

[◀ Precedente](#)[Home page](#)[▶ Successiva](#)

Serafina Spinelli

I thesauri nelle biblioteche dell'Università []*



Fra gli anni Sessanta e gli Ottanta, percorrendo quella strada che veniva contestualmente spianata dallo sviluppo delle tecnologie informatiche - ed in particolare dei sistemi di *Data Base Management* - i thesauri si sono affermati come gli strumenti per eccellenza dell'indicizzazione assegnata, ma anche come espressioni dell'avanguardia della riflessione teorica nell'ambito della rappresentazione della conoscenza.

E' un volgere d'anni nel corso del quale s'innesca una sorta di circolo virtuoso, tale per cui si crea l'aspettativa che in un ragionevole lasso di tempo ogni campo della conoscenza sarà coperto dal suo thesaurus, che questo si verificherà almeno nelle principali aree linguistiche del mondo, e che i passi successivi potranno essere da una parte la costruzione delle connessioni fra i thesauri disciplinari di una certa area linguistica in un unico macrothesaurus, dall'altra la costruzione delle connessioni fra i diversi thesauri monolingue in un unico thesaurus multilingue. E queste non sarebbero state che le premesse per altre e più avveniristiche applicazioni, dallo scambio di record catalografici e bibliografici comprensivi di indicizzazioni semantiche a livello planetario, alla possibilità per i ricercatori di interrogare basi di dati multilinguistiche utilizzando per l'*information retrieval* la propria lingua nativa, sempre sulla base di procedure di *switching* fra linguaggi basate su strutture thesaurali.

Ma già all'inizio degli anni Novanta cominciano ad emergere alcuni elementi sulla scorta dei quali nel giro di pochi anni la situazione si modifica radicalmente:

- i costi di costruzione, sviluppo e manutenzione dei thesauri, nonostante l'aiuto dei *tools* informatici, rimangono molto elevati;
- la "mappatura" dei campi disciplinari è un lavoro ad alto contenuto sia tecnico che intellettuale, e sono pochi i terreni di convergenza di tutte le competenze necessarie (paradossalmente, non tanto le Università - in cui evidentemente non solo in Italia la collaborazione fra tecnici e specialisti della disciplina è cosa inusuale - quanto gli enti commerciali, ad esempio i grandi produttori di banche dati);
- alcuni dei "vecchi" linguaggi di indicizzazione si "travestono" da thesauri, inibendo lo sviluppo di thesauri veri e propri (un esempio per tutti, quello di LCSH, le intestazioni di soggetto della Library of Congress);
- l'aumento sempre più veloce della massa informativa e documentaria in Internet fa nascere motori di ricerca ed altri strumenti che cambiano radicalmente i modelli di ricerca delle informazioni degli utenti e i loro "orizzonti di attesa" (l'uso implicito e "debole" degli operatori booleani, l'approccio

per similarità e non per corrispondenza esatta, la presentazione dei risultati in ordine di rilevanza), riverberandosi in qualche modo anche sulle pratiche indicizzatorie;

□ l'evoluzione degli studi, fortemente condizionata dalla sempre crescente disponibilità di documenti in formato digitale, porta l'indicizzazione verso ambiti a più spinta connotazione tecnologica, come la costruzione automatica di strumenti di espansione lessicale delle *query* attraverso l'applicazione di *corpora* linguistici, come l'esplorazione, l'indicizzazione ed il reperimento dell'informazione nella ragnatela mondiale di Internet tramite le varie generazioni di agenti dai comportamenti sempre più sofisticati ("adattivi", "fuzzy", "evolutivi", o, in una parola, "intelligenti"), come la rappresentazione del contenuto concettuale di collezioni di documenti digitali attraverso l'indicizzazione semantica latente (che effettua la rappresentazione del contenuto semantico dei documenti in uno spazio vettoriale ed il reperimento delle informazioni pertinenti attraverso il confronto con una rappresentazione vettoriale delle *query*) o le tecniche di *information retrieval* su larga scala e di valutazione della pertinenza sorte dall'alveo delle *Text REtrieval Conferences*.

Conseguenza di tutto ciò è stata una più o meno lenta, ma progressiva riduzione degli spazi di sviluppo, dei terreni di coltura dei thesauri.

Ma proviamo a calarci più specificamente nella realtà italiana, ad analizzare la situazione da quel particolare - ma non poi troppo esclusivo - osservatorio che sono le biblioteche delle nostre università, e ad identificare i concorrenti più forti del modello thesaurale:

- sul piano indicizzatorio classico, il modello "Soggettario di Firenze", la cui pervasività è stata negli ultimi anni amplificata da SBN (vero e proprio agente "trasportatore" del Soggettario nel mondo delle biblioteche universitarie), e la cui capacità di attrazione si è dimostrata in grado di indebolire il peso (o meglio, di affievolire la capacità di percezione del peso...) di una serie di elementi negativi, come il mancato controllo semantico, l'insufficiente controllo sintattico, la cripticità delle basi concettuali di molte scelte;
- sul piano dei modelli emergenti di organizzazione della conoscenza, quello ipertestuale e "distribuito";
- sul piano della ricerca dell'informazione, il modello dei motori di ricerca per Internet, e, in qualche misura, degli opac in ambiente web.

