

EMILIO BERTOCCI

*L'ESAME DEI DOCUMENTI*  
*(Norma ISO 5963)*

2004

## INDICE

I. La norma ISO 5963	3
II. L'esame dei documenti	5
III. L'identificazione dei concetti	6
IV. La selezione dei termini di indicizzazione	8
V. Il controllo della qualità	9

## I. La norma ISO 5963

1. Un documento è un insieme di informazioni contenute in un supporto materiale qualsiasi. Un documento può, quindi, essere un libro, un film, un disco, una diapositiva, un file, ecc.

L'analisi concettuale dei documenti è finalizzata al recupero da parte degli utenti delle informazioni riguardanti i più significativi argomenti trattati. Pertanto, interessa solo l'aspetto informativo e si prescinde dall'aspetto materiale (physical entity), oggetto della descrizione bibliografica.

2. La norma ISO 5963 *Documentation - Methods for examining documents, determining their subjects, and selecting indexing terms* (UNI ISO 5963 nella versione in lingua italiana) regola i procedimenti per l'esame dei documenti, la determinazione dei loro ambiti tematici (subjects) e la selezione di appropriati termini di indicizzazione (indexing terms).

La norma non presuppone alcun sistema di indicizzazione

3. In via preliminare, la norma fornisce alcune importanti definizioni dei termini di uso ripetuto. In particolare, bisogna segnalare la differenza fra concetto (concept) e soggetto (subject).

IL **concetto** è una unità di pensiero (a unit of thought), il cui contenuto semantico può essere riformulato ricorrendo a una combinazione di altri concetti.

Il **soggetto** è qualsiasi concetto o combinazione di concetti che, in un documento, rappresentino un tema (theme).

Tutti i concetti sono soggetti potenziali, tutti i soggetti sono concetti in atto, dove l'*actualitas* è data dai temi del documento.

In altri termini, il soggetto è costituito dai concetti che riguardano l'ambito tematico di un documento.

4. **L'indicizzazione** (indexing) è l'atto di creazione di un indice e consiste nel descrivere o identificare un documento sotto l'aspetto concettuale.

Il termine di indicizzazione è la rappresentazione di un concetto nella forma di un termine derivato dal linguaggio naturale o di un simbolo di classificazione, in uno specifico linguaggio di indicizzazione.

Un indice è un elenco di soggetti che si riferiscono alla posizione di ciascun soggetto nei documenti.

Le fasi dell'indicizzazione sono tre: esame del documento e individuazione del suo ambito tematico (subject content), identificazione dei concetti principali presenti nel soggetto, esplicitazione dei concetti principali nei termini del linguaggio di indicizzazione prescelto.

## II. L'esame dei documenti

Se i documenti sono a stampa, sviluppiamo l'esame considerando soprattutto i seguenti elementi:

- a) titolo;
- b) abstract;
- c) sommario (list of contents);
- d) introduzione, inizio di capitoli e paragrafi, conclusione;
- e) aspetti grafici;
- f) termini sottolineati o comunque evidenziati con caratteri speciali;

Titolo e abstract non devono essere assunti in modo acritico in quanto potrebbero anche condurre fuori strada come nel caso di titoli di fantasia e di abstract di sola propaganda editoriale.

L'importanza dell'introduzione è variabile essendo meno diffusa, rispetto al passato, la distinzione fra prefazione e introduzione, in cui la prima inseriva l'opera nel contesto culturale dell'epoca e discuteva le eventuali obiezioni, la seconda conduceva il lettore all'interno del testo spiegandone lo svolgimento e la partizione della materia trattata.

I documenti non a stampa hanno minori margini di analisi (il libro si può sfogliare, consultare, ecc.) e l'indicizzatore utilizza titolo e riassunto, quadro d'insieme (synopsis).

### III. L'identificazione dei concetti

1. Dopo l'esame dei documenti, la fase successiva consiste nella identificazione dei concetti principali presenti nel soggetto. A tal fine è importante preparare liste di controllo (check lists) di quei fattori rilevanti nell'ambito (field) coperto dall'indice. La norma fornisce, senza la pretesa di esaurire la materia, un questionario contenente i fattori generali che una lista di controllo dovrebbero stabilire.
2. I fattori da controllare sono i seguenti:
  - a) Il prodotto dell'attività svolta;
  - b) il concetto di attività (active concept) nel soggetto (azione, operazione, processo);
  - c) l'oggetto che subisce l'attività identificata;
  - d) l'agente che compie l'azione;
  - e) i mezzi per agire (strumenti, tecniche, metodi speciali);
  - f) i fattori precedenti contestualizzati;
  - g) la presenza di variabili;
  - h) la considerazione del soggetto da una angolazione atipica.
3. La scelta dei concetti dipende dalla finalità per cui i termini di indicizzazione dovrebbero essere usati e può essere influenzata dal supporto materiale (item).  
Le due caratteristiche in un indice che maggiormente subiscono l'influsso di questa scelta sono l'eshaustività e la specificità.

