

Costruzione di thesauri

Dante R. Matteucci

Rapporto interno C81-13
Agosto 1981

Istituto CNUCE
Via S. Maria, 36
56100 P I S A

La presente nota e' stata scritta a fronte di attivita' svolta nell' ambito del programma internazionale CNR-National Museums of Canada di Ottawa. (Attivita' finanziata attraverso il Servizio Relazioni Internazionali del CNR).

1. Necessita' di un linguaggio normalizzato

La rappresentazione di un documento, con tutti i relativi significati, in una base di dati puo' essere realizzata sia riportando il documento originale per intero (linguaggio naturale), sia rappresentando il documento a mezzo di frasi, descrittori e parole-chiave ricavati dal documento o mediante una serie di codici di classificazione prestabiliti (linguaggio di classificazione).

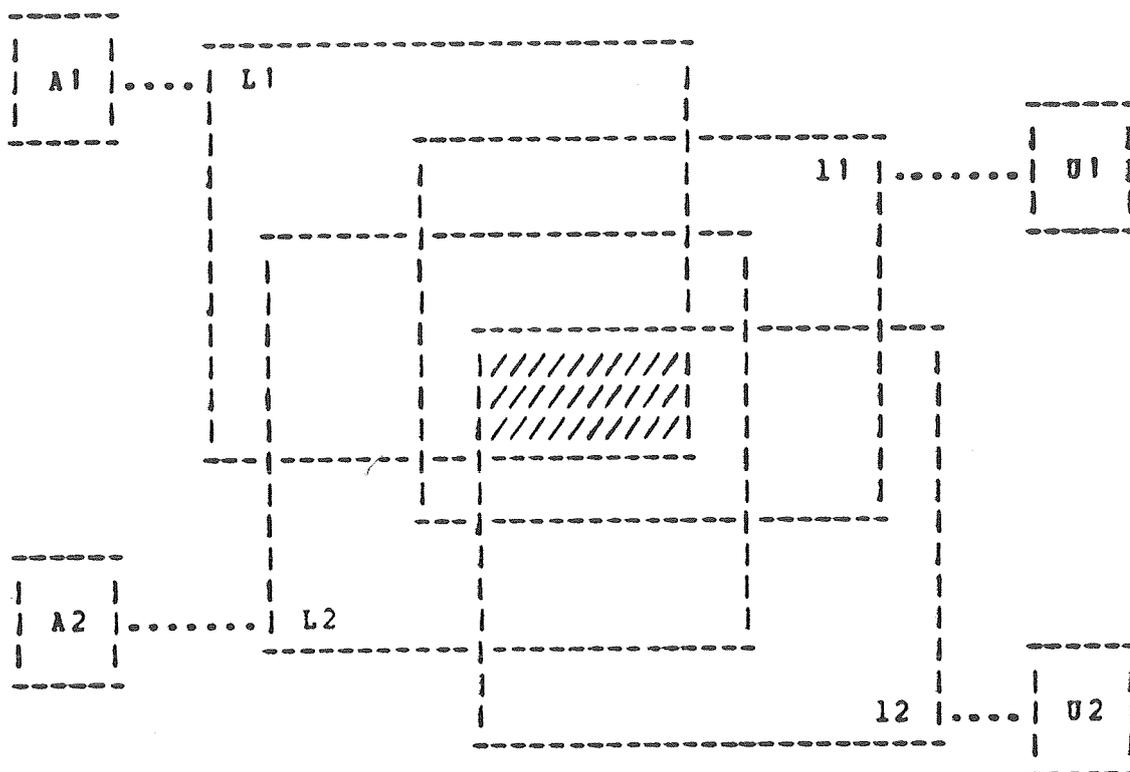
L' utilizzo diretto del linguaggio del testo, offre il vantaggio, nei confronti di un linguaggio di classificazione, della certezza di inserire tutte e sole le componenti semantiche del documento stesso, senza introdurre errori di classificazione o altre distorsioni dovute ai vincoli del linguaggio normalizzato.

