHS

==> COEFFICIENTE DI ATTINENZA ALPHA= 31
    INFORMAZIONI RELATIVE AL DOCUMENTO:
COLLOCAZIONE: T10 ANNO DI STAMPA: 1975
TIT: DATA STRUCTURES
AUT: BERZTISS,A.T.
CASA ED.: ACADEMIC

==> COEFFICIENTE DI ATTINENZA ALPHA= 31
    INFORMAZIONI RELATIVE AL DOCUMENTO:
COLLOCAZIONE: U15 ANNO DI STAMPA: 1975
TIT: DATA BASE DESCRIPTION
AUT: AA,VV.
CASA ED.: NORTH-HOLLAND

==> VOLETE CHE LA STAMPA DELLE INFORMAZIONI CONTINUI? <SI O NO>
? NO
```

Fig. 3 a.

```

==> VOLETE PORRE UNA DOMANDA CONTINUAZIONE?(SI O NO)
? SI

==> CONTINUAZIONE : NON DATA-BASE

==> LA SOTTOCLASSE= G-INFORMATION RETRIEVAL
E' STATA SELEZIONATA CON UN VALORE DI ATTINENZA= 43

==> LA SOTTOCLASSE= DATA STRUCTURES
E' STATA SELEZIONATA CON UN VALORE DI ATTINENZA= 41

==> VOLETE LA STAMPA DELLE INFORMAZIONI DEI DOCUMENTI SELEZIONATI?(SI O NO)
? SI

==> SONO STATI SELEZIONATI 16 DOCUMENTI

==> LISTA DEI DOCUMENTI SELEZIONATI :

==> COEFFICIENTE DI ATTINENZA ALPHA= 43
INFORMAZIONI RELATIVE AL DOCUMENTO:
COLLOCAZIONE: U7 ANNO DI STAMPA: 1973
TIT: INFORMATION SYSTEMS
AUT: VICKERY,B.C.
CASA ED.: BUTTERMORTHS

==> COEFFICIENTE DI ATTINENZA ALPHA= 43
INFORMAZIONI RELATIVE AL DOCUMENTO:
COLLOCAZIONE: U3 ANNO DI STAMPA: 1973
TIT: INFORMATION-RETRIEVAL
AUT: DOUDNIKOFF,B.
CASA ED.: AUERBACH

==> COEFFICIENTE DI ATTINENZA ALPHA= 41
INFORMAZIONI RELATIVE AL DOCUMENTO:
COLLOCAZIONE: H123 ANNO DI STAMPA: 1968
TIT: THE ART OF COMPUTER PROGRAMMING - VOL.1
AUT: KNUTH,D.E.
CASA ED.: ADDISON-WESLEY

==> COEFFICIENTE DI ATTINENZA ALPHA= 31
INFORMAZIONI RELATIVE AL DOCUMENTO:
COLLOCAZIONE: U10 ANNO DI STAMPA: 1968
TIT: INFORMATION RETRIEVAL SYSTEMS
AUT: LANCASTER,F.W.
CASA ED.: WILEY

==> COEFFICIENTE DI ATTINENZA ALPHA= 30
INFORMAZIONI RELATIVE AL DOCUMENTO:
COLLOCAZIONE: U15 ANNO DI STAMPA: 1975
TIT: DATA BASE DESCRIPTION
AUT: AA.VV.
CASA ED.: NORTH-HOLLAND

==> VOLETE CHE LA STAMPA DELLE INFORMAZIONI CONTINUI?(SI O NO)
? NO

==> VOLETE PORRE UNA DOMANDA CONTINUAZIONE?(SI O NO)
? NO

```

Fig. 3 b.

```
==> CI SONO ALTRE RICHIESTE?(SI O NO)
? SI

==> TIPO DI RICERCA : AUTORE

==> AUTORE : KNUTH,D.E.
COLLOCAZIONE: R55      ANNO DI STAMPA: 1974
TIT: STRUCTURED PROGRAMMING WITH GOTO STATEMENTS
AUT: KNUTH,D.E.
CASA ED.: STANFORD UN.
COLLOCAZIONE: H123    ANNO DI STAMPA: 1968
TIT: THE ART OF COMPUTER PROGRAMMING - VOL.1
AUT: KNUTH,D.E.
CASA ED.: ADDISON-WESLEY
COLLOCAZIONE: H123    ANNO DI STAMPA: 1973
TIT: THE ART OF COMPUTER PROGRAMMING - VOL.3
AUT: KNUTH,D.E.
CASA ED.: ADDISON-WESLEY

==> CI SONO ALTRE RICHIESTE?(SI O NO)
? SI

==> TIPO DI RICERCA : CASAEDITRICE

==> CASA EDITRICE : PRENTICE-HALL
COLLOCAZIONE: U9      ANNO DI STAMPA: 1975
TIT: DYNAMIC INFORMATION AND LIBRARY PROCESSING
AUT: SALTON,G.
CASA ED.: PRENTICE-HALL
COLLOCAZIONE: U4      ANNO DI STAMPA: 1971
TIT: THE SMART RETRIEVAL SYSTEM
AUT: SALTON,G.
CASA ED.: PRENTICE-HALL
COLLOCAZIONE: U13     ANNO DI STAMPA: 1975
TIT: COMPUTER DATA-BASE ORGANIZATION
AUT: MARTIN,J.
CASA ED.: PRENTICE-HALL

==> CI SONO ALTRE RICHIESTE?(SI O NO)
? NO
STOP
R
```

Fig. 3 c.

BIBLIOGRAFIA

- [1] MICHAJLOV A. I., CERNYJ A. I., GILJAREVSKIJ R. S. - *Principi di informatica*. Editori Riuniti, Roma, 1973.
- [2] DA RIVA D., RICCI G. - *Elementi di organizzazione delle informazioni*. Franco Angeli Editore, Milano, 1972.
- [3] LANCASTER F. W. - *Information Retrieval Systems*. John Wiley, 1968.
- [4] *Categories of the Computing Sciences, Revised Classification System for Computing Reviews*, CR of ACM vol. 17, n. 5, maggio 1976.
- [5] MAURER W. D. - *An Improved Hash Code for Scatter Storage*. Communications of ACM, vol. 11, pagg. 35-38, gennaio 1968.
- [6] BELL J. R. - *The Quadratic Quotient Method: A Hash Code Eliminating Secondary Clustering*. Communications of ACM, vol. 13, pagg. 107-109, febbraio 1970.
- [7] TREUSCH B., LESTUZZI F., ROVA S. - *Information Storage and retrieval. Modello di un sistema integrato*. W. P. Fondazione Dalle Molle, Venezia.
- [8] LESTUZZI F. - *SCRIN: Un sistema colloquiale di ritrovamento dell'informazione*. Congresso AICA 1975, Genova.
- [9] LESTUZZI F. - *SCRIN: System Reference Manual*. Rapporto interno Centro Studi della Barbariga, Stra (Venezia).
- [10] AGOSTI M., KING M., LESTUZZI F., WETTLER - *Ulteriori sviluppi dei due sistemi di documentazione automatica SCRIN e DOC4*. Congresso AICA 1977, Pisa.