#### Il concetto di exhaustività.

L'eshaustività si riferisce al numero dei fattori di cui sopra, che sono rappresentati dai quei termini di indicizzazione assegnati a un documento.

La norma raccomanda di non tralasciare quegli aspetti, i quali, pur presenti nel documento, non sono di diretto interesse per la biblioteca, in quanto lo sviluppo delle reti di informazioni supera la chiusura di una indicizzazione specialistica monotematica.

Nel selezionare i concetti occorre preferire quelli sia di maggiore valore potenziale per esprimere il soggetto sia più appropriati per le esigenze degli utenti nel rispetto delle finalità dell'indice, modificando anche, se necessario, strumenti e procedure di indicizzazione.

#### Il concetto di specificità.

La specificità si riferisce all'estensione della corrispondenza di un concetto con il termine che lo rappresenta nel linguaggio di indicizzazione. L'identificazione dei concetti dovrebbe essere la più specifica possibile. Tuttavia, possono essere preferiti concetti più generali, quando l'indicizzatore riscontri un eccesso di specificità in contrasto con il buon funzionamento (performance) del sistema di indicizzazione oppure ritenga che un concetto, pur presente nel documento, abbia un rilievo minore così da giustificare l'indicizzazione a un livello più generale.

#### IV. La selezione dei termini di indicizzazione

Come si è già accennato a proposito della specificità, è necessaria la migliore corrispondenza fra concetti e termini del linguaggio di indicizzazione.

In caso di termini rappresentanti concetti nuovi occorre cautela e soprattutto la consultazione di autorevoli fonti primarie di informazione (dizionari, enciclopedie) e di attendibili linguaggi di indicizzazione (thesauri, schemi di classificazione). E' la competenza dell'indicizzatore nel trattare i linguaggi di indicizzazione, che fa la differenza fra buona o cattiva scelta dei termini di indicizzazione.

Nel quadro delle definizioni, ricordiamo che la norma distingue fra termine preferito e termine non preferito in un linguaggio di indicizzazione controllato. Il primo ("descrittore") è adoperato per rappresentare un dato concetto, il secondo è un punto di accesso nell'indice che rinvia l'utente al termine assegnato con appositi segni di rinvio. Nel *Soggettario di Firenze* il termine non assegnato rimanda a quello assegnato mediante le istruzioni di rinvio diretto e indiretto. In questo modo è possibile orientare convenientemente l'utente nell'uso dell'indice.

## V. Il controllo della qualità

Sono due i fattori che determinano la qualità e la coerenza dell'indicizzazione:

- a) qualificazioni e esperienze dell'indicizzatore
- b) qualità degli strumenti di indicizzazione.

L'obiettivo della coerenza (consistency), anche se forse solo ideale, non può mancare nel buon funzionamento di un sistema di indicizzazione soprattutto quando le informazioni scambiate sono in rete. E la coerenza richiede completa imparzialità dell'indicizzatore, evitando le deformazioni dei giudizi soggettivi accresciute dalle presenze di più indicizzatori. Di qui la raccomandazione di un controllo centralizzato.

E' importante, anche in questo ambito, la competenza dell'indicizzatore, ma la qualità dell'indicizzazione dipende anche dal riscontro pratico di essa, che l'indicizzatore riesce a verificare presso gli utenti.

D'altra parte, il linguaggio di indicizzazione garantisce migliore qualità se è flessibile. Di ciò si dovrebbe tenere conto nella scelta di esso.

Un controllo della qualità dell'indicizzazione si effettua con l'analisi dei risultati del recupero (retrieval) delle informazioni, per esempio calcolando richiamo (recall) e precisione (precision).

Il richiamo consiste nel recupero dei documenti che corrispondono ai soggetti contenuti nella richiesta dell'utente.

La precisione consiste nella effettiva pertinenza dei documenti richiamati alla richiesta dell'utente.

Richiamo e precisione sono inversamente proporzionali: ad un alto richiamo corrisponde una bassa precisione e viceversa.

Sulla risposta influisce il modo in cui è stata condotta l'indicizzazione: una indicizzazione approfondita (impiego di molti termini) e l'uso di termini troppo generici determinerà una scarsa precisione. Una soluzione, almeno teorica da proporre è l'impiego di termini specifici e indicanti il soggetto principalmente trattato nel documento (sommario specifica).