

Dai thesauri ai vocabolari controllati: alcune novità introdotte nell'ultima edizione dello standard ANSI/NISO Z39.19-2005*

EMANUELA CASSON

Nel luglio 2005 è stata pubblicata la 4ª edizione dello standard ANSI/NISO Z39.19-2005, Guidelines for the construction, format, and management of monolingual controlled vocabularies. Quest'ultima revisione inserisce considerevoli novità rispetto alle precedenti edizioni, basandosi anche sui risultati emersi dal Workshop on Electronic Thesauri (4-5 novembre 1999) e dall'indagine condotta dal NISO (febbraio 2003) sull'utilizzo dello standard. Gli aspetti più innovativi sono senza dubbio l'ampliamento dell'ambito di copertura dello standard, dai thesauri ai vocabolari controllati, e la maggior attenzione al contesto digitale, sia con precise indicazioni sul design del display a schermo sia con una nuova sezione sull'interoperabilità.

Parole chiave: Standards - Thesauri - Vocabolari controllati monolingue - ANSI/NISO Z39.19 - Display - Interoperabilità

Un po' di storia

Nel luglio dello scorso anno, è stata pubblicata la 4ª edizione dell'ANSI/NISO Z39.19-2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*¹, ma la sua storia inizia oltre trent'anni fa...

La prima edizione di uno standard americano per la costruzione e l'utilizzo dei thesauri risale infatti al 1974², ma l'ANSI Z39.19 venne aggiornato già alla fine degli anni '70 con una revisione che produsse, nel 1980, la 2ª edizione. È però con la nuova edizione nel 1993³ che si amplia in modo significativo l'ambito di

* Tutte le pagine web citate sono state controllate il 9 maggio 2006.

¹ ANSI/NISO Z39.19-2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*, scaricabile gratuitamente in formato PDF a <<http://www.niso.org/standards/resources/Z39-19-2005.pdf>>. Per accedere ai documenti di lavoro <<http://www.niso.org/committees/MT-info.html>>.

² La prima edizione dello standard ANSI Z39.19:1974, *Thesaurus structure, construction and use* fu prodotta nel 1974 dalla Subcommittee 25 on Thesaurus Rules and Conventions dell'American National Standards Committee.

³ ANSI/NISO Z39.19. *Guidelines for the construction, format, and management of monolingual thesauri*. Bethesda : NISO Press, 1994. Per accedere ai documenti di lavoro <<http://www.niso.org/committees/MT-info.html>>.

copertura dello standard che offre indicazioni anche su "format and management" dei thesauri monolingue, come indicato nel nuovo titolo. La 3^a edizione inoltre risente fortemente della revisione dell'analogo standard internazionale ISO 2788⁴ e della pubblicazione di quello inglese BS 5723⁵, avvenute in quegli anni.

Le altre due successive revisioni, del 1998 e del 2003, nascevano invece dalla necessità di adeguare lo standard ai continui cambiamenti che avevano interessato il mondo dell'informazione⁶. I termini del thesaurus infatti non erano più utilizzati per indicizzare solo le "classiche" risorse a stampa, ma anche nuovi tipi di documenti come brevetti, rapporti tecnici, musica, video...

Inoltre l'ambito digitale, in rapida evoluzione, premeva perché anche gli strumenti per l'indicizzazione e il recupero dell'informazione si adeguassero: i thesauri ne furono inevitabilmente coinvolti.

Workshop e indagine NISO: gli utenti sono soddisfatti dello standard?

Nel lasso di tempo tra le due revisioni del 1998 e il 2003, NISO [National Information Standards Organization] aveva promosso il *Workshop on Electronic Thesauri* (4-5 novembre 1999) al fine di studiare «the desirability and feasibility of developing a standard for electronic thesauri»⁷.

Benché l'edizione del 1993 fosse molto aggiornata per i tempi, lo standard aveva comunque ancora dei limiti innegabili dovuti al forte legame che permaneva con il passato e in particolare con il "mondo della carta".

