

Indicazioni per l'esecuzione di ricerche bibliografiche in campo bio-medico tramite un servizio di analisi dell'informazione collegato ad un Centro MEDLARS

MARCELLA MAGLIOLA (*)

Servizio Ricerche Bibliografiche

Convenzione tra l'I.S.S. e l'I.N.S.E.R.M. in materia di informazione biomedica

Nel febbraio di quest'anno sono state poste le basi per una convenzione tra l'Istituto Superiore di Sanità di Roma e l'Institut National de la Santé et de la Recherche Médicale (I.N.S.E.R.M.) di Parigi in materia di informazione in campo medico-biologico. In virtù di tale accordo l'Istituto Superiore di Sanità affida al centro di documentazione I.M.A. (Information Médicale Automatisée) dell'I.N.S.E.R.M. l'esecuzione di ricerche bibliografiche che saranno formulate in linguaggio MEDLARS. Con questo, l'Istituto assume responsabilità e funzioni definite nel quadro dell'informazione scientifica nazionale, ed istituisce un ufficio centrale, o Servizio di Ricerche Bibliografiche, per rispondere, attraverso una tecnologia avanzata, ai bisogni documentari specifici del complesso della comunità biomedica (ricerca - clinica - università - industria).

La convenzione, già operante, mira a dare un contributo concreto ed immediato alla soluzione di un problema che ha caratteri di urgenza e di universalità particolarmente acuti. A livello nazionale, i fondi documentari esistenti sono spesso inadeguati o non convenientemente sfruttati, principalmente per una mancanza quasi generale di educazione a questo tipo di lavoro; l'espansione delle conoscenze ed il trasferimento interdisciplinare dell'informazione non consentono iniziative individuali di ampio respiro. L'ufficio centrale, con funzioni di informazione, orientamento ed elabora-

(*) Indirizzo: Istituto Superiore di Sanità, Viale Regina Elena 299 - 00161 Roma - Telefono 4998

zione di richieste di ricerche bibliografiche, facilita l'accesso rapido a fondi documentari normalizzati e, soprattutto, consente di utilizzare nelle ricerche bibliografiche un sistema automatico capace di far fronte al crescente volume della letteratura. L'operatività in scala nazionale porta quasi di necessità all'impiego del MEDLARS, sistema che, pur specifico del campo biomedico, è caratterizzato da una copertura ampiamente multidisciplinare ed una diffusione mondiale.

Creato nel 1964 negli Stati Uniti (Bethesda) dalla National Library of Medicine, il MEDLARS (MEDical Literature Analysis Retrieval System) indicizza, a vario grado di profondità, 2800 giornali biomedici pubblicati in tutto il mondo e, arricchendosi ogni anno di circa 250.000 nuove analisi di articoli, opera oggi su circa 1.500.000 registrazioni. Su queste l'elaboratore esegue ricerche retrospettive che selezionano nella massa del materiale depositato in memoria i documenti che rispondono a questioni specifiche presentate dai singoli utenti. Le stesse informazioni possono assumere la forma di aggiornamento sistematico, per lo più mensile, previa indicazione da parte dell'utilizzatore di un suo profilo di interesse.

Un lavoro di tali dimensioni è reso possibile dalla flessibilità dell'organizzazione del sistema, basata sulla ripartizione dei compiti e la partecipazione tecnica di alcune filiali o centri regionali. Il servizio I.M.A. - I.N.S. E.R.M., con il quale è stato stipulato l'accordo, è il più recente (1969) tra i centri MEDLARS europei e va ad aggiungersi a quello inglese istituito presso la National Lending Library di Boston Spa ed a quello svedese presso il Karolinska Institutet di Stoccolma.

Natura e funzioni del Servizio di Ricerche Bibliografiche

In questa relazione non s'intende presentare una descrizione in dettaglio o critica di un sistema, il MEDLARS, che è già pienamente operativo da oltre 8 anni ed ha incontrato consensi e, quindi, diffusione praticamente in ogni parte del mondo. Né si vogliono illustrare, sotto l'aspetto dell'indicizzazione della letteratura nazionale richiesta, le relazioni tra l'istituzione centrale di Bethesda ed un centro della rete di diffusione del tipo di quello francese dell'I.N.S.E.R.M. Tutto questo è stato già fatto e più autorevolmente^(1, 2) e comunque esula dai nostri propositi. Scopo di questa relazione è richiamare l'attenzione della popolazione biomedica italiana sulla possibilità di utilizzare il MEDLARS a livello nazionale, attraverso la mediazione di un organismo statale di attività scientifica, l'Istituto Superiore di Sanità, che, affiancato dall'Istituto Farmacologico Mario Negri di Milano, si fa promotore dell'iniziativa.