Se facessimo un sondaggio fra le biblioteche del nostro Ateneo, non credo che avremmo grandi sorprese nel campo della costruzione e dell'uso di thesauri italiani per l'indicizzazione dei documenti posseduti (altra cosa è invece l'uso dei thesauri, soprattutto in lingua inglese, per l'*information retrieval* su banche dati prodotte da enti terzi...), se escludiamo quello che presentiamo oggi [\[1\]](#) e quello disciplinarmente adiacente in corso di sviluppo da parte del Dipartimento di Politica. A questi - certo - si aggiungono, ma è ancora altra cosa, i già esistenti thesauri che alcune biblioteche utilizzano come fonti di terminologia disciplinare.

La domanda che viene spontaneo porsi, atteso che i costi di costruzione di un thesaurus sono notevoli e non sottovalutabili, così come lo sforzo organizzativo correlato, è dunque: i thesauri hanno ancora un valore aggiunto rispetto agli altri strumenti d'indicizzazione e di ricerca? E, se la risposta a questa prima domanda è "sì", l'interrogativo successivo è: gli utenti sono in grado di apprezzare questo valore aggiunto, oppure ormai le tecniche di costruzione delle *query* si sono irrimediabilmente appiattite sui modelli di fatto (anche se non necessariamente!) semplificatori imposti dai motori di ricerca, i cui algoritmi di *ranking* fra l'altro non solo sono diversi da produttore a produttore, ma spesso celati alla stregua di segreti industriali? Insomma, detta in modo semplice al limite della brutalità: vale ancora la pena di costruire ed utilizzare thesauri?

Sarebbe un atto di presunzione da parte mia pretendere di dar risposta a questa domanda: troppe sono le variabili in gioco per poter azzardare previsioni, soprattutto in ambiti a così rapido e tutto sommato imprevedibile sviluppo come quelli delle tecnologie informatiche, e nel momento in cui sono scesi in campo, oltre a nostri ormai storici compagni di viaggio come gli informatici, i cognitivisti o i linguisti computazionali, anche i "bioingegneri" con le loro popolazioni di agenti o i fisici ed i matematici con i loro spazi semantici vettoriali. Però mi sento di dire che a mio parere ci sono tuttora diversi elementi che fanno del thesaurus un modello concettuale ancora valido e forse insuperato: la capacità di rappresentare i rapporti fra i concetti, e quindi in sostanza di strutturare il campo di conoscenza, in maniera rigorosa ma

non rigida, formale ma non arbitraria, risolvendo con eleganza gli elementi di ambiguità, di sovrapposizione, di equivalenza, della nostra lingua naturale.

Il thesaurus è dunque non solo uno strumento d'indicizzazione ma anche uno strumento portatore di conoscenza, sia per l'indicizzatore che per l'utente; il thesaurus è una rete per scandagliare la massa degli oggetti documentari e "catturare" l'informazione desiderata, una mappa per seguire percorsi semantici in grado di ampliare o restringere la ricerca, allargarla ai concetti più comprensivi o focalizzarla sugli argomenti più specifici. Che domani si arrivi a costruirlo automaticamente selezionando un *corpus* di testi, applicandovi qualche strumento lessicale precostituito o una matrice a n dimensioni e spingendo uno o due tasti, questa è un'altra storia, e, se ci saremo, speriamo di vederla anche noi.

Vorrei terminare identificando per i thesauri un possibile alleato, e vorrei farlo prescindendo una volta tanto dalle sempre più pirotecniche emergenze tecnologiche: sul piano della classicissima indicizzazione preordinata, i thesauri sarebbero strumenti ampiamente competitivi quando se ne perseguisse l'abbinamento a strumenti di controllo sintattico più sofisticati e rigorosi del Soggettario di Firenze. Il vero punto debole di questi metodi di costruzione della stringa di soggetto (GRIS primo fra tutti), ed il motivo per cui a mio parere non riescono ad affermarsi sul "mercato" dell'indicizzazione, è proprio la mancanza di strumenti di controllo lessicali con i quali realizzare il "connubio" necessario alla messa a punto di uno strumento di indicizzazione e di ricerca veramente completo. Se c'è una categoria professionale in grado di favorire questo connubio, è proprio quella dei bibliotecari italiani: nutriti da sempre col pane della "preordinazione", non hanno (o almeno non hanno ancora!), come i colleghi americani, ceduto alle lusinghe di uno smantellamento, per molti versi miope e falsamente determinato dal contesto tecnologico, della vecchia, ma solida e concettualmente motivata tradizione della coestensione della stringa. Abbiamo forse, per una volta, la possibilità di trasformare un ritardo (se si può definire ritardo la lentezza nel seguire un modello d'indicizzazione che sa più di *trend* che di scelta motivata) in un vantaggio: non lasciamocelo sfuggire.

Serafina Spinelli, Area della Biosfera - Università di Bologna, e-mail: spinelli@mail.cib.unibo.it

Note

[1] *Thesaurus italiano di sociologia*. Progetto di Daniele Danesi, a cura di Everardo Minardi e Pierina Bonvecchio. Firenze, Ifnet, 1999.

[*] Questo articolo riprende, con qualche modifica, il testo della relazione tenuta in occasione del Seminario di studio [*Un thesaurus per la sociologia italiana. Il ruolo dei thesauri per lo sviluppo delle scienze sociali in Italia*](#), Bologna, 11 ottobre 1999.

«Bibliotime», anno III, numero 1 (marzo 2000)

[← Precedente](#) [Home page](#) [Inizio](#) [→ Successiva](#)