

Interfaccia di ricerca

Albarosa Fagiolini
Luciana Franci

Biblioteca nazionale centrale di Firenze

Le caratteristiche generali del Thesaurus sono già state illustrate nei precedenti interventi. Descriveremo ora a grandi linee le funzioni di ricerca presenti nel vocabolario, l'organizzazione dei record dei termini, la visualizzazione delle gerarchie, le differenze tra l'interfaccia per l'utente e il data base gestionale. Non entreremo invece nel merito delle ragioni e dei criteri delle scelte metodologiche che stanno dietro ad ogni visualizzazione poiché sarà l'insieme dei contributi dell'intera giornata a fare luce su questo.

Accesso al Thesaurus e sua home page

Come già annunciato, all'interfaccia web del Thesaurus si accede tramite il cd-rom allegato al volume che stiamo presentando. La pagina di apertura del cd, oltre a permettere il collegamento alla base dati, tramite apposita registrazione dell'utente, consente di visualizzare la *Guida*¹ in formato pdf.

Dalla pagina di apertura del cd si transita (*Entra*) nella home page che comprende due parti distinte:

- la prima, su sfondo rosso, si riferisce al Thesaurus del *Nuovo soggettario*;
- la seconda, su sfondo bianco, è riservata al data base che cumula le voci (sia principali, che suddivisioni) dei precedenti strumenti di indicizzazione, cioè *Soggettario*² e *Liste di aggiornamento 1956-1985*.³

Descriveremo qui di seguito caratteristiche e funzionalità del Thesaurus del *Nuovo soggettario* rimandando alla parte finale dell'intervento una breve presentazione della base dati che cumula i precedenti strumenti.

Dalla parte superiore dell'home page, quella dedicata al Thesaurus del *Nuovo soggettario*, sono attivabili:

- la funzione di ricerca dei termini di indicizzazione (*Ricerca*);
- alcune informazioni sulla natura e gli scopi dell'intero progetto *Nuovo soggettario* (*Progetto*);
- l'elenco delle persone e enti che hanno costituito il gruppo di lavoro o partecipato a vario titolo (*Crediti*);
- le indicazioni su: standard e altri riferimenti, scelta dei termini contenuti nel vocabolario, modello strutturale, categorie e faccette, tipologie di note, ecc. (*Criteri*);
- la guida alla navigazione nel Thesaurus, le sigle delle fonti, delle relazioni semantiche, ecc. (*Aiuto*).

Funzioni di ricerca

Il programma prevede una navigabilità completa, fra tutti i termini collegati tra loro da una qualsiasi relazione semantica. La pagina per la ricerca si presenta come nella figura che segue.



Le possibilità di ricerca sono:

RICERCA PER TERMINE. Può avvenire con le opzioni: *Comincia per*; *Termine esatto*; *Contiene*. Nelle liste che si ottengono in risposta, i termini evidenziati in grassetto sono preferiti (cioè utilizzabili nella costruzione delle stringhe di soggetto), gli altri sono forme non accettate dalle quali si rinvia a termini utilizzabili.

RICERCA PER TOP TERM. Si realizza attraverso i termini che stanno al vertice delle gerarchie e corrispondono alle "faccette". Questa opzione permette di visualizzare l'elenco a tendina dei 13 Top term e, scegliendone uno, di accedere al suo record completo, compresa l'intera gerarchia sottostante, attraverso l'icona a triangolo.

RICERCA PER CATEGORIA. Permette di visualizzare l'elenco a tendina di tutte le faccette all'interno delle categorie e, scegliendone una, di accedere alla lista alfabetica dei termini che condividono la medesima categoria/faccetta. Gli stessi termini che, nella ricerca per Top term, sono presentati in ordine gerarchico, appaiono qui in ordine alfabetico.

RICERCA PER DDC. Avviene attraverso le classi Dewey attribuite alla maggior parte dei termini completamente strutturati del vocabolario, cioè quelli corredati di ogni tipo di relazione e di note. Può svolgersi attraverso le opzioni:

- *Comincia per*; interrogazione per troncamento a destra, che restituisce tutti i numeri, in ordine crescente, che iniziano con la sequenza numerica digitata seguiti dai relativi termini. Digitando ad esempio "30", si ottengono tutti i termini di ambito sociologico (da 300 a 307.776)
- *Numero esatto*
- *Finisce per* interrogazione, per troncamento a sinistra che restituisce tutti i numeri che terminano con la sequenza numerica digitata seguiti dai relativi termini. Ad esempio, digitando "092", si ottengono tutti i termini che rappresentano categorie di persone.