Durante il *Workshop* furono individuati alcuni aspetti di fragilità di un ANSI/NISO Z39.19 ritenuto eccessivamente "biblioteca-centrico" e "documento-centrico". I limiti individuati erano:

- difficoltà di comprensione per i non tecnici. Molti potenziali utenti che avrebbero avuto interesse per lo standard non avevano però un'adeguata formazione biblioteconomica o in discipline affini per capirlo;

⁴ ISO 2788:1986. *Documentation - Guidelines for the establishment and development of monolingual thesauri*. Geneva : International Organization for Standardization, 1986.

⁵ BS 5723:1987. *British standard guide to establishment and development of monolingual thesauri*. London : British Standard Institution, 1987.

⁶ Per una panoramica sulla storia dell'ANSI/NISO Z.39-19, Emily Fayen, *An new standard for controlled vocabularies*. "The Indexer", 24 (2004), n. 2, p. 62-65.

⁷ Il report del *Workshop on Electronic Thesauri* (4-5 novembre 1999) è consultabile a <http://www.niso.org/news/events_workshops/thes99rprt.html>.

- applicabile all'indicizzazione documentaria, mentre gli altri supporti e formati erano ritenuti poco o per nulla rilevanti;
- indicato prevalentemente per le risorse a stampa, come dimostra il fatto che non si approfondissero possibili tipologie di *display* a schermo dei thesauri;
- troppo rivolto alla costruzione e alla manutenzione dei thesauri, accennando solo marginalmente ai principi di base del loro utilizzo;
- indirizzato al recupero post-coordinato dell'informazione. Lo standard infatti partiva dall'assunto che i vocabolari controllati hanno tra i loro fini principali quello di essere usati in sistemi post-coordinati di recupero dell'informazione e questo naturalmente era andato a discapito di altre indicazioni per altri tipi di recupero, compresi il *browsing* e la navigazione.

I punti deboli dello Z39.19 vennero ulteriormente approfonditi con un'indagine del febbraio 2003, promossa dal NISO, che mirava a capire l'utilizzo che veniva fatto dello standard⁸: i risultati dimostrarono che gli interpellati si aspettavano molto dalla revisione in corso. Tuttavia la versione del 2003 trasse solo parziali benefici dai *feedback* avuti in questo modo dalle comunità interessate, poiché il lavoro di aggiornamento era già prossimo alla conclusione.

Come indica Amy J. Warner, *project leader* del Thesaurus Development Team per la versione del 2003, c'erano infatti dei tempi da rispettare, poiché si voleva ridurre il ciclo di aggiornamento a soli 8 mesi, tanto che si era scelto, per non dover attendere il benessere per una nuova edizione, di limitarsi ad una nuova versione⁹.

Tra le numerose esigenze emerse dall'indagine, era evidente la forte necessità di uno standard che si potesse applicare tanto ai tradizionali thesauri dell'ambito cartaceo quanto a quelli digitali e quindi, aumentando la quantità e di conseguenza la varietà dei destinatari, si sentiva anche l'urgenza di interoperabilità tra protocolli, strutture e semantica.

Nello stesso periodo anche il BSI working group, nel corso della revisione del BS 5723:1987¹⁰, stava affrontando problemi simili: il contesto digitale, l'interoperabilità, la mappatura tra i thesauri e altri tipi di vocabolari controllati, i formati per lo scambio di dati e le varie necessità degli utenti, oltre che quelle dei professionisti dell'informazione.

⁸ Il *Summary of survey results* è consultabile a <<http://www.niso.org/committees/TRAG/Z39-19-Survey.pdf>>

⁹ Alan R. Thomas - Sandra K. Roe, *An interview with dr. Amy J. Warner (June 2003)*. "Cataloging & Classification Quarterly", 37 (2004), n. 3/4, p. 193-198.

