

DISCUSSIONI

Le nuove scelte di catalogazione semantica della biblioteca dell'Istituto superiore di sanità

di Maria Alessandra Falcone e Paola Ferrari

Nel corso degli ultimi anni si è più volte presentata la necessità di riflettere sui criteri strutturali e metodologici applicati nella scelta e nella costruzione delle relazioni sintattiche dei termini presenti nell'archivio automatizzato della nostra biblioteca.

L'evoluzione terminologica e disciplinare in campo scientifico, l'affermarsi di nuovi strumenti d'interrogazione online e la reale esigenza dell'utente di effettuare una ricerca il più possibile tempestiva e inerente alla specificità delle proprie richieste hanno ulteriormente contribuito a motivare nuove scelte di indicizzazione.

Inizialmente le nostre attenzioni si sono concentrate su una politica di radicale cambiamento di impostazione strutturale del soggetto: da un sistema post-coordinato a uno pre-coordinato qual è quello proposto dal Gruppo di ricerca sull'indicizzazione per soggetto (GRIS) che rappresenta tuttora un valido strumento di indicizzazione in campo semantico.

Tale metodo prevede tre fasi distinte: 1) analisi del documento; 2) costruzione della stringa di soggetto; 3) controllo terminologico.

L'aspetto innovativo di tale procedura, essendo le altre fasi momenti essenziali di una qualsivoglia corretta metodologia di indicizzazione, è da ricercarsi nel secondo punto. Questo processo prevede, infatti, una serie di operazioni analitiche per la costruzione di una stringa definita come sequenza logica di termini. Questi esprimono il soggetto del documento in modo univoco ed esaustivo attraverso un ordine di citazione dei termini scelti, regolato da principi logico-sintattici (concetto chiave, azione, agente ecc.).

Dopo più di un anno di sperimentazione ed applicazione pratica del metodo GRIS, sono emerse carenze e problematiche che hanno imposto un ulteriore approfondimento e una soluzione che nel tempo avesse potuto soddisfare le peculiari esigenze della nostra biblioteca.

Le difficoltà più evidenti da superare erano determinate dall'analisi del materiale bibliografico posseduto, costituito da testi per lo più a carattere pluridisciplinare e marcatamente specialistico in campo biomedico.

Infatti, l'uso di stringhe pre-coordinate regolate da relazioni logiche tra termini, specie per questo tipo di pubblicazioni scientifiche, non si presentava tale da eludere le difficoltà oggettive di esaustività descrittiva, dando spesso luogo, nella fase di consultazione del catalogo, ad ambiguità semantiche dell'enunciato di soggetto.

MARIA ALESSANDRA FALCONE, Biblioteca dell'Istituto superiore di sanità, viale Regina Elena 299, 00161 Roma, e-mail mfalcone@iss.it.

PAOLA FERRARI, Biblioteca dell'Istituto superiore di sanità, viale Regina Elena 299, 00161 Roma, e-mail paola.ferrari@iss.it.

Tale struttura terminologica, inoltre, presupponeva la costruzione di una rete di connessioni gerarchiche tra termini di indicizzazione preferiti nonché la possibilità di individuare, tramite una serie di relazioni sinonimiche o di equivalenza, i termini esclusi.

La difficoltà tecnica di creare, in presenza di stringhe di soggetto, un vero e proprio thesaurus correlando i vari termini selezionati ha contribuito a considerare la possibilità di adottare una diversa metodologia di indicizzazione.

L'utilizzo, in fase di ricerca, dei rinvii dai termini esclusi offre infatti maggiori garanzie di esattezza e completezza del soggetto, sebbene venga a volte da noi operata una scelta terminologica rispetto al MeSH, per non incorrere, nel caso ad esempio dei sinonimi, in inutili ripetizioni determinate dalla non sovrapponibilità dei termini inglesi con quelli italiani corrispondenti, che appesantirebbero inutilmente l'archivio.

Non del tutto irrilevante, anche se inerente ad un aspetto meramente tecnico del nostro sistema automatizzato, era la consultazione del soggetto, espresso in stringhe pre-coordinate la cui lettura non contestuale, ma frazionata, non permettendo la visualizzazione in un'unica schermata, contribuiva a dare poca chiarezza all'enunciato.