RICERCHE COMBINATE. Opzione che permette di incrociare vari campi: *Ricerca per termine e categoria*; *Ricerca per termine e numero DDC*; *Ricerca per categoria e numero DDC*.

STRINGHE DI SOGGETTO COLLEGATE AI TERMINI. Permette l'accesso alla lista alfabetica delle stringhe di soggetto collegate ad alcuni termini del Thesaurus. Le stringhe sono state scelte tra quelle elaborate nell'ambito della sperimentazione sintattica del nuovo linguaggio. Si può navigare dalle stringhe ai termini che le compongono, esclusi i nomi propri. Attra-

verso l'icona  si visualizzano i record bibliografici BNI collegati alla stringa.

Anche se già stato detto è bene ricordare che il Thesaurus non è al momento integrato con l'OPAC e che i legami presenti tra termini e stringhe e tra stringhe e descrizioni BNI sono una simulazione dimostrativa di quanto pensiamo debba avvenire a regime.

Organizzazione del record di un termine

Vediamo ora come è organizzato il record di un termine, ricordando che i tipi di termine presenti nel vocabolario sono: termini preferiti (e tra questi i Top term), termini non preferiti, etichette di nodo, cioè termini introdotti per le gerarchie ma non utilizzabili nell'indicizzazione, riconoscibili perché racchiusi tra parentesi quadre.

Il record di ciascun termine comprende generalmente cinque aree, visivamente separate, alcune delle quali non sempre presenti:

Area del termine, nella quale compare il termine seguito dall'icona della sua gerarchia, dalla categoria/faccetta, dall'eventuale Nota d'ambito e, in qualche caso, dall'icona  che permette di visualizzare stringhe di soggetto in cui il termine figura.

Alcuni termini sono affiancati dall'espressione, in caratteri rossi, che ne identifica la particolare natura: Termine non preferito, Etichetta di nodo, Top term.

Area delle relazioni standard, che comprende le relazioni di equivalenza (identificate dal simbolo **UF** per i preferiti e dal reciproco **USE** per i non preferiti), le relazioni gerarchiche (identificate dai simboli **TT**, **BT**, **NT**) e le relazioni associative (identificate dal simbolo **RT**).

Area delle relazioni di tipo storico, che contiene i legami che mettono in relazione termini non più usabili del *Soggettario* 1956 (e suoi aggiornamenti) con termini del nuovo che li sostituiscono.

La scelta di dedicare a queste relazioni un'area visivamente distinta corrisponde al fatto che la differenza tra il termine dei vecchi strumenti e la nuova forma è anche segnalata nella nota storica del termine preferito scelto per il Thesaurus.

I legami di tipo storico sono di due tipi:

- Il primo è stato attivato in presenza di differenze morfologiche o sinonimiche tra i vecchi e i nuovi termini. È identificato, per i termini preferiti (nuovi), dal simbolo **HSF** (Historical seen for) (ha come variante storica) e, per i corrispondenti non preferiti (vecchi) dal suo simbolo reciproco **HSEE** (Historical see, cioè variante storica di). Ad esempio nel record del termine **Bambini** è presente il legame **HSF** Fanciulli (forma del *Soggettario* non più accettata) e viceversa, nel record di Fanciulli è presente il legame **HSEE** **Bambini**.

La maggior parte dei termini che hanno un legame di variante storica (**HSF/HSEE**) hanno anche una relazione di equivalenza (**UF/USE**). Ed è il caso di **Bambini** appunto.

Tuttavia, quando i due termini differiscono esclusivamente per il numero (singolare/plurale), è presente solo il legame di variante storica.