Per contro, il linguaggio naturale presenta svantaggi legati alla ambiguita', soggettivita' e non univocita' del linguaggio stesso. L' esistenza di espressioni sinonime ed omonime richiede un notevole impegno nella fase di ricerca per assicurare la corrispondenza tra il linguaggio di richiesta ed il linguaggio del documento, cosi' da recuperare tutte le forme (sinonimia), senza dar luogo a recuperi errati (omonimia). Tale situazione e' schematizzata in figura 1, dove la zona tratteggiata rappresenta la coincidenza tra i diversi vocabolari.

Per un recupero efficace e fedele delle informazioni si richiede quindi l' adozione di un linguaggio normalizzato in cui ogni termine abbia un solo significato ed ogni concetto sia espresso in una sola forma. Tale linguaggio normalizzato, o documentario, determina l' efficacia del sistema informativo in quanto opera come strumento di controllo terminologico, sia nella fase di classificazione che in quella di ricerca. In altre parole esso agisce da ponte tra il linguaggio (naturale) usato nei documenti e quello (pure naturale) dell' utente della documentazione, facendo coincidere i rispettivi vocabolari (figura 2).

La scelta del linguaggio di classificazione piu' idoneo dipende dalle caratteristiche del sistema in cui esso deve operare. Nel seguito di questa nota si forniscono alcune indicazioni per impostare l'opera di normalizzazione delle informazioni pertinenti il catalogo dei beni culturali.

Nei metodi di rappresentazione basati sull' uso di codici di classificazione prestabiliti, molto usati nei primi sistemi informativi, il coordinamento dei concetti effettuato nella fase di analisi (precoordinamento), permette una classificazione molto precisa del documento. Cio' riduce il rischio di recuperare informazioni errate, ma richiede, per l' utente, una precisa conoscenza delle classificazioni usate. Inoltre tali metodi sono piuttosto rigidi per la difficolta' di prevedere le forme che saranno usate dall' utente in fase di richiesta.



A1,L1 autore1 e suo linguaggio
 A2,L2 autore2 e suo linguaggio
 U1,l1 utente1 e suo linguaggio
 U2,l2 utente2 e suo linguaggio

Figura 1 Aree di coincidenza tra i linguaggi naturali di autori e utenti diversi.

Anche i metodi di rappresentazione a parole-chiave possono creare qualche perplessita' nell' utente se, questi, non conosce precisamente le chiavi usate. Inoltre, anche se il coordinamento dei termini e' effettuato nella fase di ricerca (postcoordinamento), per il fatto che molte parole, prese singolarmente, mancano di significato proprio od hanno un significato poco preciso, si puo' determinare un elevato indice di recupero errato.

Le rappresentazioni ritenute piu' idonee sono pertanto quelle che fanno uso, per dare una descrizione semantica dei documenti, di descrittori o unita' elementari di informazione. E' opportuno che tali descrittori non abbiano un elevato livello di preordinamento (due-tre termini), per lasciare all' utente sufficiente possibilita' di coordinamento in fase di ricerca, cosi' da ridurre, in definitiva, la probabilita' di recupero errato.

