

EMILIO BERTOCCI

*I LINGUAGGI DI INDICIZZAZIONE*

2004

# INDICE

§ 1	3
Definizioni	3
§ 2	3
Distinzioni	3
§ 3	4
Soggettari e thesauri	4
§ 4	6
Il Soggettario di Firenze	6
§ 5	8
I sistemi di classificazione	8

## § 1

### Definizioni

L'indicizzazione è la creazione di liste di termini secondo un certo ordine e in un appropriato linguaggio per il recupero delle informazioni.

Un linguaggio di indicizzazione è una grammatica con termini convenzionali e regole sintattiche finalizzata all'organizzazione delle informazioni e alla capacità di risposta alle richieste dell'utente.

## § 2

### Distinzioni

Distinguiamo i linguaggi di indicizzazione in linguaggi alfabetici e linguaggi sistematici.

Dei linguaggi sistematici si dirà nel paragrafo 5.

I linguaggi alfabetici sono ulteriormente distinguibili in linguaggi a termini derivati e linguaggi a termini assegnati.

I linguaggi a termini derivati sono una sorta di ponte fra l'uso del linguaggio ordinario e l'uso di un linguaggio convenzionale. Infatti consistono nel selezionare termini derivati dal titolo o da altre parti del documento, accettando qualunque forma ordinaria con esclusione di una lista di termini non rilevanti (articoli, preposizioni, ecc.). Fra i più noti linguaggi a termini derivati ricordiamo quelli denominati KWIC (key word in context) e KWOC (key word of context) che si basano appunto sui termini del titolo.

Riepiloghiamo le distinzioni con il seguente schema generale:

#### Linguaggi di indicizzazione

alfabetici		sistematici	
a termini derivati	a termini assegnati	analitici	enumerativi

### § 3

## Soggettari e thesauri

Con i linguaggi a termini assegnati è necessario tradurre i termini ordinari secondo i principi e le regole del linguaggio adottato.

Per i limiti didattici di questo testo è sufficiente accennare ai soggettari e ai thesauri.

Un **soggetto** è l'argomento trovato in un documento mediante l'analisi del contenuto del documento stesso. Può essere espresso nel linguaggio ordinario o in linguaggio indicizzato.

Soggettari e thesauri assolvono la funzione di tradurre i termini del linguaggio ordinario in un linguaggio di indicizzazione.

In particolare, un **thesaurus** è un vocabolario di termini controllati (descrittori) collegati da una rete di relazioni esplicitate e riguardanti un certo ambito di studio.

Lo schema di relazioni dei descrittori è il seguente: dato un descrittore, esso è collegato almeno a tre tipi di termini: BT, NT, RT.

BT	Broader term (termine più largo)
NT	Narrower term (termine più ristretto)
RT	Related term (termine correlato)

Es.	Filosofia greca	BT	Filosofia antica
	Filosofia antica	NT	Filosofia greca
	Filosofia greca	RT	Filosofia romana

Un **soggettario** è un vocabolario controllato di voci generali e suddivisioni con le quali possono essere costruite stringhe secondo regole sintattiche e con una rete di collegamenti.

La soggettazione è l'attività di traduzione di un soggetto espresso in linguaggio ordinario nei termini accettati dal Soggettario adottato.

Un catalogo alfabetico per soggetti è un elenco di voci relative ai soggetti dei documenti posseduti ordinate e collegate secondo le regole del Soggettario adottato. E' uno strumento di informazione bibliografica che consente al lettore di controllare quanti documenti sono presenti nella biblioteca su un determinato argomento a prescindere dall'ambito disciplinare in cui è inserito. E' comunque da notare che il vantaggio dell'immediata individuazione del soggetto è bilanciato dallo svantaggio della dispersione dei termini che si riferiscono alla stessa disciplina.

## § 4

### Il Soggettario di Firenze

In Italia, la Bibliografia Nazionale Italiana (BNI) adotta il Soggettario di Firenze, elaborato dalla Biblioteca Nazionale Centrale di Firenze nel 1956, con successivi aggiornamenti. Le singole voci di soggetto possono avere suddivisioni di varia natura (generiche, geografiche, cronologiche, formali) che permettono la costruzione di una stringa in un ordine stabilito (precoordinato).

Esempio di stringa: Italia – Storia – Sec. 20. – Saggi (soggetto su una raccolta di saggi sulla storia d'Italia nel secolo XX).

Le voci sono collegate fra loro con appositi segni di rinvio e di richiamo, che costituiscono precise istruzioni per l'utente.

Il **rinvio** è effettuato da una voce non accettata, ma diffusa nel linguaggio ordinario, alla voce accettata, con i seguenti segni:

v. (vedi), rinvio dalla voce esclusa alla voce accettata

\* (asterisco), rinvio indiretto dalla voce accettata a quella esclusa

Esempio:

Arte preistorica v. Arte primitiva

Arte primitiva (voce accettata)

\* Arte preistorica (voce esclusa)

I **richiami** riguardano voci correlate e sono indicati con i segni:

v.a. (vedi anche), collegamento discendente: da voce generale a voce specifica

\*\* (doppio asterisco), collegamento ascendente: da voce specifica a voce generale

v.a. e \*\* possono indicare anche voci affini.

