

The Ovid Nursing Subject Thesaurus: A Comprehensive Guide

by

Bella Hass Weinberg, D.L.S.

Thesaurus Consultant

New York

Ovid Technologies

16 December 2008

Table of Contents

	Key to Abbreviations	3
I.	Purpose of the Thesaurus	4
II.	Anticipated Audience of the Thesaurus	4
III.	Scope of the Thesaurus	4
	A. Subject Scope	4
	B. Linguistic Scope	4
	C. Geographic Scope	4
IV.	Compilation Methodology	4
V.	Relationship to MeSH	5
VI.	Structure and Presentation	5
	A. Description	5
	B. Rationale for Decisions on Structure and Presentation	6
VII.	Adherence to Thesaurus Norms and Standards	6
VIII.	The Role of Scope Notes	6
IX.	Application of the Thesaurus within Nursing@Ovid	7
X.	Relationship to the Lexicon and Ovid Nursing Database	7
XI.	Limitations	7
XII.	Development Steps	7
	About Professor Bella Hass Weinberg	9

Key to Abbreviations

BT	=	Broader Term
MeSH	=	Medical Subject Headings
N@O	=	Nursing at Ovid
NISO	=	National Information Standards Organization
NST	=	Nursing Subject Thesaurus
NT	=	Narrower Term
OND	=	Ovid Nursing Database
RT	=	Related Term
UF	=	Used For

I. Purpose of the Thesaurus

A *thesaurus* - in the information science sense of the word - is a controlled vocabulary designed for *postcoordination*, that is, combination of terms at the searching stage.

The Ovid Nursing Subject Thesaurus (NST) is intended to serve as an aid to online searching, and as the basis of automatic indexing of the Ovid Nursing Database (OND).

One of the modern synonyms of thesaurus is *ontology*. At the recent conference on "Text Comparison and Digital Creativity," a historian of logic defined ontology as "knowledge which makes knowledge possible." Ovid's NST is designed to enhance the retrieval of relevant documents, and so to add to the knowledge of the nursing community.

II. Anticipated Audience of the Thesaurus

The audience of the thesaurus is anticipated to include nursing students and educators, registered nurses, allied health students and professionals, and the librarians who assist members of these groups in searching.

III. Scope of the Thesaurus

A. Subject Scope

The terms in the NST cover the full range of nursing specialties, and the majority of disciplines in allied health (veterinary medicine, for example, is excluded). As additional content is acquired for the OND, the subject scope of the NST may be expanded.

B. Linguistic Scope

The NST at present includes only English-language terms. American terms and spelling are favored, but the first edition of the NST includes cross references from British variants. Many dictionaries of nursing and allied health have been acquired for the thesaurus team, to facilitate the addition of cross references from English-language variants used in Australia and India.

C. Geographic Scope

The thesaurus is international in scope. It includes descriptors for the names of many countries as well as headings for associations of nursing and allied health professionals found in many countries.

In light of the first feature, it is possible to search a topical descriptor in combination with the name of a country in order to learn how a nursing specialty is practiced in that country, e.g., Midwifery AND Morocco.

IV. Compilation Methodology

The first edition of the NST was compiled after consultation of available controlled vocabularies in relevant disciplines as well as special classification schemes for nursing.

Many older controlled vocabularies have numerous inverted terms. These terms were flipped to direct order to facilitate matches with natural-language search terms. In some

cases, Google was searched to verify that the direct form of a term is used in natural language. In such cases, Google is recorded in the Source field of the descriptor record. Selected headings from nursing classifications were also modified to match natural-language expressions.

All terms in the draft NST were searched against a machine-readable file of Library of Congress Subject Headings, which provided Source notes for many descriptors. Medical Subject Headings were searched as well (see following section).

V. Relationship to MeSH

The NST is based largely on Medical Subject Headings (MeSH): two-thirds of the NST's descriptors derive from MeSH, and the Source field notes "MeSH-08" for these terms. Inverted terms in the latter vocabulary have generally been flipped to direct form in the NST. In such cases, the Source field notes "MeSH-08 (inverted)." MeSH definitions have been extracted from a machine-readable version of the vocabulary, and stored in the Scope Note field of the NST. More than 3,000 descriptor records, i.e., a third of the terms in the NST, currently include MeSH definitions.