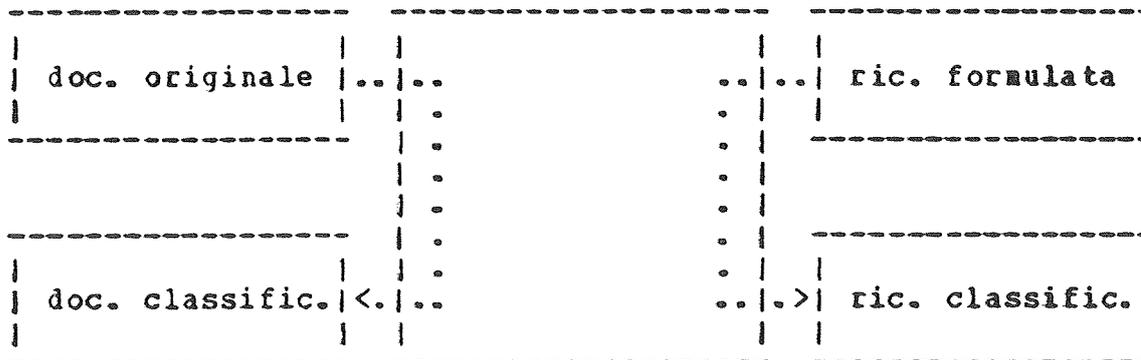
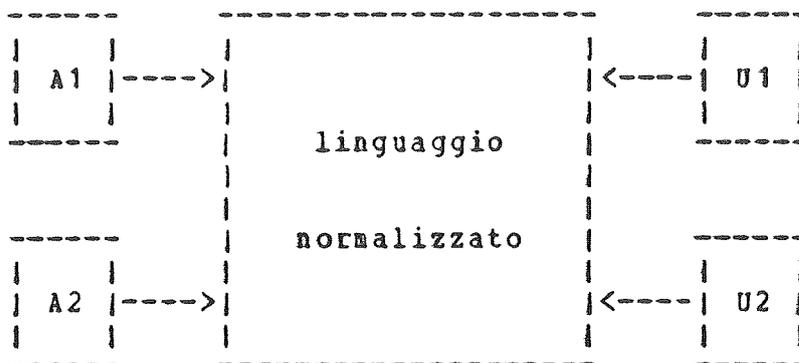


Figura 2 Funzione del linguaggio normalizzato

2. Formalizzazione e thesauri

Il linguaggio formalizzato e' composto dai descrittori usati e dalle relazioni che tra questi intercorrono. Una lista di descrittori normalizzati nella quale per ciascuno di questi sono date tutte le relazioni con gli altri termini ad esso, per qualche motivo, connessi, da origine ad un thesaurus.

Queste relazioni possono essere relative al significato della parola, o possono voler connettere ad un dato descrittore altri termini ad esso affini. Le relazioni piu' usate sono le seguenti:

- a) Relazioni di equivalenza o sostituzione: usate per stabilire la sinonimia tra i termini;
- b) Relazioni gerarchiche: usate per guidare da un dato termine ad uno di significato piu' ampio o ad uno di significato piu' ristretto;
- c) Relazioni di affinita': usate per associare un termine ad un' altro cui e' intimamente connesso, senza che tra' essi esista una relazione di gerarchia.

In un thesaurus si trovano tre livelli di termini:

- descrittori o termini preferiti,
- specificatori,
- termini di accesso.

I descrittori, che rappresentano i veri termini di lavoro del thesaurus, sono i termini usati nella classificazione per descrivere e nella ricerca per reperire i documenti; essi possono essere generici, specifici, di una parola o di piu' parole. Ogni descrittore identifica (specifica) una classe di documentazione.

Nei sistemi postcoordinati vi sono inoltre termini, gli specificatori, che servono ad identificare una classe di documentazione combinando diversi descrittori.

Nel thesaurus si trovano infine termini, detti termini di accesso, che non sono ne' descrittori ne' specificatori, ma che guidano verso un appropriato termine descrittore.

Esiste poi un quarto tipo di termini, gli identificatori, utili nella ricerca. Gli identificatori, o determinativi, sono nomi di oggetti o persone, che si possono aggiungere ai termini descrittori. Essi non sono normalmente inseriti nel thesaurus, eventualmente potranno essere ordinati in apposite liste o files.

3. Considerazioni per la costruzione di un thesaurus

Al fine di avviare la costruzione di un thesaurus e' necessario conoscere le esigenze del servizio di documentazione che deve utilizzarlo.

