

2.3 Indexing Process

- _ When an organization with multiple indexers decides to create a public or private index, some procedural decisions assist the indexers and end users on how to create the index terms.
 - _ scope of the indexing
 - _ linking index terms

Scope of Indexing

- _ defines the level of detail that the subject index will contain.
- _ This is based upon usage scenarios of the end users.
- _ When performed manually, the process of determining the bibliographic terms that represent the concepts in an item is difficult.
- _ Problems arise from interaction of two sources: the author and the indexer.
- _ The vocabulary domain of the author may be different than that of the indexer, causing the indexer to misinterpret the importance.

- _ The indexer is not an expert on all areas and has different levels of knowledge in the different areas being presented in the item.
- _ This results in different quality levels of indexing.
- _ The indexer must determine when to stop the indexing process.
- _ There are two factors involved in deciding on what level to index the concepts in an item:
 - _ Exhaustivity
 - _ Specificity.
- _ Exhaustivity of indexing
 - _ is the extent to which the different concepts in the item are indexed.
 - _ For example, if two sentences of a 10-page item on microprocessors discusses on-board caches, should this concept be indexed?
- _ Specificity
 - _ relates to the preciseness of the index terms used in indexing.
 - _ For example, whether the term “processor” or “microcomputer” or “Pentium” should be used in the index of an item.

- _ using general index terms yields low exhaustivity and specificity.
- _ This approach requires a minimal number of index terms per item.
- _ reduces the cost of generating the index.

- _ Another decision on indexing is what portions of an item should be indexed.
- _ The simplest case is to limit the indexing to the Title or Title and Abstract zones.

- _ Weighting
 - _ Weighting of index terms is not common in manual indexing systems.
 - _ Weighting is the process of assigning an importance to an index term's use in an item.
 - _ The weight should represent the degree to which the concept associated with the index term.
 - _ The process of assigning weights adds additional overhead on the indexer and requires a more complex data structure to store the weights.

Pre coordination and Linkages

- _ Another decision on the indexing process is whether linkages are available between index terms for an item.
- _ Linkages are used to correlate related attributes associated with concepts discussed in an item.
- _ *Precoordination*
 - _ process of creating term linkages at index creation time.
 - _ When index terms are not coordinated at index time, the coordination occurs at search time.
- _ *Post coordination*
 - _ coordinating terms after (post) the indexing process.
 - _ implemented by “AND” ing index terms together, which only finds indexes that have all of the search terms.
- _ Figure 3.2 shows the different types of linkages.
- _ It assumes that an item discusses the drilling of oil wells in Mexico by CITGO and the introduction of oil refineries in Peru by the U.S.

- _ When the linked capability is added, the system does not erroneously relate Peru and Mexico since they are not in the same set of linked items.
- _ It still does not have the ability to discriminate between which country is introducing oil refineries into the other country.
- _ Introducing roles in the last two examples of Figure 3.2 removes this ambiguity.

<u>INDEX TERMS</u>	<u>Methodology</u>
oil, wells, Mexico, CITGO, refineries, Peru, BP, drilling	No linking of terms
(oil wells, Mexico, drilling, CITGO)	linked (Precoordination)
(U.S.,oil refineries, Peru, introduction)	
(CITGO, drill, oil wells, Mexico) (U.S., introduction, oil refineries, Peru)	linked (Precoordination) with position indicating role
(SUBJECT: CITGO; ACTION: drilling; OBJECT: oil,wells MODIFIER: in Mexico)	linked (Pre-coordination) with modifier indicating role
(SUBJECT:U.S.; ACTION:introduces; OBJECT: oil refineries; MODIFIER: in Peru)	

- _ Positional roles treat the data as a vector allowing only one value per position.
- _ Thus if the example is expanded so that the U.S. was introducing oil refineries in Peru, Bolivia and Argentina, then the positional role technique would require three entries, where the only difference would be in the value in the “affected country” position.
- _ When modifiers are used, only one entry would be required and all three countries would be listed with three “MODIFIER”s.