The NST does not include the check tags or subheadings (qualifiers) of MeSH because it is difficult for automatic indexing software to infer such concepts from free text. While subheadings are a powerful feature of MeSH, they are not found in true thesauri, as the combination of headings and subheadings is a characteristic of precoordinate vocabularies. Some concepts represented by MeSH subheadings are precoordinated in NST descriptors, e.g., "Foot Injuries," while others can be searched in a postcoordinate fashion, e.g., "Beta Carotene AND Deficiency."

The NST does not include MeSH tree numbers, although its hierarchical structure closely mirrors that of MeSH, and an Explode capability is part of the search interface. The NST's structure is described in the following section.

VI. Structure and Presentation

A. Description

The NST is a flat-format thesaurus. This means that (unlike MeSH) it shows only one level of hierarchy for each descriptor - one level of broader terms, and one level of narrower terms.

The NST includes Used For (UF) references - a record of the synonyms from which USE references are provided. In MeSH, the command "see" is the equivalent of USE, and X is the equivalent of UF.

The NST also provides Related Term (RT) references. "See related" is the rough equivalent of RT in MeSH, as this type of cross reference is not always provided in both directions in MeSH. In the NST, RT references are always symmetrical.

The relationship indicators (BT, NT, RT, UF) are spelled out in the display of the NST, which renders it user-friendly. Non-preferred terms are presented in italics, and descriptors in Roman type. All terms are hyperlinked, which makes it easy to move around the thesaurus. The number of Hits for each descriptor is displayed, to tell the user how many documents have been indexed with that term.

The software on which the NST is compiled - Thesaurus Master - provides a full hierarchical display, with top terms arranged alphabetically, but the Ovid platform cannot provide two views of a thesaurus at present. Plans to modify the presentation to include a multilevel display of the hierarchy are described in the final section of this paper.

Within the software used for the NST, Scope Note and Source are two separate fields. In the thesaurus linked to the Ovid Nursing Database, they are combined in one field.

B. Rationale for Decisions on Structure and Presentation

The flat-format structure was selected for its simplicity and flexibility. It is a simpler structure than a tree structure, as it does not require the user to toggle between two displays of a vocabulary; all the basic semantic relationships are displayed in a single view of the descriptor record.

The flat format was also chosen for its flexibility - from the perspective of the thesaurus designer. A tree structure requires careful planning and a clear definition of the scope of a vocabulary. The scope of the OND is expected to expand as new content is added, and the NST will change accordingly. If notation were assigned to the hierarchy, it would quickly break down when a new branch of allied health is added, for example. A flat-format thesaurus is infinitely hospitable to new terms and even entire hierarchies.

Given the current limitation of the Ovid platform to display only one view of the thesaurus, it was felt that the alphabetical view is more important than a hierarchical one.

VII. Adherence to Thesaurus Norms and Standards

The consultant for Ovid's NST chaired the committee of the National Information Standards Organization (NISO) that developed the *American National Standard for the Construction, Format, and Management of Monolingual Thesauri* (ANSI/NISO Z39.19-1993), which was published in 1994 and reaffirmed in 1998. This standard was a NISO best-seller and influenced many developers of thesaurus software.

The NST complies with the NISO standard through its provision of descriptors in direct order and the inclusion of the standard term relationships.

VIII. The Role of Scope Notes

In many controlled vocabularies, such as Library of Congress Subject Headings, scope notes are provided only for descriptors whose scope is unclear, and for cases in which indexers and searchers have to be instructed on the differentiation of descriptors.

In the NST, the goal is to provide scope notes for the vast majority of descriptors. At present, more than 3,000 terms have definitions taken from MeSH. Licenses are being negotiated with the publishers of machine-readable medical dictionaries, nursing thesauri, and nursing classifications to extract their definitions in order to populate the Scope Note field of the NST. We believe that the inclusion of definitions for the majority of terms in the NST will be a substantial aid to end-users as well as search intermediaries.