- Il secondo tipo di legame storico è stato attivato nel caso di scomposizione di termini precedentemente composti. È identificato, per i termini composti non più accettati dal simbolo **USE+** (scomposto in) e, per i termini preferiti derivanti dalla scomposizione, dal suo reciproco **UF+** (presente nel termine composto non preferito). Ad esempio nel record di Partecipazione degli adolescenti (termine composto non più accettato) è presente il legame **USE+** **Adolescenti**, **Partecipazione** e viceversa, nei record di **Adolescenti** e di **Partecipazione** è presente il legame **UF+** Partecipazione degli adolescenti.

Questa relazione è prevista anche dagli standard, ma nel *Nuovo soggettario* è graficamente collocata tra i legami di tipo storico, perché finora utilizzata esclusivamente per i termini composti nei precedenti strumenti. Può anche darsi che in futuro il legame **UF+/USE+** possa comparire visivamente accanto a tutte le altre relazioni standard.

Area delle note e altri elementi. Sono riportate in questa parte:

- le Note, quando ritenute necessarie, e in un ordine stabilito che vede per prima la Nota di definizione, seguita dalla Nota sintattica e infine dalla Nota storica. L'ordine di presentazione delle note segue le indicazioni di BS 8723-2.⁴ Da osservare che la Definizione compare in questa area in quanto la sua funzione è riferita essenzialmente alla gestione del vocabolario, mentre la Nota d'ambito è presente nell'area del termine perché rappresenta un ausilio nell'indicizzazione dei documenti;
- il numero DDC, cioè la notazione Dewey attribuita alla maggior parte dei record con status Termine strutturato;
- la Fonte, cioè le sigle corrispondenti ai precedenti strumenti d'indicizzazione, dai quali il termine proviene e ai repertori sui quali il termine è stato controllato. Posizionandosi sulle sigle delle fonti, si visualizzano le loro citazioni complete; nel caso di fonti elettroniche, è possibile visualizzarne l'indirizzo e collegarsi direttamente alla risorsa indicata.

Area dei dati gestionali comprende:

- il nome dell'Agenzia catalografica che ha strutturato il termine;
- lo Status del record (status di lavorazione) che può essere: Termine strutturato, MIN e BASE. Per il significato dettagliato di queste sigle rimandiamo all'*Aiuto* on line e alla *Guida*;
- l'Identificativo del termine.

Visualizzazione delle gerarchie

Abbiamo già accennato alla possibilità di visualizzare la gerarchia dei termini preferiti, che si sviluppa sia verso l'alto (sequenza dei termini più generali, sovraordinati, fino al TT), sia, se presente, verso il basso (sequenza completa dei

termini più specifici, subordinati). Il numero dei cerchietti, che precedono ciascun termine della gerarchia, indica il livello di "profondità" della relazione di quel termine rispetto a quello di partenza, sia verso l'alto, sia verso il basso. Nel caso di termini poligerarchici (che hanno cioè più di un sovraordinato) le singole gerarchie vengono visualizzate separatamente e in sequenza. Ad esempio, **Poesia dadaista tedesca** ha un doppio BT (**Poesia dadaista** e **Poesia tedesca**) e sviluppa due gerarchie che conducono al medesimo sovraordinato di terzo livello **Poesia** e, attraverso la gerarchia di questo, al TT **Forme**.

Differenze tra l'interfaccia per l'utente e il data base gestionale

La base web del Thesaurus riproduce il contenuto della base gestionale così come si presentava nel novembre 2006, alla conclusione del lavoro prototipale. Il data base gestionale, cioè l'archivio in cui si è costruito il Thesaurus in questi due anni di lavoro, sarà aggiornato e implementato a partire dai prossimi mesi e le future edizioni accoglieranno tutte le modifiche e le aggiunte apportate.

Abbiamo scelto di pubblicare un'interfaccia utente diversa da quella del data base gestionale sia per limitare l'abbondanza di dati presenti, che, se indispensabile nell'attività di implementazione del Thesaurus, non è necessaria per le ricerche di indicizzatori o utenti, sia perché volevamo che gli stessi dati fossero visualizzati in un ordine stabilito e in aree anche visivamente distinte.

Esaminiamo quali sono le differenze più importanti tra l'interfaccia pubblicata, messa a punto da Edigeo, e il data base gestionale, versione adattata del software AGROVOC della Fao.