L'Istituto Superiore di Sanità, sensibilizzato ai problemi della documentazione dalle esigenze dei suoi propri ricercatori, ha ritenuto suo compito

specifico, nel quadro delle sue molteplici attività in campo sanitario, sociale ed educativo, studiare la possibilità di offrire uno strumento potenzialmente in grado di servire tutti i settori della ricerca biomedica pubblica e privata, compresi quelli delle industrie con interessi di ricerca biologica, quali le industrie farmaceutiche, alimentari, ecc. Per questo fine mette a disposizione, oltre alla sua biblioteca tra le più ricche di opere e periodici nel campo, anche tecnici qualificati con lunga esperienza in fatto di ricerche bibliografiche, cui ha affidato l'incarico di realizzare nel modo più semplice e proficuo la saldatura tra l'utenza ed il sistema.

L'ufficio centrale, o Servizio di Ricerche Bibliografiche istituito presso l'Istituto Superiore di Sanità, vuole essere programmaticamente un organismo informale ed aperto ad ogni collaborazione in materia di documentazione, non semplicemente un centro di raccolta ed inoltra di domande di ricerche bibliografiche. Solo così può esplicare la funzione che si è assunta, quella di intermediario, sufficiente e necessario, tra il richiedente e l'elaboratore. Intermediario sufficiente, in quanto, in base all'accordo, il suo intervento deve essere tale da tradurre, senza distorsioni di contenuto, le richieste di ricerche che gli pervengono in una forma che sia compatibile con la struttura del sistema e da questo accettabile senza ulteriore elaborazione a livello di formulazione. Intermediario necessario, in quanto il suo intervento è essenziale all'elaborazione della richiesta. Non è un compito facile, ma un impegno che i documentalisti si assumono nella piena coscienza di quanto esso comporti di lavoro e di specializzazione. Il documentalista partecipa dello scienziato di cui condivide le conoscenze fondamentali e del tecnico dell'elaborazione elettronica, con cui ha in comune le nozioni di base dell'automazione integrate da una formazione specialistica in relazione ai sistemi documentari automatizzati.

Il materiale oggetto di ricerca da un lato rispecchia l'infinita variabilità del pensiero umano, mentre dall'altro deve obbedire a criteri rigidi. Due ricerche bibliografiche, per quanto relative ai medesimi concetti, riflettono inevitabilmente le situazioni diverse che le hanno determinate. E queste situazioni, più spesso lette tra le righe o dedotte intuitivamente che espresse in modo esplicito dal richiedente, hanno un peso determinante nella formulazione che ne risulta condizionata in modo sostanziale. Ad esempio, è ovvio che una ricerca fatta ai fini della compilazione di una tesi di laurea ha caratteristiche del tutto diverse da quelle della medesima ricerca fatta da un medico che voglia documentarsi per l'esercizio della sua professione, da un tecnico che voglia utilizzarla nell'esecuzione di lavoro sperimentale, da un professore che voglia servirsene per compilare un testo didattico. Questo significa che non esistono due richieste assolutamente identiche nella sostanza e, di conseguenza, due formulazioni perfettamente sovrapponibili.

Questa mancanza di schemi standard in cui incasellare le singole richieste non deve essere considerata una deficienza del sistema, ma, al contrario, uno dei suoi maggiori pregi. La flessibilità che gli consente di inquadrare la richiesta nella situazione e specificazione professionale dell'individuo che la presenta è di per sé garanzia della validità dei risultati. Quando la lista dei riferimenti bibliografici non risponde alle aspettative non è quasi mai il sistema che ha mancato, ma il richiedente che non è riuscito ad esprimere quello che voleva o l'intermediario che non ha saputo interpretarlo e tradurlo. Molto spesso l'analisi obiettiva dell'insuccesso porta all'esposizione iniziale del problema, o troppo vaga, o troppo ristretta, o persino reticente (non tutti sanno che esiste in materia un segreto professionale rigorosamente rispettato). Di qui spesso l'opportunità di un dialogo chiarificatore tra il tecnico della ricerca bibliografica e l'utente, meglio se in un secondo momento, quando il richiedente abbia già meditato il problema e l'abbia formulato narrativamente, nel linguaggio a lui più naturale. (Tentare di conformare dall'inizio la questione alle limitazioni ed alla logica del sistema significa darne un'immagine viziata che spesso tradisce le esigenze che la hanno dettata). Il compito dell'utente si esaurisce nella chiarezza dell'esposizione della questione: la formulazione di una opportuna strategia, la scelta della terminologia più confacente, lo studio degli articoli indicati come pertinenti, la definizione delle categorie generali, sono tutte operazioni che spettano all'intermediario. Questo non facile lavoro di traduzione del linguaggio naturale in linguaggio del sistema ha il suo peso sulla validità dei risultati, che restano però sempre condizionati in larga misura dall'esposizione iniziale. Un dubbio nella presentazione del quesito comporta una formulazione imprecisa e sfocia in una ricerca insoddisfacente, se non del tutto inaccettabile. Nel sistema non c'è posto per incertezze, ma solo per scelte univoche e ben precise. L'apparente possibilità di risolvere interpretazioni alternative ampliando la richiesta fino a comprendere tutte le potenziali soluzioni risulta all'atto pratico in una proporzione intollerabile di pubblicazioni non pertinenti.