Un altro fattore, di non secondaria importanza, che ha influito su un cambiamento di scelta, riguarda l'interrogazione tramite operatori logici booleani, utilizzata per lo più da un'utenza biomedica ed effettuabile tramite sistemi automatizzati.

Questo metodo permette di selezionare i record contenenti almeno uno dei termini di ricerca tramite l'operatore OR, entrambi i termini con l'operatore AND ed escludere tutti i record contenenti uno dei due termini con l'operatore NOT.

In questo modo si effettua infatti una ricerca più precisa e mirata, essendo l'utente parte attiva nell'individuazione dei descrittori più idonei, mentre la visualizzazione di stringhe, precedentemente formulate dal catalogatore e non sempre così inerenti alla specificità dei contenuti, ne disorientava e limitava a volte la ricerca.

La revisione del soggetto, iniziata operativamente a partire dal marzo 2001, è stata anche favorita dall'avvenuta traduzione in italiano, da parte del Servizio Documentazione dell'ISS, del *Medical Subject Headings* (MeSH) prodotto dalla National Library of Medicine (NLM) di Bethesda, centro dei servizi di informazione in medicina degli USA.

Questo strumento rappresenta per noi un valido supporto thesaurale, costituendo una fonte autorevole di termini di carattere tecnico-scientifico selezionati da specialisti delle singole discipline. La revisione annuale, il continuo aggiornamento, l'introduzione di nuove voci con note di ambito complete, nonché l'aggiunta di rinvii a cui è sottoposto il MeSH, ne garantiscono l'attualità, la validità scientifica e la verifica qualitativa della terminologia.

Il thesaurus MeSH è composto infatti da:

- circa 20.000 descrittori principali (*main headings*);
- quasi 800 sottodescrittori (*subheadings* o *qualifiers*);
- più di 100.000 voci supplementari (*supplementary concept records*).

Il MeSH costituisce inoltre un sicuro ausilio per l'utente in quanto riporta una serie di termini non-MeSH, di descrittori cioè che, pur essendo parti essenziali della struttura ad albero perché passaggi di divisione necessari, è improbabile che siano utilizzati come termini di indicizzazione o di ricerca. Essi quindi rinvieranno a termini semanticamente correlati consentendo una ricerca a più ampio spettro.

Il MeSH non si serve dei codici di collegamento oggi convenzionali quali BT, NT, RT, USE FOR ecc., bensì di un insieme di segni espliciti realizzati specificatamente per il sistema MEDLARS (insieme di fonti di informazione in formato elettronico sulla biomedicina prodotto dalla National Library of Medicine di Bethesda).

Sono stati infatti introdotti legami quali:

- *see / seen from*, che rimanda da termini non preferiti a termini scelti (Cancer *see* Neoplasms / Neoplasms X Cancer);

– *see related / seen also from*, che indica la presenza di altri descrittori correlati concettualmente alla voce in esame (Blood pressure *see related* Hypertension / Hypertension XR Blood pressure);

– *consider also*, che rimanda a termini che possono essere linguisticamente correlati (Brain *consider also* Cerebr- e Encephal-).

La scelta della terminologia del MeSH ha quindi comportato la registrazione nel nostro archivio di tali codici di collegamento tramite un sistema di *cross-references*.

La presentazione dei termini che indicizzano ciascuna unità bibliografica segue inoltre, come per il MeSH, un ordine gerarchicamente strutturato in base al grado di rilevanza rispetto al contenuto concettuale del documento.

Così come proposto dal record bibliografico catalogato dalla NLM, verranno indicati nell'ordine le voci relative agli aspetti formali e di connotazione geografica dei documenti.

Inoltre, abbiamo ritenuto importante mantenere un archivio relativo alle voci di congresso nelle quali indichiamo anche la città e la data di svolgimento dello stesso, permettendo così un ulteriore punto di accesso per il recupero dell'informazione, qualora si conoscessero unicamente dati parziali.

La suddetta scelta si discosta da quella operata dalla NLM, in quanto quest'ultima fornisce solamente un'indicazione generica di "congressi".

Una medesima voce potrà inoltre svolgere la doppia funzione di voce principale o di suddivisione, a seconda del ruolo specifico in cui verrà utilizzata.

La politica di indicizzazione ha riguardato essenzialmente il rispetto del criterio di specificità nella selezione del descrittore MeSH più idoneo, l'assegnazione di tutti quei termini necessari a descriverne in modo esaustivo il contenuto, nonché la coordinazione tra due o più descrittori.