¹⁰ Il BSI working group ha aggiornato recentemente il BS 5723:1987 che è diventato BS 8723-1:2005. *Structured vocabularies for information retrieval. Guide. Definitions, symbols and abbreviations*. London : British Standard Institution, 2005 e BS 8723-2:2005. *Structured vocabularies for information retrieval. Guide. Thesauri*. London : British Standard Institution, 2005.

L'ANSI/NISO Z39.19 oggi

Con il costante aumento delle quantità di informazioni disponibili, diventa sempre più importante saper individuare quelle cercate. È in questa direzione che il controllo terminologico assume ulteriore rilevanza, così da migliorare da una parte l'efficacia dell'archiviazione dell'informazione e dall'altra i sistemi di recupero che consentano sia di identificare che di localizzare il contenuto desiderato.

In questo contesto, l'ANSI/NISO Z39.19-2005 amplia il concetto di "documento" preferendo l'uso del termine "content object", per indicare ogni entità portatrice di informazione, che può esistere in forma fisica o digitale. Sebbene questa distinzione sia rara, ci sono due classi di *content objects*, quelli primari e quelli secondari:

- un *content object* primario è l'item stesso, sia che esista in forma fisica (per es. periodico, audiocassetta, DVD...) sia che esista solo in formato digitale;
- un *content object* secondario consiste invece nei metadati che descrivono il *content object* primario.

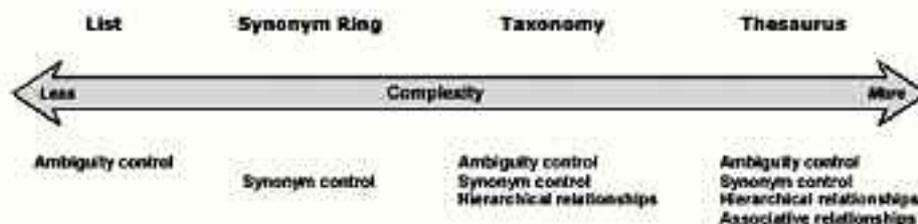
La nuova edizione dell'ANSI/NISO Z39.19 manifesta così subito una variazione di approccio, ma è anche più rigorosa nel normare non solo quello che lo standard consente e quello che non consente, ma anche quello che suggerisce. C'è infatti un set di disposizioni di base sulle prassi e le tecniche preferite, oltre ad alcune opzionali. Le indicazioni sono formalizzate in modo convenzionale: *must* (obbligatoria), *should* (consigliata) e *may* (opzionale); vi sono inoltre *must not* (non consentito in conformità con lo standard) e *should not* (non consigliato).

Sono però almeno tre, fra gli altri, gli aspetti di questa ultima edizione dello Z39.19 su cui è opportuno soffermarsi: non si occupa più solo di thesauri, ma dei vocabolari controllati; si sofferma su numerose possibili modalità di visualizzazione a schermo del vocabolario controllato e infine segnala alcune soluzioni per l'interoperabilità.

Fin dal nuovo titolo, lo standard segnala quella che è la principale novità e cioè l'ambito di applicazione: non più i soli thesauri, ma l'intera classe dei vocabolari controllati.

I vocabolari controllati sono strutturati in modo da rendere chiari i diversi tipi di relazioni tra i termini che li compongono e ci sono diversi tipi di vocabolari controllati a seconda del crescente grado di complessità della struttura: liste, *synonym rings*, tassonomie e thesauri¹¹.

¹¹ È utile scorrere la sezione 4 dell'ANSI/NISO Z39.19-2005 che fornisce definizioni, abbreviazioni e acronimi.



ANSI/NISO Z39.19-2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*, p. 17.

Una lista è un semplice gruppo di termini preferiti privi di struttura e di solito presentati in ordine alfabetico o in altre sequenze logiche, per es. Extra Large può essere selezionato come termine preferito da includere in una lista al posto di XL. Spesso la lista viene usata nei siti web in quelle che sono chiamate *pick list*.