L'insuccesso dovuto ad una formulazione errata non implica necessariamente la proposizione inversa e cioè che, presentata la richiesta in modo corretto, il risultato sarà soddisfacente al 100% o, in altri termini, le pubblicazioni reperite saranno tutte quelle pertinenti e solo quelle. Un siffatto risultato ideale non è raggiungibile. L'analisi statistica eseguita su campioni significativi di risposte indica che il MEDLARS opera in media al 58% di recupero e 50% di precisione. In pratica, più alto è il recupero, più bassa la precisione e viceversa. Entro determinati limiti è possibile operare sulla strategia della ricerca in modo da ottenere un maggior numero di citazioni bibliografiche e quindi una minore perdita di documenti rilevanti, ammettendo però in partenza un numero relativamente grande di pubblicazioni

scarsamente pertinenti. Inversamente, si può restringere la ricerca per aumentarne la precisione, correndo il rischio di perdere materiale di qualche valore. È una questione di scelta preliminare di cui ancora una volta è arbitro il richiedente, ma strumento il bibliografo specialista.

Nella sua forma più schematica, la ricerca documentaria può essere riassunta nelle seguenti operazioni :

- analisi concettuale della questione;
- messa in equazione;
- traduzione nel linguaggio documentario;
- ricerca propriamente detta;
- edizione o stampa delle indicazioni bibliografiche.

Molti di questi passaggi sottintendono la partecipazione attiva dell'intermediario della ricerca, o documentalista, o bibliografo.

Del suo ruolo nell'analisi concettuale della domanda è stato già detto e con questo, implicitamente, anche della sua funzione nell'organizzazione dei differenti concetti in equazione di ricerca secondo la logica dell'algebra di Boole. Quello che invece non risulta immediatamente è l'importanza di una buona esperienza in campo di ricerche bibliografiche quando si debbano stabilire quali termini, o categorie di termini, esprimano più o meno fedelmente il concetto.

È questo il punto in cui l'elaborazione concettuale assume forma concreta, ma anche il punto in cui il sistema diventa rigido ed esigente. I termini ammessi dal MEDLARS sono esclusivamente quelli derivati da un dizionario scientifico normalizzato (Medical Subject Headings o MESH), con aggiornamento periodico che ne assicura la vitalità, ma crea nel contempo complicazioni pratiche. Nel MESH i termini, oltre ad essere riportati in sequenza alfabetica integrata da rinvii e rimandi, sono riuniti in categorie il cui impiego globale o, come si dice negli ambienti MEDLARS, la cui « esplosione » risparmia lunghe liste di vocaboli in qualche modo raggruppabili, e, quindi, prezioso tempo di impiego dell'elaboratore, ma può portare anche ad errori sostanziali, sia per eccesso che per difetto. La scelta dei termini non è però unicamente una questione di vocabolario. Più importante a riguardo è la ricerca e, possibilmente, la verifica di una presumibile coincidenza tra i termini suggeriti e quelli con cui quel concetto è stato caratterizzato all'atto dell'indicizzazione dei lavori che lo contengono e nella successiva registrazione in memoria. Non v'è dubbio che a questo livello il bibliografo abbia un ruolo determinante, soprattutto quando il concetto non ha espressione diretta ed inequivoca in termini accettabili o quando, al contrario, ammette più espressioni le cui differenze, non sempre sottili, possono non essere colte

da chi non abbia sufficiente esperienza in merito. Il processo di ricerca propriamente detto, o di paragone tra la formulazione della ricerca e le parole-chiave descrittive dell'articolo registrate in memoria, non è di stretta pertinenza dell'intermediario bibliografo, mentre nella definizione dei particolari di edizione un suo suggerimento può essere di una certa utilità. Così, ad esempio, la scelta della lingua di pubblicazione dei lavori non sembra dare adito a dubbi: si indicano solo le lingue conosciute in misura sufficiente a consentire almeno per sommi capi la lettura di una pubblicazione scientifica. Eppure, se risulta che pubblicazioni di merito sono scritte in lingua non conosciuta, non è consigliabile escluderla a priori, poichè dalla lista delle parole chiave che seguono l'indicazione bibliografica, in inglese, si possono derivare indicazioni interessanti.

Anche a livello della lettura dei risultati la collaborazione dell'esperto è utile, almeno la prima volta che l'utente si accosta al sistema. Le indicazioni bibliografiche, complete di ogni particolare, sono di tipo standard e come tali non richiedono delucidazioni. Alla stessa stregua è ovvio il significato della lista dei termini di indice che, distinti in primari e secondari, seguono la citazione bibliografica di cui costituiscono in sostanza un abstract. Di spiegazione ed interpretazione ha invece bisogno la sequenza delle diverse strategie entro la medesima ricerca, sia nei casi in cui le sub-ricerche sottintendono livelli di specificità progressivamente crescente che in quelli in cui indicano differenti angolazioni dello stesso problema.