Nell'interfaccia utente ad esempio il codice Tipo di termine prevede, come già accennato in precedenza, solo quattro tipologie (Termine preferito, Termine non preferito, Etichetta di nodo, Top term) a fronte delle 11 presenti nel data base gestionale. Altri dati, importanti per una base di lavoro, non compaiono nell'interfaccia utente: non sono visibili le sigle degli operatori che hanno collaborato alla costruzione del Thesaurus né il Commento, area in cui possono essere annotati, nella fase di strutturazione dei termini, appunti, ricerche, dubbi, ecc. Inoltre, a differenza di ciò che avviene nella base gestionale, nell'interfaccia l'ordine di presentazione dei legami gerarchici e associativi, che è quello stabilito dagli standard, si integra con la presentazione alfabetica dei termini aventi lo stesso tipo di legame e, tra le relazioni gerarchiche, è visibile anche il Top term. Anche le note sono presentate, come già detto, in un ordine stabilito, coerente con quanto previsto dagli standard internazionali. Al contrario, nel data base gestionale, compaiono in un ordine, determinato dalla cronologia di immissione dei dati. Infine, l'interfaccia utente consente di visualizzare in maniera più snella le poligerarchie che hanno invece un aspetto alquanto complesso nel data base gestionale. Un'ultima osservazione: nella base gestionale sono previsti campi e codici che saranno utilizzati in futuro, e anche queste informazioni, quando attivate, con-

fluiranno nell'interfaccia utente. È già prevista, ad esempio, la possibilità di collegare un termine con il suo equivalente in inglese (e in prospettiva anche in altre lingue), oppure quella di utilizzare altri codici "tipo di termine" come *Termine proposto* e *Variante*.

Prima di concludere, facciamo una rapida incursione nella base che raccoglie, cumulandole, le voci dei termini del *Soggettario 1956* e delle *Liste di aggiornamento 1956-1985*. Come già anticipato da Anna Lucarelli, la digitalizzazione e il riconoscimento ottico dei vecchi strumenti, messi a punto solo nella fase finale dei lavori sul Prototipo, vogliono sottolineare la continuità tra vecchio e nuovo, pur nel quadro di un profondo rinnovamento. Già una parte dei termini dei vecchi strumenti è transitata nel *Nuovo soggettario* (al momento circa 3.600). A regime tutto il patrimonio terminologico dei vecchi strumenti confluirà nel *Nuovo soggettario* e stiamo pensando a dispositivi che rendano visibile questo passaggio da una base all'altra.

Pur essendo già stata effettuata una rapida revisione di questo data base non si esclude la possibilità di mancanze o lievi incongruenze rispetto ai testi in formato cartaceo. Nel corso dei prossimi mesi la base sarà periodicamente controllata e i refusi (compresi quelli del testo originario) progressivamente corretti.

La pagina web per la ricerca nell'archivio dei vecchi strumenti si presenta come nella figura che segue:

La funzione di RICERCA dei termini in questo archivio può avvenire per PAROLA CHIAVE attraverso cui si accede non solo al termine cercato ma anche a tutte le voci nel cui corredo esso è contenuto, oppure per INDICE VOCI o INDICE TERMINI. Ad esempio la voce **Beni mobili** cercata come parola chiave compare più volte nel data base: una volta per indicare la prima voce del *Soggettario* con tutto il suo corredo; una seconda per indicare la medesima voce come suddivisione nelle *Liste di aggiornamento 1956-1985*; varie altre volte all'interno del corredo (struttura sindetica e suddivisioni) di altri termini, risultanti dalla ricerca, compresi quelli composti. ■

Note

¹ BIBLIOTECA NAZIONALE CENTRALE DI FIRENZE, *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto. Prototipo del Thesaurus*, Milano, Editrice Bibliografica, 2006.

² *Soggettario per i cataloghi delle Biblioteche italiane*, a cura della Biblioteca nazionale centrale di Firenze, Firenze, Stamperia Il cenacolo, 1956.

³ *Bibliografia nazionale italiana. Soggetti. Liste di aggiornamento, 1956-1985*, Firenze, Biblioteca nazionale centrale, 1987.

⁴ BRITISH STANDARDS INSTITUTION, *Structured vocabularies for information retrieval. Guide. Part 2: Thesauri*, London, BSI, 2005, BS 8723-2:2005.