Ad un livello leggermente maggiore di complessità si colloca il *synonym ring*, che è una lista di sinonimi e quasi sinonimi, senza però l'identificazione di un termine preferito, usati per migliorare il recupero dell'informazione in sistemi in cui il contenuto non è indicizzato o in cui il vocabolario di indicizzazione non è controllato.

Una tassonomia invece è una lista di termini preferiti, in cui ogni termine appartiene almeno a una struttura gerarchica, ed è spesso usata nei sistemi di navigazione nel web.

All'apice della complessità, troviamo il thesaurus che ha tutte le proprietà dei precedenti vocabolari controllati e in più altri tipi di relazioni, che sono in genere definiti nel momento in cui viene costruito.

Di solito un thesaurus è composto da una lista alfabetica di tutti i termini preferiti e non preferiti, con tre tipi di relazioni per ogni termine (equivalenza, gerarchica e associativa). Oltre al *display* alfabetico, un thesaurus può organizzare i termini anche in rappresentazioni gerarchiche: sono proprio questi modi di visualizzazione che contraddistinguono profondamente un thesaurus dagli altri vocabolari controllati più semplici.

Vista la varietà di vocabolari controllati ci sono quindi anche diversi tipi di *display* per i quali vengono definiti dei requisiti e delle raccomandazioni: ecco un'altra sezione¹² dello standard ampiamente rinnovata in questa 4ª edizione.

¹² La sezione 9 "Displaying controlled vocabularies" dell'ANSI/NISO Z39.19-2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*, p. 57-81.

Il *display* di un vocabolario controllato innanzi tutto deve tenere in considerazione una serie di elementi per un'efficace visualizzazione:

- la presentazione, che deve essere precisa nel rappresentare le relazioni tra termini e nell'indicare le distinzioni tipografiche (uso delle maiuscole, i corsivi, i grassetto...);
- i tipi di visualizzazione che possono essere utilizzati, dalla semplice lista alfabetica al complesso *display* grafico¹³;
- il formato in cui può essere prodotto e distribuito il vocabolario controllato, cioè in forma stampata o digitale, oppure entrambe. Il web e l'ipertestualità inoltre ammettono *display* aggiuntivi il cui potenziale naturalmente non ha uguale validità nella forma stampata;
- la documentazione, che deve essere sempre fornita all'utente per indicargli come utilizzare il vocabolario controllato.

Oltre al *display* alfabetico, le tassonomie e i thesauri devono includere un *display* gerarchico che illustri le relazioni. Queste possono variare a seconda di come il *display* gerarchico è stato strutturato: a multilivello, ad albero, a *top term*, a faccette...

Un tipo di *display* notevolmente avvantaggiato dall'ambito digitale è quello grafico; inoltre numerosi studi hanno dimostrato che questa forma di visualizzazione può comunicare le relazioni tra concetti in modo più efficace rispetto ad una visualizzazione in un *display* lineare.

Dell'ampia sezione sulla visualizzazione dei termini dei vocabolari controllati, è soprattutto la parte sui formati a schermo e web ad essere di straordinaria novità.

Lo standard parte dalla considerazione di base che la tecnologia ha prodotto molti *software* che offrono la possibilità di creare *display* di termini e relazioni tra termini, ma, se le opzioni dei *display* elettronici offrono significativi miglioramenti rispetto al formato a stampa, richiedono anche considerazioni aggiuntive sulla progettazione.

Per la stesura di questa parte dello standard infatti si sono considerati, oltre agli standard esistenti sull'interazione uomo-macchina, anche studi per le interfacce utente, per gli *screen layout*, per i menu, per le finestre, per le pagine web e per gli altri formati di interfaccia. Naturalmente anche i risultati degli studi e le relative norme per l'accessibilità da parte di persone con disabilità devono essere applicati rigorosamente a tutti i *display* progettati per i vocabolari controllati.