IX. Application of the Thesaurus within Nursing@Ovid

Envisioned as the centerpiece of Nursing@Ovid, the NST is accessible from the results presentation in Basic Search as well as by clicking on “Search Tools” within the OND and selecting the radio button “Thesaurus”. The icon *i* (for “information”) indicates that the descriptor record includes a scope note—containing a definition and/or a source note. The display shows the number of postings (hits) in OND for each descriptor.

A Permuted Index to the thesaurus is also available; for any single word, this index shows all the NST terms in which it occurs, whether preferred (descriptor) or non-preferred (cross reference).

X. Relationship to the Lexicon and the Ovid Nursing Database

Ovid’s Lexicon is a natural-language processing feature that is activated when the user selects Basic Search. The Lexicon is currently separate from the NST, but there are plans to integrate the two, as they have similar purposes: enhancing the user’s search by adding or suggesting related terms.

XI. Limitations

Some limitations of the NST have been mentioned in the preceding sections. Some of the limitations emanate from the use of automatic indexing software; others from the current limitations of the Ovid platform and its search interface. The rule base for automatic indexing is revised continuously on the basis of USE references added to the thesaurus. Change requests have been submitted to Ovid programmers to enhance and simplify the interface.

XII. Development Steps

Many of the preceding sections mention plans to expand the NST and enhance its interface.

As with any thesaurus, NST is a work-in-progress. Already work is nearly complete for the next release, with several specific enhancements to the thesaurus itself and the way that OvidSP uses it:

- Addition of slang terms and helpful synonyms
- Addition of standard acronyms and initialisms
- Adjustment of indexing rules for acronyms that are also common words
- Improving rules for terms containing apostrophes or commas
- Improving rules for handling selected frequent terms during indexing

As content is added to the OND, new descriptors and cross references will be added to the NST. As licenses are approved, definitions will be added to the Scope Note field, and new nursing classifications will be incorporated into the NST.

The addition of descriptors will entail the establishment of semantic relationships with terms included previously. Thus, the thesaurus will undergo continuous revision, although new versions will be mounted only periodically by Ovid's Content team.

As for presentation, there are plans to provide a multilevel format within the alphabetic display, i.e., to show all levels of broader and narrower terms for each descriptor. This is not the exact equivalent of a tree structure, but a good compromise as long as the Ovid platform cannot provide two displays of a thesaurus.

As Ovid's thesaurus consultant, I continue to submit proposals for simplification of the search interface of the NST.

In sum, the NST loaded with the initial release of Nursing@Ovid complies with thesaurus standards, but many enhancements to it are envisioned. Suggestions from the user community are welcome.

About Professor Bella Hass Weinberg

Bella Hass Weinberg received her doctorate in library-information science from Columbia University in 1981. The subject of her dissertation was "Word Frequency and Automatic Indexing." She joined the faculty of the Division of Library and Information Science at St. John's University in 1982 and was promoted to Full Professor in 1992. Dr. Weinberg teaches in the area of organization of information, including courses on cataloging, indexing, classification, thesaurus design, and information architecture. She has presented workshops and numerous conference papers on these topics internationally. Professor Weinberg also teaches a graduate course on Medical Information and Librarianship on the basis of her experience in medical libraries. She has published papers on Medical Subject Headings and has consulted on pharmaceutical vocabularies.

Professor Weinberg served as President of the American Society of Indexers (1988-89), and received its Hines Award for distinguished service to the profession. She chaired the committee of the National Information Standards Organization (NISO) that produced the widely-used standard for thesaurus construction (1993; reaffirmed 1998). Bella Hass Weinberg consults on large-scale indexing projects, thesauri, and search interfaces in a variety of domains. She has published extensively on many aspects of the organization of information. Her reviews of books in this field were collected in an anthology entitled *Can You Recommend a Good Book on Indexing?* (Information Today, 1998).