In particolare e' necessario prendere in considerazione quanto segue:

- a) Settore di interesse: occorre delimitare il campo che deve essere coperto, indicando le aree che possono essere trattate superficialmente e quelle ove e' necessaria una maggiore specificazione;
- b) Tipo di documentazione: occorre specificare il materiale da classificare (monografie, pubblicazioni periodiche, documentazione specialistica, cataloghi, ecc.); occorre anche fornire indicazioni sulla documentazione esistente e sul relativo incremento.
- c) Tipologia dell' utenza: si deve dare una specificazione qualitativa e quantitativa dell' utenza e della relativa percentuale di uso del servizio; occorre anche fornire indicazioni sul tipo di servizio offerto (retrospettivo, disseminazione selettiva);
- d) Tipologia delle domande: specificazione delle esigenze dell' utenza;

Sulla base delle precedenti considerazioni si potra' determinare:

- a) se e' necessario un thesaurus;
- b) il livello del thesaurus stesso;
- c) il settore disciplinario che deve essere coperto.

4. Costruzione di un thesaurus

Nella impostazione di un thesaurus e' opportuno passare attraverso le seguenti fasi:

- Selezione dei termini descrittivi,
- Definizione delle regole e convenzioni per la formalizzazione;
- Costruzione delle relazioni tra i termini,

4.1. Selezione dei descrittivi

La selezione dei termini o descrittivi si puo' ottenere in due modi:

a) Metodo induttivo

Consiste nell' estrarre i descrittivi dai documenti man mano che si procede alla loro classificazione. Il thesaurus risultante e' detto 'costruito a posteriori' e prendera' corpo dopo che sara' stata classificata una quantita' sufficiente di documenti.

b) Metodo deduttivo

Consiste nel ricavare i termini pertinenti un dato campo documentario da dizionari, indici, liste di classificazione, ecc. Il thesaurus risultante e' detto 'costruito a priori' e si crea prima di iniziare la classificazione dei documenti.

In realta', e' piu' opportuno ricavare il thesaurus definitivo combinando i risultati ottenuti dall'applicazione dei due metodi su esposti.

Nella selezione dei termini si dovra' tenere presente quanto segue:

- l' utilita' del descrittore nella classificazione e nella ricerca sara' determinata sulla base della frequenza relativa di uso sia nella letteratura, sia nel linguaggio dell' utente;
- l' accettabilita' del termine sara' determinata consultando dizionari, enciclopedie, monografie specifiche, esperti in materia; in ogni caso dovra' esserne verificata la forma esatta e l' espressione corretta;
- un descrittore puo' essere formato da una sola parola o da piu' parole, tante quante ne sono necessarie per esprimere il concetto. Si dovra' cercare di limitare, per quanto possibile, il numero di parole usate;
- per ciascun termine si dovra' verificare l' esistenza (ed eventualmente il tipo) di relazioni con altri descrittori gia' selezionati.

4.2. Definizione delle regole e convenzioni per la formalizzazione

La compilazione di un thesaurus comporta l' adozione di una serie di regole e convenzioni per controllare la forma dei termini, le sinonimie ed omonimie, la punteggiatura, ecc. Nel seguito di questo paragrafo sono indicate alcune delle principali regole che si possono adottare.

Il controllo della forma e' basato sulle seguenti regole:

- sara' usata, preferibilmente, la forma sostantiva; gli aggettivi saranno uniti ad un sostantivo per formare una espressione sostantiva;

Es. Attivita' artistica;

- i verbi non saranno mai usati, preferendo la forma sostantiva;

Es. Donazione, non Donare;

- la forma singolare o plurale dei descrittivi sarà determinata dalle normali regole idiomatiche; termini che abbiano diversi significati al singolare ed al plurale, saranno entrambi inseriti; In linea di massima si può dire che è preferibile la forma singolare per concetti che esprimono processi, proprietà e condizioni specifiche, nomi propri, discipline, ecc. Sarà usata invece la forma plurale per i termini che rappresentano proprietà caratteristiche e materiali generici, oggetti numerabili, eventi, ecc.
- per le parole con diversa ortografia sarà usata l'espressione più largamente ammessa, considerando l'altra come forma sinonima.