Prima ancora di collaborare alla compilazione della richiesta di ricerca bibliografica, il bibliografo può rendersi utile — sembrerebbe un paradosso — proprio quando il sistema, il MEDLARS, non è utile o lo è meno di altri sistemi. Il MEDLARS è una delle più autorevoli bibliografie secondarie in campo biomedico, ma non è l'unica e, in determinati casi, nemmeno la migliore. A prescindere dall'evidente limitazione rappresentata dall'anno di inizio, non è escluso che lavori di un certo rilievo possano essere pubblicati su periodici non indicizzati dal MEDLARS e quindi non reperibili né ad una ricerca manuale sull'Index Medicus né attraverso elaborazione elettronica. Inoltre, la limitata definizione chimica dei composti biologicamente attivi considerati nel MESH obbliga ad impiegare denominazioni molto ampie, talvolta quasi generiche, per cui la ricerca risulta in una massa di citazioni relative ad individui chimici che con quello che interessa hanno legami molto tenui. In altri casi, come quando si voglia conoscere la produzione scientifica di una determinata istituzione o ricostruire, attraverso i suoi successivi sviluppi, la storia di una teoria, l'esclusione del MEDLARS è addirittura tassativa. Altri sistemi, quali il Chemical Abstracts Service,

l'Excerpta Medica, i servizi dell'I.S.I. (Citation Index, ASCA IV), offrono possibilità e presentano limitazioni che ne caratterizzano l'applicazione, permettendo, con una scelta oculata, di dare risposta praticamente ad ogni tipo di ricerca bibliografica. La scelta del sistema più rispondente presuppone conoscenza ed esperienza, o, in altri termini, quelle cognizioni che concorrono a formare la specializzazione del bibliografo. È anche questa specializzazione che il Servizio di Ricerche Bibliografiche mette a disposizione della comunità perchè nell'interesse comune integri altre specializzazioni.

Efficienza operativa del MEDLARS : cause di errori a livello dei diversi subsistemi che lo compongono.

Il sistema opera, in media, a circa il 58% di recupero (recall) ed il 50% di precisione (*) e, su questa base di precisione, fornisce il 65% della letteratura di maggior valore. Riportando nel diagramma precisione/recupero i valori medi registrati nei diversi centri MEDLARS ed estrapolando, si può costruire una curva il cui significato più interessante è che in sostanza il rapporto medio non è casuale, ma, coscientemente o incoscientemente, il risultato di una politica di compromesso tra il recupero e la precisione. L'espressione « in media » non vuole qui significare che i valori si distribuiscono con andamento gaussiano intorno alle percentuali indicate, poichè poche delle singole ricerche cadono nell'area dei rapporti medi $\pm 5\%$. Alcune ricerche risultano particolarmente buone, con alto recupero accompagnato da alta precisione, altre, invece, danno risultati assolutamente insoddisfacenti. L'esame accurato di campioni di ricerche alle due estremità della curva di distribuzione ha permesso di identificare i più importanti fattori che governano il successo o l'insuccesso di una ricerca.

Ai fini di un'analisi della funzionalità globale del MEDLARS, sotto l'aspetto delle cause di errore che su di esso più influiscono, il sistema è stato studiato separatamente nei suoi quattro componenti principali: struttura o linguaggio del sistema, processo di indicizzazione, interazione utente-sistema e processo di formulazione della ricerca. Una valutazione critica (3)

(*) Recupero = rapporto tra il numero dei documenti pertinenti in qualche modo noti ed il numero di documenti pertinenti reperiti dal sistema.

Precisione = rapporto tra il numero dei documenti pertinenti ed il numero totale di documenti reperiti.

indica come ciascuno dei sub-sistemi abbia una sua propria responsabilità in rapporto all'efficienza totale, sia in fatto di capacità di recupero che di precisione :

Cause di errore	Percentuale del totale degli errori di recupero	Percentuale del totale degli errori di precisione
linguaggio del sistema	10,2	36,0
indicizzazione	37,4	12,9
interazione utente-sistema	25,0	16,6
formulazione ricerca	35,0	32,3
elaboratore elettronico	1,4	0,1
altre	—	2,4

(La somma delle singole percentuali riflette il fatto che il medesimo errore può essere l'espressione di più cause che sono state analizzate e considerate singolarmente).

Numericamente, gli errori propri della struttura del sistema e della fase di preparazione dello stock di informazioni su cui operare la selezione (definizione del linguaggio di indice e indicizzazione) equivalgono a quelli della fase di utilizzazione del sistema (interazione utente-sistema e formulazione della ricerca). Dei primi, basti dire che sono oggetto di studio costante e che in parte, in verità piuttosto limitata, potranno essere eliminati. I secondi invece, sono di più stretta pertinenza dell'utenza ed in misura molto maggiore suscettibili di essere evitati.