Esempio:

Sappiamo che filosofia greca è generale rispetto ad aristotelismo e specifica rispetto a filosofia antica. Dunque, i collegamenti saranno i seguenti:

Filosofia greca

v.a. Aristotelismo (collegamento discendente)

\*\* Filosofia antica (collegamento ascendente)

Filosofia antica

v.a. Filosofia greca (collegamento discendente)

Aristotelismo

\*\* Filosofia greca (collegamento ascendente)

Per evitare una eccessiva dispersione dei termini sono previste schede guida secondo un ordine differente da quello usato per la costruzione delle stringhe. Particolarmente importanti sono i soggetti biografici che raggruppano non solo le opere di un autore (sarebbero un doppione del catalogo alfabetico per autori), ma anche biografie, studi critici, inquadramenti storico-teorici.

La Biblioteca Nazionale Centrale di Firenze ha annunciato il superamento del soggettario tradizionale con la progettazione di un Thesaurus Italiano. E' l'epilogo di una situazione in cui gli aggiornamenti del Soggettario di Firenze sono stati rappresentati dall'inserimento di nuove voci, dalla diffusione della forma diretta, ma sono mancate nuove voci guida che avrebbe dato un senso logico al soggettario anche nell'età dell'information retrieval.

## § 5

### I sistemi di classificazione

I sistemi di classificazione privilegiano l'ordine sistematico per materie all'ordine alfabetico per soggetti.

Se nella soggettazione i legami disciplinari fra i soggetti erano dispersi, nella classificazione i soggetti sono distribuiti fra le discipline di appartenenza.

I sistemi di classificazione fanno uso di notazioni. La notazione può essere definita come un insieme dei simboli usato da un qualunque sistema organizzato di informazioni. E' pura o mista a seconda della uniformità o varietà dei simboli (numeri, lettere, segni speciali).

I sistemi di classificazione possono essere suddivisi in due gruppi: sistemi a classificazione enumerativa e sistemi a classificazione analitico-sintetica.

Il più diffuso sistema di classificazione, la Classificazione Decimale Dewey (DDC), è enumerativo. L'enumerazione consiste nell'elencare gli aspetti del documento da classificare mediante un procedimento di successive divisioni logiche e gerarchiche della classe (disciplina). In questo modo le relazioni fra le classi dovrebbero essere facilmente visualizzabili, una volta conosciuto il meccanismo di costruzione delle notazioni.

Per quanto riguarda la classificazione analitico-sintetica, a solo titolo di prima informazione, ricordiamo che è stata teorizzata soprattutto dal matematico e bibliotecario indiano Ranganathan e consiste in un procedimento di scomposizione di un soggetto in categorie semplici (cinque secondo Ranganathan: Personality, Matter, Energy, Space, Time) e di successiva ricomposizione disciplinare secondo un ordine preordinato basato su tali categorie, anziché gerarchico.

In particolare, data una disciplina di base, essa viene suddivisa nelle sue differenti caratteristiche ottenendo così un insieme di concetti per ciascuna caratteristica. Gli insiemi sono le faccette (facets). Un concetto è detto "focus" se interno a una faccetta o "isolato" se riferibile alle varie faccette.

In questo contesto, è sufficiente fornire solo alcune informazioni preliminari sulla DDC, che è utilizzata sia nella Bibliografia Nazionale Italiana, sia in molte biblioteche, soprattutto per la



collocazione dei libri. Infatti, data la sua importanza per la catalogazione semantica, la Classificazione Decimale Dewey sarà trattata a parte nel prossimo capitolo.

La DDC usa una notazione pura, in quanto è solo numerica.

Nella collocazione dei libri in scaffali aperti al pubblico, al numero della DDC seguono di solito le prime tre lettere indicanti l'autore dell'opera o il titolo, se manca l'autore. Comunque i criteri di collocazione possono variare a seconda delle tradizioni della singola biblioteca.

Devono, invece, essere uniformi i criteri di costruzione della notazione, altrimenti si tratterà di notazione simile, ma non di una notazione secondo le norme della DDC.

Per capire come si arriva al numero della DDC indicato nell'etichetta incollata sul dorso del libro, bisogna ricordare prima di tutto che ogni numero deve avere almeno tre cifre così distribuite: la cifra di sinistra indica la classe principale; la cifra centrale indica la divisione; la cifra di destra indica la sezione.

Esempio: 195 Filosofia italiana moderna

cifra di sinistra 1 (classe principale: filosofia)

cifra centrale 9 (divisione: filosofia occidentale moderna)

cifra di destra 5 (sezione: Italia)

Questa scelta notazionale risulta immediatamente più agevole rispetto a classificazioni

Dopo la terza cifra è inserito un punto (decimal point) per una sola volta.

I numeri aggiunti dopo la terza cifra sono suddivisioni degli schemi di classificazione (Schedules) e si trovano in apposite Tavole ausiliarie (Tables) o anche in altre parti degli schemi di classificazione o in apposite tavole di addizione. E' da tener presente che la costruzione della notazione è talora complessa e avviene mediante regole generali e istruzioni specifiche che il classificatore deve osservare.

I soggetti, ordinati alfabeticamente, insieme alle corrispondenti notazioni delle classi di appartenenza sono elencati in un apposito Indice relativo. Consultando l'indice si può risalire al numero assegnato al soggetto cercato.