Es. Soprintendenza (principale)
Soprintendenza (sinonimo)

- sono possibili i seguenti tipi di forma:

- (i) parola singola
Es. Natività;
- (ii) nome + aggettivo
Es. Schiavo morente;
- (iii) nome + nome
Es. Palazzo Rucellai;
- (iv) nome + preposizione + nome
Es. Madonna del Granduca, creazione di Adamo;
- (v) nome + congiunzione + nome
Es. Castore e Polluce;

- le espressioni di più parole saranno normalmente registrate nell'ordine naturale e non in ordine invertito, senza far uso di caratteri o segni di congiunzione. Sarà opportuno, per quanto possibile, escludere preposizioni e congiunzioni.
- se per qualche motivo si debbono registrare espressioni invertite, l'attenzione deve essere posta sul termine più importante.

Es. Bartolomeo, fra' (Baccio Della Porta detto f. B.)
Il controllo dei sinonimi deve essere effettuato allo scopo di far corrispondere un solo termine a ciascun concetto; esso si realizza stabilendo opportune relazioni di equivalenza verso i termini preferiti.

Esistono due tipi di sinonimie: la sinonimia vera e propria, cioè l'identità sostanziale di significato, sotto ogni punto di vista, tra due o più parole o espressioni, e la quasi-sinonimia, relativa ai termini che, senza avere lo stesso significato, possono essere considerati sinonimi in determinate situazioni o ambiti.

Rientrano nelle sinonimie i seguenti casi:

- espressioni con forme grammaticali diverse;

- espressioni con diverso ordine delle parole;
- forme locali e popolari;
- forme superate
- nomi registrati

Rientrano nella classe delle quasi-sinonimie i seguenti casi:

- espressioni rappresentanti un diverso punto di vista della stessa proprieta';
- espressioni rappresentanti concetti con una considerevole sovrapposizione semantica.

Tra' i diversi sinonimi sara' scelto il termine principale preferendo, nell' ordine:

- la forma piu' comprensibile o piu' familiare all' utilizzatore,
- la forma che ha il minor numero di significati,
- una forma locale o un termine scientifico.

In ogni caso una adeguata relazione guidera' dai termini secondari a quelli principali.

Per quanto attiene il trattamento dei termini contrari si possono distinguere due casi: quello dei termini esprimenti reciprocita' o complemento che possono essere trattati come quasi-sinonimi e quello delle espressioni opposte o inverse che saranno trattate come termini separati.

Il controllo degli omonimi deve essere effettuato allo scopo di far corrispondere una sola nozione a ciascun termine. Le diverse accezioni di un omonimo potranno essere distinte a mezzo di qualificatori posti tra parentesi dopo il descrittore, ovvero ricorrendo a termini modificatori. Descrittori e qualificatori formeranno un descrittore composto.

Se un semplice qualificatore non e' sufficiente potranno essere inserite delle annotazioni piu' complete per spiegare, sia il significato del termine, sia le sue modalita' di applicazione. Tale annotazioni non dovranno fornire indicazioni di tipo generale, ma dovranno essere indirizzate al contesto del sistema.

Il controllo della punteggiatura perseguira' il criterio di evitare, per quanto possibile, caratteri speciali di punteggiatura. Si usera' di norma la parentesi per introdurre qualificatori e annotazioni; la virgola sara' usata come elemento separatore nelle espressioni invertite.

Es. Giovanni da Piesole, fra' (detto Beato Angelico)

Il tratto di congiunzione tra due parole sara' di norma eliminato, sia introducendo uno spazio, come nel caso delle combinazioni 'nome + nome' e 'lettera + parola', sia eliminandolo del tutto, come nel caso delle combinazioni 'prefisso + parola'.

Es. Palazzo Medici-Riccardi ==> Palazzo Medici Riccardi
Eta' pre-romana ==> Eta' preromana

Anche le forme abbreviate e le sigle saranno normalmente evitate per non introdurre ambiguita'. Esse infatti non sono sempre riconosciute, ed il loro significato puo' dipendere dal contesto.

Per la catalogazione delle opere d' arte (schede OA) sono ammesse le seguenti abbreviazioni:

attribuito	==> attr.
altezza	==> alt.
lunghezza	==> lungh
larghezza	==> largh.
diametro	==> diam.
circa	==> ca.
secolo, secoli	==> sec.
frazione (comune)	==> fraz.
localita'	==> loc.