Della necessità, più che opportunità, di una adeguata interazione utente-sistema è stato già detto. Quello che forse è meno chiaro è il significato di quel 25% degli errori di recupero e 16,6% degli errori di precisione attribuiti ad essa in tabella. Un errore di recupero dovuto ad una interazione difettosa implica che la richiesta, come è stata formulata, è molto più specifica dell'area di informazione effettivamente desiderata. Allo stesso modo, un errore di precisione indica che la formulazione è molto più generale dell'informazione voluta. Se poi per una inadeguata interazione la sovrapposizione è solo parziale, si possono avere errori sia di recupero che di precisione. L'errata interazione è essenzialmente conseguenza di due situazioni. L'una, ben nota ai bibliotecari che la vivono nella loro esperienza di ogni giorno, si verifica quando l'utente presenta una richiesta che non rispecchia quello di cui ha realmente bisogno o, alternativamente, quando le sue esigenze non sono ben comprese da chi formula la richiesta. Nell'altra situazione, l'utente presenta una richiesta che riflette sufficientemente l'informazione voluta, ma non è a conoscenza di quali tipi di articoli esistano sull'argomento

e potrebbero essere utili. (Questo tipo di interazione difettosa può essere risolto da richieste iterative o, più semplicemente, da uno spoglio preliminare dell'Index Medicus).

Essendo responsabile del 35% e 32% degli errori rispettivamente di recupero e precisione, il processo di formulazione della ricerca è la fase che più contribuisce a limitare la funzionalità del sistema. Lo studio analitico menzionato individua ad un livello più particolareggiato le singole cause che determinano tali percentuali. In fatto di recupero si ha :

ricerca che non considera tutti i possibili approcci al reperimento	21,5%
formulazione troppo esaustiva (*)	8,4%
formulazione troppo specifica	2,5%
« selective printout »	1,6%
uso di termini « ponderati » (weighted) (**)	0,2%
altri errori	0,8%

In fatto di precisione, la stessa analisi dà :

formulazione non specifica	15,2%
formulazione non esaustiva	11,7%
uso di termini, o combinazioni di termini, non appro- priati	4,3%
errori nella logica della strategia	1,1%

Considerando insieme gli errori di recupero e di precisione, nella fase di formulazione della ricerca si possono distinguere tre tipi fondamentali di errori :

- 1) errori nel recupero dovuti al fatto che la formulazione non copre tutti gli approcci ragionevoli per il reperimento di articoli rilevanti ;
- 2) errori puri derivati dall'impiego di termini inappropriati o da un'errata logica nella strategia della ricerca;
- 3) errori dovuti ai livelli di specificità e/o esaustività adottati nella strategia della ricerca.

Il primo tipo di errore, che rappresenta la seconda maggiore causa di mancato recupero, implica che il 21,5% degli articoli rilevanti non sono stati ritrovati perchè non sono stati usati termini, o combinazioni di termini,

(*) Una ricerca si dice « esaustiva » quando la sua formulazione comporta la coesistenza di tutte le categorie di concetti, impliciti o espliciti nel quesito, o di tutte le « faccette » nel senso di Ranganathan.

(**) Termini sotto cui la pubblicazione è reperibile nell'Index Medicus o rispecchianti esclusivamente gli aspetti principali della pubblicazione stessa.

sotto cui si sarebbe potuto sortire materiale di un certo valore. Ovviamente, l'errore ha un peso diverso a seconda che consista nel mancato impiego di un termine o combinazione di termini in una formulazione per il resto completa o consista invece nell'aver trascurato tutto un aspetto della richiesta. Il primo ha conseguenze meno drastiche del secondo, ma può essere anche sostanziale.

Il secondo tipo di errore, causato dall'impiego di termini o combinazioni di termini inappropriati, interessa la precisione e contribuisce per il 4,3% alla deficienza di precisione del sistema globale. La natura di questi errori è facilmente comprensibile, ma merita forse sottolineare come l'inappropriata combinazione di termini tenda a verificarsi in formulazioni piuttosto complesse in cui una lista di termini in somma logica (legati tra loro dalla relazione « o ») è unita in prodotto logico (relazione « e ») ad una seconda lista di termini in somma logica. In questi casi, mentre la strategia generale è valida, alcune delle combinazioni risultanti possono essere irrazionali in relazione alla richiesta.

Il terzo tipo di errore è il più difficile a controllare in quanto relativo allo spostamento in un senso o nell'altro di un equilibrio basato più sull'intuizione che su quantità definite. L'estensione della esaustività e/o specificità della formulazione è parte essenziale e problema centrale della strategia della ricerca. Quanto meno specifica ed esaustiva è una formulazione, tanto maggiore è il numero dei documenti reperiti: il recupero tende ad aumentare al diminuire della precisione. Quanto più specifica ed esaustiva è una formulazione, tanto minore il numero dei documenti reperiti: col migliorare della precisione il recupero tende ad abbassarsi. Una formulazione esaustiva richiede la coesistenza di tutte le nozioni richieste, in una determinata relazione, benchè non necessariamente al livello di specificità stabilito dal richiedente. L'esaustività è correlata al livello di coordinazione, cioè al numero dei termini che debbono coesistere perchè un articolo possa essere reperito, ma non è identificabile con questo poichè tra i due il rapporto non è rigorosamente 1:1. In sostanza, formulazioni troppo esaustive sono responsabili dell'8,4% degli errori di recupero, così come quelle non esaustive dell'11,7% degli errori di precisione. In rapporto alla specificità, il 2,5% degli errori di recupero sono dovuti ad una formulazione eccessivamente specifica ed il 15,2% degli errori di precisione ed una formulazione non specifica. La riduzione di specificità, se comporta generalizzazioni, tende a migliorare il recupero a spese della precisione. Non è possibile indicare criteri di validità generale in materia di strategia della ricerca e soprattutto in relazione al giusto livello di generalità da adottare. La scelta della politica da seguire va decisa caso per caso e, al di fuori di alcune situazioni limite, è affidata esclusivamente all'utente che, nel riempire il formulario, esprime preferenze che segnano direttive.

