



AICA' 77

L77-5

RICUPERO DELL'INFORMAZIONE BIBLIOGRAFICA: UN'INTERFACCIA
UTENTE-SISTEMA PER LA DEFINIZIONE DELLE STRATEGIE DI RICERCA

M. B. Baldacci, R. Sprugnoli
Istituto di Elaborazione dell'Informazione
del C.N.R., Pisa

Tra i fattori che influenzano l'efficienza di un sistema per il ricupero dell'informazione bibliografica riveste importanza preminente i) il modo in cui sono descritti i documenti dal punto di vista semantico e ii) il modo in cui sono formulate e trattate le domande sottoposte al sistema. Sono infatti queste due componenti "esterne" al sistema, più che il sistema stesso, a determinare il grado di precisione e di completezza delle risposte [1].

Mentre è difficile influire sui fattori umani che condizionano i) in modo quasi esclusivo, è possibile approntare alcuni strumenti che rendano più efficiente la componente ii). Trascurando volutamente i sistemi che si avvalgono di vocabolari controllati (o thesaurus) per la descrizione e la ricerca dei documenti (in questi sistemi è in genere giudicato insoddisfacente il rapporto fra costo di costruzione e aggiornamento del vocabolario e grado di specificità della ricerca che esso permette), possiamo citare, come esempio di strumenti largamente usati nei sistemi con vocabolario libero, la tecnica di troncamento delle parole e l'uso di connettivi per aumentare rispettivamente completezza e precisione delle risposte. Tuttavia, per l'utente di tali sistemi, e in modo particolare per l'utente occasionale, l'impossibilità di avere dal sistema informazioni che lo guidino nella definizione della strategia di ricerca costituisce frequentemente una grossa barriera, tale da scoraggiare la prosecuzione delle ricerche nel caso di tentativi infruttuosi: molti sistemi infatti non offrono molto più che la possibilità di interrogare il dizionario per conoscere se un termine è o non è in esso compreso.

Una strada per arrivare ad una organizzazione dei termini di descrizione tale che possa validamente sostituire il thesaurus senza avere alcuno dei costi da questo imposti discende da un modo di descrivere i documenti, che è sempre più diffuso, ma che non sembra aver trovato fino ad oggi corrispondenti applicazioni pratiche: esso consiste nell'associare al documento uno o più descrittori (o codici) appartenenti ad un vocabolario controllato di dimensioni modeste (quale ad es. lo schema di classificazione dell'ACM) e una serie di termini estratti liberamente dal testo. Questa descrizione "a due livelli" - che risulta uno strumento potente ed agevole perchè permette di raggiungere, attraverso i termini "liberi" e il vocabolario controllato un grado elevato di specificità in un contesto ben definito - pensiamo possa costituire la base per stabilire in modo "automatico" relazioni fra termini semanticamente correlati.

Il progetto attualmente in corso tende a verificare sperimentalmente la validità di queste assunzioni dotando il sistema RESP [2] di una interfaccia con l'utente che permetta a questi di chiedere al sistema:

1. quali sono i campi di interesse della base di dati; il sistema risponde mostrando lo schema di classificazione ai vari livelli di dettaglio, secondo le indicazioni dell'utente;
 2. se un termine è presente nel dizionario e quante sono le sue occorrenze;
 3. la lista dei termini associati a un determinato codice e le loro frequenze (questa lista può essere presentata sia in ordine alfabetico sia in ordine di frequenza);
 4. la lista dei codici associati con un dato termine e relative frequenze;
 5. la lista dei termini associati con un dato termine e relative frequenze;
 6. il numero delle associazioni fra più termini;
- dove sono chiamati "associati" codici e/o termini attribuiti allo stesso documento. In tal modo si crea, pur senza il supporto di un thesaurus, quella funzione di sistema chiamata "browsing" che, nel caso del sistema RESP, dà all'utente la possibilità di
- a. orientarsi sul contenuto della base di dati (passo 1);
 - b. sapere se un termine è presente nel dizionario (passo 2);
 - c. trovare eventuali sinonimi (oppure termini in altre lingue) in caso di risposta negativa al punto b), e di valutare il numero di documenti da essi puntati (passo 3);
 - d. di espandere la ricerca dei termini correlati attraverso il gioco alterno dei passi 3, 4, 5;
 - e. di valutare preventivamente il numero dei documenti di possibile interesse (passo 6).

RIFERIMENTI

- [1] B. C. VICKERY: Information Systems. London, Butterworths, 1973; pp. 123-128.
- [2] R. SPRIGNOLI: "RESP: A Programmable Retrieval and Data Base Management System". Nota Interna IEI, in preparazione.