4.3. Costruzione delle relazioni tra i termini

Nel paragrafo 2 sono state definite le diverse relazioni che possono intercorrere tra i vari descrittori di un thesaurus. Nel seguito di questo paragrafo sono indicate le forme di queste relazioni.

Le relazioni di equivalenza possono essere rappresentate con l' indicatore:

USA per guidare da un termine che non deve essere usato (secondario) a quello che gli e' preferito; puo' anche condurre a piu' termini da combinare;

Es. S. Maria Maddalena
USA La Maddalena

Es. Incensiere barocco
USA Incensiere + Barocco
dove il simbolo + indica l' operatore AND.

Esistera' poi l' indicatore reciproco:

UP (usato per) per guidare dai termini principali a quelli secondari.

Es. Scalpellino
UP Marmorario
Lapicida

Palmatoria
UP Bugia

Le relazioni di tipo gerarchico sono rappresentate mediante i simboli BT - NT, reciproci l' uno dell' altro:

BT (Broader Term - termine piu' ampio)

NT (Narrower Term - Termine piu' ristretto)

Le classiche relazioni gerarchiche sono quelle genere/specie e oggetto/tipo, come ad esempio:

Arti grafiche
NT Pittura

Pittura
BT Arti grafiche

Lapide
NT Lapide sepolcrale

Lapide sepolcrale
BT Lapide

In particolari casi vengono considerate gerarchiche anche certe relazioni intero/parte, come ad esempio:

Pistoia (provincia)
NT S. Marcello Pistoiese (comune)

Molto spesso i termini possono apparire in piu' di una gerarchia; si avra' cosi' ad esempio, la rappresentazione:

Ostensorio
BT Suppellettile liturgica
Oreficeria sacra

Le relazioni di tipo associativo sono generalmente rappresentate dal simbolo:

RT (Related Term - termine affine)

La relazione RT e' reciproca. La classica relazione RT e' quella di tipo parte/intero, ma ne esistono molte altre come indicato negli esempi che seguono:

oggetto/parte

Chiesa
RT Abside

oggetto/oggetto (come attributo)

Marmo
RT Scultura marmorea

oggetto/oggetto (attinenza)

Calice
RT Patena

oggetto/applicazione

Palmatoria
RT Processione

oggetto/processo

Dipinto
RT Pittura

processo/oggetto (agente)

Pittura
RT Tempera

Nella maggior parte dei thesauri i diversi tipi di relazione sono rappresentati in liste separate come segue:

Incensiere

UP Turibolo
Turribolo
BT Utensili sacri
NT Incensieri fittili
Incensieri barocchi
RT Navicella
Incenso

5. Thesaurus e prestazioni del sistema informativo

Le caratteristiche del thesaurus influenzano le prestazioni complessive del sistema informativo. Un thesaurus ben strutturato puo' migliorare infatti la classificazione dei documenti, conducendo ai termini specifici piu' appropriati per rappresentarne il contenuto e, d' altra parte, puo' migliorare la ricerca delle informazioni guidando l' utente ai termini usati nella descrizione semantica dei documenti.

Non e' pero' semplice definire la corretta struttura del thesaurus in quanto esistono diversi fattori che, se migliorano certe prestazioni, ne deteriorano altre.

L'efficacia di un sistema informativo puo' essere misurata per mezzo di due parametri:

il rapporto di richiamo, definito come:

$$RR = \frac{\text{n. di documenti pertinenti selezionati/}}{\text{n. totale di documenti pertinenti contenuti nel sistema}}$$

ed il rapporto di precisione, definito come:

$$RP = \frac{\text{n. di documenti pertinenti selezionati/}}{\text{n. totale di documenti selezionati}}$$

Questi due parametri sono legati tra loro da una relazione inversa, nel senso che se si cerca di aumentare il rapporto di richiamo (allargando la ricerca) si diminuisce la precisione, mentre se si tende a migliorare il rapporto di precisione (restringendo la ricerca) si riduce il richiamo.