Ricerca bibliografica automatica versus ricerca bibliografica manuale

Una pubblicazione scientifica è patrimonio comune della società che dall'insieme dei singoli contributi trae incentivazione a nuovi progressi. Tuttavia, proprio questa universalità dell'apporto alla conoscenza crea problemi di gestione dell'informazione: il dato, la teoria, l'osservazione hanno significato unicamente se reperibili nella massa delle pubblicazioni al momento in cui possono trovare utilizzazione. Di qui la necessità di una organizzazione razionale della ricerca bibliografica. Ricercare l'indicazione di qualcosa che è stato pubblicato non significa eseguire, manualmente o meccanicamente, lo spoglio di uno schedario di dimensioni più o meno grandi. Significa studiare in una certa profondità il suo contenuto concettuale in modo da cogliere gli elementi essenziali che la caratterizzano e ne permettono l'individuazione. Sotto questo profilo, la distinzione tra ricerca manuale o tradizionale e ricerca meccanizzata o automatica diventa meno netta: l'elaborazione intellettuale che ne è alla base identica nei due casi e la differenza riguarda unicamente le modalità tecniche con cui le indicazioni sono sortite. Con questo perde il significato anche il preteso antagonismo uomo-macchina nella soluzione di problemi documentari complessi. Al di fuori di entusiasmi eccessivi o di scetticismi ingiustificati, non va dimenticato che l'elaboratore si assume una parte relativamente modesta dei compiti di carattere intellettuale richiesti da un'analisi del contenuto. In essenza, la funzione dell'elaboratore consiste nel ricevere dall'esterno segnali che rappresentano i dati del problema da risolvere, sottoporli ad un trattamento secondo un programma prestabilito ed inviare all'esterno i segnali che rappresentano i risultati del trattamento. Queste operazioni avvengono senza l'intervento dell'uomo, che però conserva un ruolo essenziale nella realizzazione dell'elaborazione. È di pertinenza umana pensare e definire la questione (una macchina non pone di per sé problemi da risolvere); scegliere dati e valutarne il significato; creare programmi che permettano di effettuare il compito nelle migliori condizioni di precisione, di tempo, di costo. Da questo non deriva necessariamente che l'automazione è limitata ai soli lavori di selezione e compilazione, ma parimenti non v'è dubbio che la ricerca bibliografica non ha ancora perso il suo carattere di prodotto critico del pensiero, fatto di logica, ma anche di intuizione.

L'asserire che i diversi tipi di ricerca documentaria hanno un denominatore comune nell'analisi concettuale non implica che siano tra loro interscambiabili ed equivalenti. Ciascun tipo ha le sue proprie caratteristiche che, a seconda delle circostanze, ne rendono consigliabile o sconsigliabile l'adozione. Per fissare le idee, si fa riferimento al caso particolare della bibliografia secondaria ad interesse prevalentemente medico, elaborata dalla National Library of Medicine nella duplice veste di fascicoli mensili stampati,

o Index Medicus, e ricerca su richiesta mediante elaboratore, o MEDLARS, tenendo però presente che le considerazioni hanno carattere assolutamente generale e, con qualche piccola modifica, valgono per tutti i sistemi similari. Nel caso particolare dell'esempio in questione, occorre premettere che allo atto dell'immissione di qualsiasi pubblicazione nel sistema, o, più precisamente, in fase di indicizzazione, la pubblicazione stessa è caratterizzata da un certo numero di termini di indice o descrittivi (con una media di 10 e più negli articoli dei giornali indicizzati in profondità) che ne riassumono il contenuto di informazione. Di questi termini di indice, quelli che esprimono gli aspetti fondamentali del lavoro (in genere 1-3) vanno a costituire le voci sotto cui il lavoro sarà riportato nell'Index Medicus, mentre i termini rimanenti, relativi ad aspetti secondari, sono registrati in memoria e potranno costituire oggetto di ricerca per elaborazione elettronica dei dati. Premesso questo, è ovvio che se l'argomento di cui si vuol conoscere la bibliografia può essere caratterizzato in modo esauriente da uno o due termini, sarà indubbiamente conveniente la ricerca manuale sotto quei termini nell'Index Medicus nella cumolazione annuale o fascicoli mensili. Analogamente, la ricerca per autore, tranne nei casi in cui l'autore abbia dimensioni tali da essere di per sé oggetto di studio, non è giustificata né conveniente per elaborazione elettronica. Le ricerche retrospettive anteriori al 1964, anno di inizio del MEDLARS, devono necessariamente essere condotte nell'Index Medicus che, sia pure con denominazioni lievemente diverse, si estende sufficientemente a ritroso nel tempo. Per queste stesse considerazioni, sarà invece opportuna l'elaborazione elettronica quando l'argomento della ricerca può essere espresso solo da un certo numero di termini correlati tra loro. A maggior ragione, la ricerca all'elaboratore si impone senza alternativa quando ci siano motivi per ritenere che la questione che interessa possa costituire in molti lavori solo un aspetto marginale e pertanto non avrebbe possibilità di essere reperita in essi nell'Index Medicus. Ovviamente, questi casi limite sfumano in una molteplicità di situazioni intermedie in cui la scelta dell'uno o dell'altro procedimento può essere suggerita da fattori meno categorici quali, ad esempio, l'urgenza di una risposta o la richiesta di un'indicazione orientativa.