I vocabolari controllati infatti possono essere prodotti in molteplici formati: quello a stampa, una riproduzione digitale del documento a stampa (di solito in for-

¹³ Le caratteristiche e gli usi di ciascun tipo di *display* sono presentati nell'*Appendix C. Characteristics and uses of controlled vocabulary display options*, in ANSI/NISO Z39.19-2005, p. 137-141.

mato PDF), una versione interattiva *on line* che è ricercabile tramite una *query* e/o *browsing* e infine un'applicazione *software* locale che può essere installata sul PC dell'utente.

La visualizzazione di informazioni su uno schermo si differenzia da quella in forma stampata, che rende di sicuro più faticosa la lettura di un testo; tuttavia vi sono degli innegabili vantaggi che si riflettono anche nei *display* dei vocabolari controllati.

La ragione per cui un utente si rivolge ad un vocabolario controllato è quella di individuare rapidamente i termini appropriati da usare nella sua ricerca o nell'indicizzazione, ma, se nella forma a stampa questo compito è assolto dagli indici permutati, l'ambito digitale offre l'opportunità di ricercare per parole chiave tra tutti i termini presenti nel vocabolario. Anzi, il maggior vantaggio in questo senso è che le ricerche con parola chiave richiamano tutte le occorrenze del termine, anche nei termini composti.

La maggior parte delle opzioni di ricerca per parola chiave fornisce parecchie alternative per gli utenti, come il Medical Subject Headings (MeSH) Browser¹⁴.

Enter term or the beginning of any root fragments: or

Search for these record types:

- Main Headings
- Qualifiers
- Supplementary Concepts
- All of the Above
- Search as MeSH Unique ID
- Search as text words in Annotation & Scope Note

Search in these fields of chemicals:

- Heading Mapped To (HM) (Supplementary List)
- Indexing Information (II) (Supplementary List)
- Pharmacological Action (PA)
- CAS Registry/EC Number (RN)
- Related CAS Registry Number (RR)

Individuato il termine desiderato, un vocabolario controllato in ambito digitale deve permettere anche di far visualizzare all'utente le note storiche, le note d'ambito, le eventuali definizioni oltre che, ovviamente, tutte le relazioni gerarchiche, associative e di equivalenza di un termine, così da avere una visione generale della struttura del vocabolario che invece di solito risulta parziale.

¹⁴ MeSH <<http://www.nlm.nih.gov/mesh/MBrowser.html>>.

Vi sono poi altre norme dedicate nello specifico a vocabolari controllati che utilizzano un *browser* web per la visualizzazione e sono indicazioni relative alle tecniche di navigazione, al *browsing* e all'ipertestualità.

L'uso dell'ipertestualità e degli altri strumenti di navigazione nel web può rendere molto più facile muoversi in un vocabolario controllato, permettendo agli utenti di:

- navigare di livello in livello fino a trovare quello dalla specificità desiderata;
- collegare le definizioni, le note d'ambito, le note storiche soltanto quando necessario;
- muoversi facilmente da un tipo di *display* all'altro (per esempio da un elenco alfabetico ad una struttura gerarchica o da un formato grafico di una mappa concettuale a una visualizzazione dettagliata del termine).

Questo interesse dell'ANSI/NISO Z39.19-2005 per le tecnologie informatiche si conferma in particolare nella nuova sezione dedicata all'interoperabilità¹⁵, cioè la capacità di due o più sistemi di scambiare informazioni tra loro.

Visto che è cresciuta, e cresce continuamente, la quantità delle risorse informative e dei vocabolari controllati usati per indicizzarle, lo standard ritiene che non si possa prescindere da strumenti che rendano possibile un sistema di *cross-access*.

Due interessi conflittuali infatti guidano lo sviluppo e l'uso dei vocabolari controllati. Da una parte, le comunità di indicizzatori sviluppano vocabolari controllati specifici per i loro ambiti concettuali e terminologici; dall'altra, gli utenti desiderano invece effettuare un'unica ricerca per trovare tutte le risorse, indifferentemente tramite l'utilizzo di uno o più vocabolari controllati, anche se coprono diversi ambiti concettuali.

