

Barbara H. Kwaśnik
Syracuse University

**Revisiting the Preserved Context Index System (PRECIS):
The Bridge between Hierarchically Structured Thesauri and Facetted Classifications**

This presentation will address the difficult task of representing complex concepts in a text in a way that reflects their contextual meaning. The preservation of context enables the disambiguation of a term's possible multiple senses, and also shows how the term is being used. In developing these ideas we revisit an indexing system called PRECIS, which was developed by Derek Austin in the early 1970s for subject indexing for the *British National Bibliography*, and subsequently developed by him with the assistance of Mary Dykstra into a adaptable method of linking both the semantics and syntax of indexing terms.

In the traditional approach to indexing there are two primary methods of deriving the indexing vocabulary:

1. terms are chosen from the text itself – typical of back-of-the-book indexes, or
2. terms are chosen from a controlled vocabulary, such as a thesaurus or a subject headings list.

In the first instance, besides the important “content bearing” words, the indexer supplies syntax in the form of subentries. For example:

librarians
 education of
 job satisfaction of
 poor pay for

This sort of contextual information is usually limited to indexes created one work, that is, the index that appears at the end of the work. In the case of shared or continuing indexing, such as that which occurs in dynamic collections of works and serially produced works, however, the indexers typically must choose terms from a controlled vocabulary. Many traditional controlled vocabularies comprise subject terms that are structured in the form of hierarchies. For example:

librarians
 academic librarians
 community college librarians
 university librarians
 school library media specialists
 special librarians
 law librarians
 music librarians

The strength of such representations is that they offer semantic context. By knowing that music librarians are a kind of special librarian we know more about the meaning of the term than if it were isolated. The obvious issues with such a structure are that:

- typically only one aspect of meaning is revealed; the “context” is that of increasing or decreasing specificity along one set of discriminatory dimensions. But, what if a librarian is a music librarian at a university (and therefore, an academic librarian as well)?
- such vocabularies include only nouns, or noun phrases of nouns and adjectival modifiers. Even actions are transformed into nouns: “running,” “management,” “searching.”
- associations among terms can only *imply* syntactic relationships. E.g., “pasteurization” and “milk.”

Facetted analysis, developed as a notion by Ranganathan, attempted to remedy the limitations of one-dimensionality by enabling the representation of an object from a number of perspectives, such as time, geographical association,

processes, materials, and so on. For instance, in using the *Art & Architecture Thesaurus* to represent artefacts, we can construct a descriptive string such as “embroidered, felt, 12-Century, Celtic slippers,” in which the terms and their citation order offer a rich dimensionality, and can then be permuted in order to enable searching on any one of the component terms.

Both faceted and traditional hierarchically structured indexing vocabularies rely on the subject terms (typically nouns and their adjectives) to convey the meaning of the text. PRECIS, on the other hand, also captures the subject terms, but adds one more dimension, and that is the syntax. For example: a document about the storage of powdered milk would generate the following entries (Dykstra, 1987):

Milk Powdered milk. Storage
Powdered milk Storage
Storage. Powdered milk.

An indexer using PRECIS asks the following questions of the text:

Did anything happen?
If yes, to whom or what did it happen?
Who or what did it?
Where did it happen?

In enabling a syntactic analysis of the text, the indexer can then represent not only the subjects, but also the story behind the subject. In doing so, it bridges the context-enhancing strategies of subject hierarchies and faceted strings, and returns some of the richness provided by a good back-of-the-book index that is inevitably lost in controlled-vocabulary indexing.

The rules and guidelines for PRECIS indexing are quite complex, however, requiring a sophisticated understanding of the logic of syntax, as well as the linguistic mechanics that reveal meaning at this level. In addition, this is a manual system, in the sense that a person must assign the topic, as well as the permutations, while a computer can assist with only the housekeeping chores of alphabetizing and maintaining consistency. Perhaps it is time to pull PRECIS into the era of natural language processing where texts can be parsed for syntax, and then offered to the indexer for further intellectual manipulation.