

Raccomandazioni per l'indicizzazione semantica nella Rete bibliotecaria di Romagna e San Marino

1. Oggetto del documento

Da alcune verifiche effettuate si è potuto constatare come l'attività di indicizzazione semantica, in particolare di soggettazione, sia assai cospicua all'interno del Polo.

Un semplice sondaggio sui soggetti immessi in Polo nel periodo 01/06/2014 - 28/10/2014 ha prodotto infatti il rilevante risultato di 3.774 stringhe di soggetto; nello stesso periodo sono state immesse 498 classi.

Si ritiene pertanto opportuno fare il punto sui diversi aspetti che riguardano l'indicizzazione semantica, non ultimi quelli inerenti le ripercussioni sui sistemi di *Front-end*).

2. Raccomandazioni

Le possibilità di indicizzazione semantica sono molteplici e da considerare fra loro complementari.

La ripetizione delle medesime informazioni in diversi punti del record non è di alcuna utilità per l'utente ed è antieconomico dal punto di vista del lavoro del bibliotecario.

Di seguito l'elenco dei canali di indicizzazione in essere:

Soggetti (solo in Polo)
Classi (solo in Polo)
Keywords (solo in Polo)
Tabella di Genere (Libro Moderno e Libro Antico, condivisa con Indice)
Tabella di Genere (Musica, solo in Polo)
Tipo pubblicazione (solo in Polo)
Forma letteraria (solo in Polo)
Tabelle di Dati aggiunti nei documenti musicali (Moduli Moderno e Musica), e nei Tipo documento "Materiale minore" (Moderno) e "Materiale video" (Moderno) (Solo in Polo)

I diversi metodi di indicizzazione semantica producono (o produrranno a seguito delle modifiche già programmate) lo stesso esito su ScopriRete (o, in generale, sui sistemi di front-office Web 2.0 compliant) ossia implementano le faccette o la nuvola di *tags*, grazie alle quali gli utenti possono facilmente navigare nel catalogo e raffinare i risultati delle proprie ricerche.

Si suggerisce pertanto che il catalogatore adotti - di volta in volta - il metodo di indicizzazione più efficiente (con particolare riguardo alla rapidità) ed efficace (con particolare riguardo al recupero dell'informazione), anche in relazione alle proprie competenze e specializzazioni. Da questo punto di vista è difficile fornire indicazioni di validità generale, se non qualche linea di comportamento come nel box sottostante.

- 1) Se il codice di genere contiene già l'indicazione di Atti di congressi, Mostre ed esposizioni, Letteratura per ragazzi, è inutile ripetere tali informazioni nel soggetto.¹
- 2) Se dalla natura bibliografica o dal tipo documento (tipo record Unimarc) si evince che si tratta di periodico, DVD, CD, Tesi di laurea, è inutile ripetere tali informazioni nel soggetto.
- 3) E' importante usare la classificazione quando si tratta di materiale "non soggettabile" (opere di fantasia, classici della materia: non soggetteremo "La Repubblica" di Platone, ma lo potremo utilmente classificare 321.07 STATO IDEALE).

E' possibile inoltre indicare alcuni principi di ragionevolezza cui attenersi:

- ⇒ Tenendo conto che la soggettazione, in particolare, richiede una conoscenza approfondita non solo della semantica del Soggettario, ma anche e soprattutto di una sintassi specifica, appare evidente che dovrebbe essere una attività svolta solo da colleghi che padroneggiano pienamente tali strumenti. In mancanza di competenze approfondite risulterà comunque utile e proficuo collegare *keywords* ai titoli. Si tratta di una attività consigliabile in molti casi: attività comunque impegnativa e da non sottovalutare. Allo stesso modo, se il soggetto e la classe attribuiti ad un' opera dovessero contenere gli stessi termini, sarà perfettamente inutile attribuirli entrambi: uno dei due basterà.
- ⇒ Buona pratica sarà quella di ricondurre i vari termini utilizzati nei diversi archivi semantici ad uno stesso termine. Il termine corretto di riferimento è quello previsto dal *Thesaurus* del Nuovo Soggettario, il quale può essere considerato a tutti gli effetti il vocabolario controllato "per eccellenza". A questo proposito comunichiamo che è in fase di ultimazione l'import del thesauro della BNCF in SOL e la gestione unificata dei sistemi di

¹ In caso di notizia a livello SUP è impossibile inserire un codice di genere mancante. In tal caso il soggetto potrà contenere le indicazioni formali non altrimenti esprimibili, ma converrà segnalare i BID al Servizio biblioteche che, disponendo delle necessarie autorizzazioni per intervenire in Indice dall'Interfaccia Diretta, provvederà ad inserire il codice di genere mancante. I codici di genere sono più numerosi dei 3 indicati, che sono i più ricorrenti (ci sono anche Bibliografie, Biografie, Cataloghi, Dizionari, Enciclopedie che potrebbero NON essere ripetuti nei soggetti).

indicizzazione semantica. La razionalizzazione della terminologia in uso migliora le risposte del catalogo, diminuendone il "rumore". Relativamente ai TAG di ScopriRete, si rammenta che termini uguali, indipendentemente dall'archivio semantico di provenienza, appariranno come un unico termine (che nella nuvola di tag apparirà di dimensione più grandi).

- ⇒ **L'indicizzazione semantica deve essere applicata al Titolo uniforme ove esistente.** Tenendo presente che a livello di *front-end* l'indicizzazione sarà comunque ereditata da tutte le manifestazioni collegate, è evidente il risparmio in termini di lavoro del catalogatore. A tal proposito si ricorda che è stato aggiornato il documento "Potenziamento della gestione dei titoli uniformi e loro indicizzazione semantica" (http://www.bibliotecheromagna.it/admin/PagPar.php?op=fg&id_pag_par=383&fld=file)
- ⇒ Pur con le opportune semplificazioni e ottimizzazioni, è doveroso che l'indicizzazione sia una pratica costante. Essa rientra infatti tra le **attività di natura condivisa e cooperativa** che rendono SBN un servizio e non solo un catalogo.