L'interoperabilità è dunque l'obiettivo non del futuro ma del presente, poiché le varie tipologie di utenza hanno necessità diverse. In un mondo ideale del recupero dell'informazione, gli utenti vorrebbero appunto una metaricerca che permetta loro di trovare tutte le risorse desiderate utilizzando solo il vocabolario controllato preferito e gli indicizzatori desidererebbero invece poter indicizzare le risorse di un ambito usando il vocabolario controllato di un altro ambito; i produttori e distributori di *content objects* si augurerebbero di poter unire insieme due o più database, indicizzati però con diversi vocabolari controllati, mentre le comunità di indicizzatori spererebbero di poter unire insieme due o più vocabolari controllati realizzandone uno nuovo che racchiuda in sé tutti i concetti ed i termini di quelli da cui sono originati.

A complicare questo quadro si aggiunge la presenza di comunità di utenti multilingue, che vorrebbero effettuare le ricerche in più di una lingua, per poter trovare

¹⁵ La sezione 10 "Interoperability" dell'ANSI/NISO Z39.19-2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*, p. 82-89.

dei *content object* che sono stati a loro volta indicizzati in più di una lingua. Pur non volendo sovrapporsi o sostituirsi agli standard esistenti per i vocabolari multilingue¹⁶, l'ANSI/NISO Z39.19-2005 offre comunque alcuni suggerimenti sull'interoperabilità tra vocabolari controllati in lingue diverse.

Naturalmente l'interoperabilità a tutti i livelli è condizionata da vari fattori e, tra quelli che lo standard individua, i più rilevanti sono:

- la presenza di argomenti simili in ambiti diversi;
- vocabolari controllati diversi, che possono anche avere vari livelli di specificità o granularità, usati per indicizzare contenuti informativi di ambiti simili;
- le strategie di ricerca previste nei database rispetto a quelle realmente compiute.

La sezione sull'interoperabilità si occupa tra le altre cose di fornire indicazioni su:

- la ricerca, prendendo in considerazione i risultati degli studi sulle strategie di ricerca degli utenti e gli stili di apprendimento cognitivo nel contesto interattivo;
- l'unione di due o più database;
- l'unione di due o più vocabolari controllati.

Lo standard stesso manifesta la consapevolezza, nonostante il considerevole lavoro sia a livello teorico che pratico, di essere innovativo ma non troppo approfondito su questo tema, poiché «all are still evolving»¹⁷.

Alcuni metodi proposti per attuare l'interoperabilità sono: derivazione da un modello già esistente, traduzione/adattamento, commutazione (*switching*), mappatura intellettuale (*direct mapping*) e mappatura automatizzata (*co-occurrence mapping*). Questi ed altri sistemi per ottenere l'interoperabilità tra linguaggi di indicizzazione possono essere comunque approfonditi con la lettura dell'appendice D¹⁸.

Lo sguardo d'insieme offerto sulla nuova edizione dell'ANSI/NISO Z39.19, lo standard più richiesto per il *download* dal sito della National Information Standards Organization, ha cercato di evidenziare, oltre alle maggiori novità rispetto alle precedenti versioni, quella che è stata l'attenzione costante del Thesaurus Development Team per rendere lo standard adeguato ai tempi e alle nuove tecnologie.

¹⁶ ISO 5964:1985 *Guide to establishment and development of multilingual thesauri*. Geneva : International Organization for Standardization, 1985 e BS 6723-1985. *British standard guide to establishment and development of multilingual thesauri*. London : British Standard Institution, 1985.

¹⁷ ANSI/NISO Z39.19-2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*, p. 86.

¹⁸ *Appendix D. Methods for achieving interoperability*, in ANSI/NISO Z39.19-2005, p. 142-146.