Quello che però sfugge a chi non abbia sufficiente familiarità con il sistema è che la ricerca meccanizzata non esclude categoricamente la ricerca manuale sull'Index Medicus, ma, al contrario, la richiede spesso quale operazione preliminare alla formulazione della richiesta che deve essere sottoposta alla elaborazione elettronica. In sostanza, poichè è ovvio che un lavoro sarà reperito solo se ricercato sotto gli stessi termini sotto i quali è stato indicizzato, il problema della ricerca si connette al problema dell'indicizzazione e trova verifica immediata nella consultazione dell'Index Medicus. L'indicizzazione non è quasi mai un'operazione univoca che ammetta determinati

termini e non altri. Alla difficoltà di individuare e graduare nella loro importanza gli elementi fondamentali di un lavoro si associa quella di tradurre concetti e sfumature di concetti nei termini accettabili dal sistema. Nè, in fondo, sorprende che sia così; l'indicizzazione di un lavoro, allo stesso modo della concretizzazione di un quesito bibliografico nel linguaggio del sistema, è in essenza un lavoro di interpretazione e come tale sottoposto alla soggettività propria dell'intelligenza umana. Ammessa quindi questa variabilità, occorre trovare il modo di ovviare agli inconvenienti che comporta, soprattutto in vista del fatto che il lavoro di selezione dovrà essere poi affidato ad una macchina che per definizione è caratterizzata dalla più assoluta mancanza di fantasia. L'unica soluzione è una verifica preliminare per consultazione diretta dell'Index Medicus, sia ricercando sotto quali termini sia stato indicizzato un lavoro noto ritenuto pertinente, che esaminando quali lavori siano riportati sotto i termini che meglio sembrano esprimere i concetti richiesti. Questa concessione sistematica al dubbio, che si traduce in una approssimazione alla soluzione ottimale per tentativi suffragati dalla ricerca manuale, rappresenta forse il momento più decisivo della formulazione di una richiesta e non può essere totalmente soppressa, nemmeno in nome di una lunga esperienza in fatto di formulazione. La consultazione manuale dell'Index Medicus, inoltre, dà un'indicazione di massima del numero delle referenze che saranno reperite ed in tal modo condiziona in modo sostanziale l'intera strategia della ricerca. È una consultazione che in effetti ha la natura di un dialogo in cui l'utente interroga proponendo una formulazione della richiesta ed il sistema risponde indicando, tramite la sua edizione a stampa, quale e quanto grande potrà essere il risultato conseguente a quella formulazione.

I.N.S.E.R.M. : altre realizzazioni e prospettive future

Lo spoglio dell'Index Medicus preliminare alla formulazione della richiesta di ricerca bibliografica anticipa una delle funzioni essenziali del MEDLINE, servizio derivato dal MEDLARS che, si prevede, l'I.M.A. renderà operativo verso la fine dell'anno. Il MEDLINE è un sistema di documentazione a interrogazione diretta, operante secondo i principi del MEDLARS, ma su uno stock di base più ridotto (1.000 in luogo di 2.800 periodici). Con esso, campioni delle indicazioni bibliografiche conseguenti alle formulazioni si leggono sullo schermo televisivo in tempi reali e, su base conversazionale, si possono introdurre correzioni nella strategia della ricerca fino ad ottenere risultati soddisfacenti. Se poi l'elenco di citazioni bibliografiche che appare sul teleschermo, per quanto pertinente, non esaurisce la richiesta, il MEDLINE può essere considerato quale banco di prova della strategia che potrà essere trasferita al MEDLARS con alte probabilità di successo.

Anche nella seconda generazione del MEDLARS, il MEDLARS II, il sistema conversazionale consentirà all'utilizzatore di modificare la sua questione in funzione delle risposte che a lui dà l'elaboratore. Tecnicamente, il passaggio dall'una all'altra generazione avverrà per gradi, tenendo conto degli investimenti necessari, sia in materiali che in programmi. Dalla condizione attuale in cui la formulazione della questione e la ricerca sono fatte «en différé» ('batch processing' o trattamento simultaneo di lotti di ricerche) si passerà ad una fase in cui la formulazione e, eventualmente, la indicizzazione saranno su base conversazionale, mentre la ricerca resterà «en différé». Nella fase ancora successiva la formulazione della questione avverrà su base conversazionale con accesso diretto alla parte più recente del materiale in memoria: il materiale rimanente sarà ancora trattato «en différé». Infine, si avrà formulazione della questione su base conversazionale ed accesso diretto alla totalità del materiale in memoria. In conseguenza del piano di sviluppo previsto, quindi, la formulazione attuale o «en différé» costituisce solo una tappa provvisoria, mentre la ricerca sarà sostituita progressivamente in un tempo che potrà essere più o meno lungo (4).