Qui di seguito sono indicati i fattori che influenzano le prestazioni finali, suddivisi in:

- a) fattori di richiamo: tendenti ad assicurare la ricerca del numero ottimale di documenti pertinenti;
- b) fattori di precisione: tendenti ad impedire la selezione di documenti non pertinenti.

Sono fattori di richiamo:

- termini di accesso: perche' possono condurre l'utente dai concetti specifici ai descrittori usati nel sistema per rappresentarli;
- controllo della forma: perche' previene la perdita di documenti pertinenti come conseguenza della dispersione dei concetti tra le diverse forme dello stesso termine descrittore;
- controllo dei sinonimi: perche' previene la dispersione dei sinonimi tra termini descrittori diversi;
- relazioni di struttura gerarchiche e associative: perche' allargano la ricerca ai termini affini;

Sono fattori di precisione:

- specificita' del vocabolario: perche' piu' alto e' il livello di dettaglio, maggiore e' il numero dei termini e piu' precisamente si puo' descrivere il contenuto dei documenti;
- postcoordinamento: perche' coordinando tra' loro diversi descrittori si consente una definizione piu' accurata dei concetti;
- preordinamento: perche' i descrittori multitermine permettono di definire piu' minutamente il contenuto dei documenti;
- strumenti ausiliari di precisione: perche' l'introduzione di opportuni indicatori di legame, di ruolo o di relazione consente di ridurre le false coordinazioni e le correlazioni incorrette tra' i termini.

La specificita' e' il principale fattore di precisione nel reperimento; tuttavia l'alta specificita' tende a ridurre il richiamo: infatti la consistenza della classificazione diventa sempre piu' difficile man mano che aumentano i termini del vocabolario, perche' piu' sottile e' la differenza di significato che il vocabolario puo' esprimere. Inoltre l'alta specificita' richiede, sia per la classificazione che per la ricerca, personale altamente qualificato ed esperto.

Ricordando che la specificita' del vocabolario e' determinata dal numero di classi univocamente identificate, cioe' dal numero di specificatori piuttosto che dal quello dei descrittori, si puo' cercare una soluzione di compromesso, limitando il numero di termini descrittivi ed inserendo numerosi termini specificatori e di accesso, per ridurre la probabilita' di generare inconsistenze e, in definitiva, per aumentare il richiamo.

6. La soluzione del 'Public Archives of Canada'

Presso la Divisione Iconografica del 'Public Archives of Canada' e' in via di realizzazione un thesaurus da usare nella descrizione iconografica della collezione d'arte custodita presso la Divisione stessa. Tale thesaurus e' un lessico strutturato di termini con tre livelli di specificita':

a) una lista ristretta di grandi categorie di classificazione:

- araldica
- architettura
- oggetti
- costumi
- fauna
- flora
- paesaggi
- personaggi
- trasporti
- vita quotidiana

b) una lista limitata di termini di classificazione, per ciascuna categoria delle dieci che costituiscono il macrothesaurus. Attualmente il thesaurus e' stato predisposto per la parte che riguarda l'architettura, gli oggetti ed i costumi; per ciascuna di queste categorie sono state definite circa dieci suddivisioni di classificazione, come mostra l'esempio che segue per l'architettura:

- architettura commerciale ed industriale
- architettura domestica
- architettura governativa
- architettura militare
- architettura religiosa
- architettura sanitaria e scolastica

- architettura culturale e ricreativa
 - dettagli architettonici
 - opere ingegneristiche
 - strutture indipendenti non abitative
- c) una lista aperta di termini di classificazione a vocabolario controllato; a differenza dei primi due livelli che sono fissi, il terzo non e' limitato ma puo' essere espanso per rispondere a nuove esigenze di classificazione o di ricerca. Ad esempio per l'architettura si troveranno i nomi di tutti i tipi di edifici ed opere architettoniche come:

- banca, albergo, (commerciale)
- stadio, museo, (culturale e ricreativa)
- prigione, prefettura, (governativa)
- ospedale, scuola, (sanitaria e scolastica)

Mediamente sono stati inseriti circa 100-150 termini in ciascuna categoria di secondo livello.

