

# Indexing and Abstracting

## Lecture 05 -- Indexing Methods and Procedures

Kuang-hua Chen  
Department of Library and Information  
National Taiwan University  
khchen@ntu.edu.tw

# Indexing

- Indexing is the process of identifying information in a knowledge record and organizing the pointers to that information into a searchable file
- The outcome of the indexing process is an index that indicates topics and possible uses for the documents and points to the location of the information

# Trade-offs

- Controlled vocabulary vs natural language
- Recall vs precision
- Specific indexing vs generic indexing
- Conceptual indexing vs keyword indexing
  
- Indented format vs run-in format
- Alphabetical display vs classified display
- Word by word vs character by character

# Aboutness

- The aboutness of a documents is not limited to the explicit keywords from the text
- Aboutness is much more than just coverage of surface content
- A major reason that indexing fails is because the indexer superficially dealt with the aboutness issue

## Subtle Aboutness

- Language is a multilayer carrier of information
- A sentence can convey different messages for different people
- Aboutness can have different interpretations according to the background and orientation of the writer and the reader, and the indexer
- Indexers have to tell major topics from minor topics in documents

## Indexing Process

- A combination of formal rules, common sense, and talent
- Indexing can be done by humans, by a computer, or by a combination of humans and computer

## Steps of Indexing

### Subject Analysis

- Decide which topics in the item are relevant to the potential user of the document
- Decide which topics truly capture the content of the document

### Subject Translation


- Determine terms that come as close as possible to the terminology used in the document
- Decide on index terms and the specificity of those terms

### Index compilation

- Group references to information that is scattered in the text of the document
- Combine headings and subheadings into related multilevel headings
- Direct the user seeking information under terms not used to those that are being used by means of *see* references and to related terms with *see also* references
- Arrange the index into a systematic presentation

## What to index

- Policies (organizational, institutional)
  - The policies and purposes of a large general indexing service will be geared to a large user group with a broad subject interest
  - In a special library or a narrowly defined information center the users will have a distinct, more specific type of information need
- Subjective value judgment (Personal)
- The quality of an index can be judged by what is left out and by what is included



## First of all for indexing one piece of work

- Recording of bibliographic data
- Content Analysis
  - Title
  - Abstract
  - Text
  - Reference



## Bibliographic Data

- An important pointer for users to reference this indexed item
- Some conventions or rules should be followed
  - Name format
  - Abbreviation usage
  - ...



## Content Analysis

- Title
- Abstract
- Text
- Reference



## Title

- Content-bearing
- But too concise and general
- Sometimes, too vague
- Even worst, not related



## Abstracts

- Good abstracts can be indicators of subject content
- Most words in a good abstract will convey subject content
- Abstracts, like titles, can be badly written and misleading



## Text

- Introduction
  - What is going to be said or done
- Conclusion
  - What has been said or done
- Section headings
- First sentence and last sentence of a paragraph
- Historical and theoretical background
- Methodology
- Charts, diagram, graphs, photographs, tables



## Reference

- Good indicators to reflect the subject content
- Citation indexing
- Titles in reference list



## Key Points

- Subject determination
  - “The mosquitoes attack with the ferocity of a tiger”
  - “The queen looked at me with her mosquitoes eyes”
  - Major ideas are repeated and minor ideas are only mentioned?
- Locator
- Term selection
- Entry Points



## Depth of Indexing

- Depth of indexing is the degree to which a topic is represented in detail
- Exhaustivity
- Specificity



## Exhaustivity

- Possible terms have been exhausted
- It seems that many index terms will be assigned
- The number of index terms reflects the *exhaustivity*



## Exhaustivity (Continued)

- The more exhaustive the item is indexed, the more likely it will be discovered because of the wider range of subject terms
- The trade-off is that the document may not be specifically pertinent to user's need
- The degree of exhaustivity depends on the policy of organization, money, time and the need of your users



## Specificity

- The preciseness with which we describe a document is another dimension in choosing descriptors
- The more specific the term, the more precise the results
- If the terms used are precise, we could say the indexing is specificity
- A very specific indexing language will have a large vocabulary with more potential descriptors

## Specificity (Continued)

- The problem of specificity begins at the designing of indexing language
- We have to considerately select vocabulary and design thesaurus
  
- Use the terms which the authors use

## Trade-offs

- User-oriented
  - For general users, broader-term approach with more exhaustivity
  - For specialized user groups, narrower terms with more specificity

## Display of Indexes

- Typographic conventions
  - Punctuation
  - Type size
  - Font
- Main entries in heavy print (capital letters)
- Subheadings are in light print (smaller letters)
- *See* references are often italicized
- Indentation for distinction of main headings and subheadings

## Indented VS Run-in Style Format

- **Indented Format:** (縮排)
  - grammar
    - author's preferences, 333, 336, 339, 362
    - computerized checking, 337
    - as cultural product, 8-9, 338, 339
    - handbooks and usage guides, 61-62, 336-37
- **Run-in Format:** (接排)
  - **When space is the concern.**
    - grammar: author's preferences, 333, 336, 339, 362; computerized checking, 337; as cultural product, 8-9, 338, 339; handbooks and usage guides, 61-62, 336-37



## Indented (cross reference at bottom)

dogs

cocker spaniels, 55


Dalmatians, 33

English setters, 66

golden retrievers, 63

Gordon setters, 39

*See also* American Kennel Club



## Indented (cross reference at top)

dogs. *See also* American Kennel Club

cocker spaniels, 55

Dalmatians, 33

English setters, 66

golden retrievers, 63

Gordon setters, 39



## Run-in (cross reference at bottom)

dogs: cocker spaniels, 55; Dalmatians, 33;  
English setters, 66; golden retrievers,  
63; Gordon setters, 39; *See also*  
American Kennel Club



## Run-in (cross reference at top)

dogs (*see also* American Kennel  
Club): cocker spaniels, 55; Dalmatians,  
33; English setters, 66; golden  
retrievers, 63; Gordon setters, 39



## Alphabetization

### ■ Letter-by-Letter

A blood group  
ABO blood group  
A factor  
allyl alcohol  
allylcysteine  
allyl sulfide  
atherosclerosis

...

endings  
endogenous  
end piece  
end zone

### ■ Word-by-Word

A blood group  
A factor  
ABO blood group  
allyl alcohol  
allyl sulfide  
allylcysteine  
atherosclerosis

...

end piece  
end zone  
endings  
endogenou