È inoltre da osservare come l'accertata maggiore consuetudine e familiarità della nostra utenza ad utilizzare i descrittori MeSH nell'interrogazione di basi di dati quali ad esempio PubMed che, tra le basi di dati bibliografiche, rappresenta il sistema più diffuso a livello biomedico, abbia semplificato la metodologia di ricerca.

Anche in passato nella nostra biblioteca si era fatto uso di tale strumento, ma esso non costituiva il principale thesaurus di riferimento, essendo molteplici le fonti alle quali il soggetto attingeva per l'individuazione dei descrittori e delle suddivisioni, con la conseguente disomogeneità dei termini thesaurali presenti nell'archivio.

Le maggiori difficoltà connesse in passato all'utilizzo del MeSH erano spesso determinate dalla non formalizzazione in lingua italiana di termini inglesi tecnici a volte difficilmente traducibili, ostacolo superato dalla recente traduzione della quale oggi ci serviamo per la costruzione del nostro thesaurus espresso per tradizione nella nostra madre lingua.

Il nostro lavoro è tuttora rivolto alla revisione sistematica e complessiva dei termini presenti in archivio, con indicazione in nota della fonte che, pur essendo nella maggior parte dei casi il MeSH, in alcune occasioni e per documenti di ambito disciplinare non pertinente alla medicina, risulta essere evinta da ulteriori e autorevoli soggetti quali ad esempio quello della Library of Congress o, nel caso di documenti italiani, prevalentemente da quello della Biblioteca nazionale centrale di Firenze.

La National Library of Medicine, consentendo la consultazione in rete del proprio catalogo ed offrendo una descrizione bibliografica completa dei documenti posseduti, costituisce per noi una valida traccia per l'ordine citazionale dei termini scelti che descrivono in modo esaustivo il contenuto concettuale del documento da indicizzare.

Con l'indicizzazione MeSH abbiamo visto aumentare considerevolmente il numero di termini, specialmente per quanto riguarda quelli relativi al settore della patologia medica, della biologia cellulare e dei composti chimici, rendendo necessario un lavoro di conseguente bonifica dei documenti precedentemente correlati a termini più generici.

Il nostro intervento ha anche riguardato le relazioni tra soggetti attraverso una revisione di tutte quelle voci che, mentre nel vecchio catalogo e nelle stringhe apparivano come voci principali o suddivisioni concettualmente legate tra loro, con la nuova soggettazione su base MeSH erano espresse con un unico termine. Ad esempio:

Accidents / prevention & control

see

Accident prevention

Incidenti – Prevenzione e controllo

vedi

Antinfortunistica

Acqua – Contaminazione radioattiva

vedi

Water pollution, radioactive (Contaminazione radioattiva dell'acqua).

Con l'utilizzazione del soggetto MeSH, abbiamo rilevato anche un'ulteriore conformità tra i soggetti prescelti per l'indicizzazione di un determinato documento e la classificazione NLM corrispondente, classificazione da noi già precedentemente utilizzata, ma che a volte presentava discrepanze rispetto ai termini di soggetto. Questo era dovuto alla scelta di termini non sempre appartenenti al Medical Subject Headings e pertanto senza una precisa correlazione con la classe da noi assegnata.

Dal momento che si è accennato alla classificazione è da notare, seppur brevemente, come il sistema di classifica, in uso presso la nostra biblioteca, sia conforme per quanto possibile al MeSH per quanto riguarda la medicina, le specializzazioni mediche e le scienze precliniche.

Per argomenti riguardanti la psicologia, la sociologia, la geografia, la tecnologia, la medicina veterinaria, la fisiologia animale nonché la biblioteconomia, si è resa necessaria l'integrazione con il sistema di classificazione della Library of Congress, sistema che utilizziamo come supporto sostanziale.

Attualmente il nostro archivio dei soggetti presenta inevitabili disomogeneità in quanto in linea continuano a coesistere termini strutturati in stringhe nonché descrittori non ancora controllati nella loro attendibilità. Il nostro obiettivo finale è che tale revisione venga al più presto completata, offrendo così non solo uno strumento di ricerca efficiente, ma auspicando che questo thesaurus possa costituire un valido esempio di scelte terminologiche-descrittive per l'indicizzazione del materiale di altre biblioteche scientifiche.