In definitiva si hanno dieci grandi categorie (macrothesaurus), circa cento sottocategorie (secondo livello) e circa 10000-15000 termini specifici (terzo livello).

Poiche' i primi due livelli sono fissi, i corrispondenti termini sono stati codificati con numeri di due cifre ciascuno (ad es. 0203 architettura domestica, 0205 opere ingegneristiche) per semplificare il lavoro di registrazione.

Per una maggiore precisione si possono poi usare dei termini di qualificazione da inserire dopo il termine di terzo livello; per l'architettura si hanno i seguenti qualificatori:

- aborigeno
- decorato
- interno
- legno
- pietra
- terra

Ai termini del thesaurus possono infine essere aggiunte delle espressioni accessorie che permettono di specificare o determinare il soggetto identificato dal termine a vocabolario controllato. Tali elementi determinativi saranno inseriti tra parentesi, dopo il termine controllato, solo quando noti con certezza a seguito di analisi iconografiche o in base a documentazione.

Con i determinativi in linguaggio naturale si riesce ad inserire una documentazione iconografica precisa mantenendo, allo stesso tempo, i vantaggi della struttura di classificazione e del vocabolario normalizzato e controllato.

Un descrittore iconografico completo sara' pertanto composto come segue:

Architettura: architettura commerciale: banca: interno
(Monte dei Paschi di Siena)

Il thesaurus del 'Public Archives of Canada' e' caratterizzato da una moderata specificita' di termini descrittori o preferiti, unita ad una molteplicita' di termini di termini di accesso per guidare all' uso dei piu' appropriati descrittori. Cio' consente di contenere gli sforzi ed i costi di creazione del thesaurus e, al tempo stesso, fa si che non si debba utilizzare, per la classificazione, personale altamente esperto e specializzati in diverse discipline.

Per controbilanciare l' apparente perdita di precisione di questo approccio metodologico si confida sia sulla disponibilita' degli elementi determinativi, sia sul fatto che il sistema di reperimento, che sara' accessibile interattivamente, consentira' di effettuare ricerche con piu' parametri simultanei cosi' da raggiungere il necessario grado di precisione. D' altra parte il thesaurus viene considerato una struttura dinamica in costante evoluzione, che puo' essere continuamente migliorato anche per mezzo di informazioni statistiche relative alla frequenza di utilizzo di tutti i suoi termini.

Il thesaurus, come strumento di ricerca nel sistema informativo, e' poi integrato da una funzione che consente di stabilire un accesso poligerarchico all' insieme dei dati. Presa allora una parola come 'agricoltura', che comprende termini appartenenti a diverse categorie di classificazione (architettura, oggetti, costumi, personaggi, ecc.), a mezzo della predetta funzione si puo' effettuare una ricerca generale che permette di collegare i termini contenuti nel thesaurus a tutti i livelli di gerarchia necessari: sara' cosi' possibile trovare il riferimento agli abiti vestiti in campagna, agli edifici, agli strumenti e ad ogni altra cosa connessa con l' agricoltura.

Bibliografia

Aitchison, j. - Gilchrist, A. Thesaurus construction: A practical manual, ASLIB, London, 1972.

ANSI American National Standards, Guidelines for thesaurus structure, construction and use, ANSI, New York, 1974.

Chenhall, R. G. Nomenclature for museum cataloguing: A system for classifying man-made objects, A.A.S.L.H., Nashville, 1978.

Castonguay, D. Vocabulary control in Iconography at the Public Archives of Canada, Working paper P.A.C., Ottawa, 1979.

Lancaster, F. W. Vocabulary control for information retrieval, Information Resources Press, Washington, 1972.

Sorgel, D. Indexing languages and thesauri: Construction and maintenance, Melville Publishing Co., Los Angeles, 1974.