Tra i servizi collaterali al MEDLARS che l'I.N.S.E.R.M. mette a disposizione degli utenti, vanno segnalati l'S.S.M. (Service Signalement et Microfiches) e il S.I.N.B.A.D. (Système Informatique pour Banque de Données). Il primo è un servizio centralizzato di segnalazione e diffusione rapida della letteratura biomedica corrente al fine di integrare la rete documentaria fornita dalle biblioteche, aumentare le possibilità di lettura e fornire, in forma di microfiches, articoli individuali nel più breve tempo possibile. Operando su uno stock centrale di 300 periodici di ogni parte del mondo importanti in campo medico, l'I.N.S.E.R.M., dopo anni di studio e sperimentazioni pilota, ha messo a punto un circuito d'informazione che diffonde sistematicamente e tempestivamente i sommari dei singoli fascicoli dei detti periodici. Da questi gli utenti possono desumere quali siano gli articoli di probabile interesse, richiederli e poi riceverli a stretto giro di posta.

Ancora più interessante per le sue prestazioni attuali e potenziali aperture è il servizio indicato in breve come Banca dei Dati, in grado di fornire una risposta precisa ed istantanea a tutti i tipi di questioni che comportino dati. Il tempo di elaborazione è dell'ordine di un secondo e quindi in scala umana la risposta risulta immediata, in forma di dialogo, pur senza escludere l'interrogazione «en différé», come ad esempio nel caso di serie di tavole statistiche. La precisione e concisione della risposta (un valore di pH, una attività enzimatica, etc.) comportano una organizzazione rigorosa dei dati immessi nel sistema ed esigono la collaborazione di specialisti della disciplina nel rispetto di un codice comune che rende l'informazione accessibile a tutti. Come contropartita, ciascuno ha accesso ai dati raccolti dagli altri analizzatori secondo il principio della «analisi cooperativa». La validità

delle premesse al sistema ed, in particolare, la possibilità di cogliere il contenuto essenziale di un documento attraverso un codice analitico formato da un insieme di « grilles » è stata studiata in un esperimento pilota applicato al campo dell'enzimologia. Successivamente, nella fase operativa, l'I.N.S.E.R.M. ha portato il suo interesse sulla tossicologia, in quanto è il campo che forse più di ogni altro richiede risposte precise ed immediate e può riguardare, senza eccezioni, tutta la comunità. La Banca è oggi uno strumento di reale efficacia e la sua sfera d'azione tende sempre più ad estendersi (in Europa partecipano al programma, con possibilità di interrogazione telefonica, 20 Centri Anti-Veleno). Tuttavia, ancora più importante dei risultati pratici già conseguiti è il fatto che il carattere generale dei programmi li rende indipendenti dalla disciplina cui possono essere applicati. In seguito potranno rientrare in questa organizzazione altri domini di interesse e la priorità in questa organizzazione toccherà certamente a quelli nei quali l'esigenza di un'informazione precisa e rapida è più sentita (immunologia, farmacologia...). (In relazione a quest'ultima va detto, per inciso, che è allo studio e già in fase di sperimentazione preliminare un sistema che, ricalcando sotto qualche riguardo l'iniziativa dell'I.N.S.E.R.M., promette di essere di utilità nell'esercizio della pratica medica (5)).

Concludendo, l'informatica, o, più in generale, la scienza della documentazione sta vivendo anni di progresso e di perfezionamento: le vie della comunicazione scientifica e biomedica sono studiate e migliorate in tutti i loro aspetti. In questa direzione si inquadra la politica promozionale dell'Istituto Superiore di Sanità, che tende a sensibilizzare la comunità biomedica sulle possibilità di una informazione moderna mettendo a disposizione mezzi adeguati in un clima di proficua cooperazione.

BIBLIOGRAFIA

- (1) UNITED STATES. National Library of Medicine. The Principles of MEDLARS. Bethesda, Dept. of Health, Education and Welfare, 77 p. (1970).
- (2) ZERAZZA, J. L'I.N.S.E.R.M. et l'Information Biomédicale. *Bull. Bibliothèques France*, 17, 35-66 (1972).
- (3) LANCASTER, F. W. MEDLARS: Report on the Evaluation of its Operating Efficiency *Am. Docum.*, 119-142 (1969).
- (4) WARNET, N., P. DEVILLERS & M. LE MINOR. Documentation Médicale: un Exemple d'Application de l'Informatique. *Rev. Inform. Méd.*, 2, 171-181 (1971).
- (5) GREMY, F. L'Informatique au Service du Médecin Praticien. *Rev. Prat.*, 22, 1369-1373 (1972).