INDEXING*

Audrey G. Anday

*paper presented during the Basic Library Management: Series 3 on 19 May 2009 during the seminar-workshop organized by the Cavite Librarians Association, Inc (CLASSI) in Trece Martires City, Cavite.

Backgrounder
**INFORMATION ANALYSIS**

⇒ Important requisite not only for storage but for searching
⇒ Determination of known attributes or properties or characteristics in an unknown entry like document, question, material etc

---

**Techniques for Analysis**

⇒ Annotation  
⇒ Abstracting  
⇒ Extracting  
⇒ Indexing  
⇒ Classifying

• These are all for identifying the thought contents. The last two are normally used for retrieval purposes.
• **Annotation**
  - brief comment or explanation about a document or its content

• **Abstracting**
  - preparation of an abbreviated accurate representation of the content of a document, without added interpretation or criticism and without distinction as who wrote the abstract

• **Extracting**
  - preparation of one or more portion of a document selected to represent as a whole

• **Summary**
  - brief statement within the document (usually at the end) of its salient findings and conclusions
Definition of Terms

- **Index** - an indicator of the position of the required information

- **Indexing** - refers to the activity of assigning terms from an indexing language to documents or parts of documents for the purpose of information retrieval

- Similar to classifying and subject cataloging in that all terms describe activities involving the assignment of subject identifiers (subject headings, descriptors, index terms) for purpose of information retrieval
Aim of Indexing

• To answer the questions a user might ask about the purpose of a document and how it can help him.
• These can be categorized under the following headings (Menou, 1982):
  • Subject
  • Approach
  • Manner
  • Time

Purposes of Indexing

• Helps discover/rediscover items of information in documents which are deemed relevant to user information needs
• Provides a systematic regular and rapid access to the body of relevant literature
Types of indexing

1. **Free word** - uses words as used by author in title, summary or text of a document. It requires no judgment or skill. Examples are:
   a) Concordances - using every word
   b) KWIC, KWAC, KWOC - using significant words in the title

---

**KWIC - Keyword-in-Context**

- Originally introduced by Andrea Crestadoro as long ago as 1964, under the name ‘Keyword in titles’, for a catalogue of Manchester Public Libraries.
- Nearly a century later it was developed by H P Luhn for computer manipulation and applied to American chemical Society’s current awareness publication, *Chemical titles*, generated on an IBM computer, wherein it was a simple matter to store a list (called a stop list) of the common function words in memory and inhibits the generation of any entries alphabetized on those words (Luhn 1960). It was widely adopted by various information services in the years following 1958.
- As early as 1962, 30 applications of the KWIC concept were in existence, including such examples as the KWIC Index to Neurochemistry and the KWIC Index to the science Abstract of China.
Example: KWIC Index

- Let’s suppose we had the following titles:
  - *Blue-eyed Cats in Texas*
  - *The Cat and the Fiddle*
  - *Dogs and Cats and Their Diseases*
  - *The Cat and the Economy*

- The KWIC index is an alphabetical list, ordered on each-conveying word in the title. Thus, taking the above titles, a KWIC index might contain the following entries:

<table>
<thead>
<tr>
<th>In Texas,</th>
<th>Blue-eyed Cats…………………………………23</th>
</tr>
</thead>
<tbody>
<tr>
<td>The</td>
<td><em>Cat</em> and the Economy………………………..12</td>
</tr>
<tr>
<td>The</td>
<td><em>Cat</em> and the Fiddle…………………………17</td>
</tr>
<tr>
<td>Dogs and Blue-eyed</td>
<td><em>Cats</em> and Their Diseases……………………3</td>
</tr>
<tr>
<td>And Their</td>
<td><em>Diseases, Dos and Cats</em>……………………3</td>
</tr>
<tr>
<td>Their Diseases</td>
<td><em>Economy, The Cat</em>……………………………12</td>
</tr>
<tr>
<td></td>
<td><em>Fiddle, The Cat</em>……………………………..17</td>
</tr>
<tr>
<td></td>
<td><em>Texas, Blue-eyed Cats</em>……………………23</td>
</tr>
</tbody>
</table>
KWOC – Keyword-out-of-Context

• Elizabeth A1-Hazzam has described a KWOC index introduced in 1972 to provide multiple access points to the map collection of Arizona State University’s Hayden Library.

• This index is based not on titles but on thesaurus, which includes some Library of Congress subject headings but consist mainly terms from class G of the Library of Congress Classification and from information on the face of maps in the collection.

• KWOC index doesn’t rotate the title, but lifts out the keyword of interest and lists it separately to the side.

EXAMPLE:

<table>
<thead>
<tr>
<th>Blue-eyed</th>
<th>Blue-eyed Cats in Texas.................................23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cat</td>
<td>The Cat and the Economy...............................12</td>
</tr>
<tr>
<td>Cat</td>
<td>The Cat and the Fiddle................................17</td>
</tr>
<tr>
<td>Cats</td>
<td>Dogs and Cats and Their Diseases.....................3</td>
</tr>
<tr>
<td>Cats</td>
<td>Blue-eyed Cats in Texas...............................23</td>
</tr>
<tr>
<td>Diseases</td>
<td>Dogs and Cats and Their Diseases.....................3</td>
</tr>
<tr>
<td>Dogs</td>
<td>Dogs and Cats and Their Diseases.....................3</td>
</tr>
<tr>
<td>Eco</td>
<td>The Cat and the Economy...............................12</td>
</tr>
<tr>
<td></td>
<td>The Cat and the Fiddle................................17</td>
</tr>
<tr>
<td></td>
<td>Blue-eyed Cats in Texas...............................23</td>
</tr>
</tbody>
</table>
**KWAC – Keyword-Augmented-in-Context**

- It is another variation of the KWIC index introduced in an attempt to overcome some of the difficulties and questions of the KWIC and KWOC index.
- This involves the addition of index words to the title of the document by a subject specialist. The thought that this augmentation of the title would permit better penetration of the subject matter of the document than is often possible by using the author’s title alone. In one system, the augmentation is called a “notation of content”. (Kent)

**Keyword – And – Context**

- Indexes where the keyword appears both as heading and again in the title. However, the distinction between KWOC and KWAC indexes is rarely made and KWOC is the *generic term* that is usually used. (Rowley)
2. **Uniterm indexing** - using single common English words, serial number and other symbols

3. **Controlled indexing** - implies a careful selection of terminology or vocabulary with a view to avoid scattering. Controlled are: subjects chosen; number of aspects chosen; and, language used to express the analytics.
Controls are exercised by:
   a) Authority list like classification schedule, subject heading list etc
   b) Formulating the rules for indexing
   c) Developing and open-ended dictionary or thesaurus

How else to effectively search?

• Truncation
• When searching keyword phrases, use positional operators to place words within same sentence (FT!)
• WITH/SAME is preferable, personally
• Disguising digits
• Natural language can still search any word or phrase in the dictionary file
The Best Method?

- Combination of freetext and controlled
- Helps solve homograph problems
- Indexers and human errors
- Authors terminology

Thesaurus and Relationships

- T = Term
- BT = Broad Term

\[ T \rightarrow \text{Term Word or phrase used to denote concept} \]
\[ BT \rightarrow \text{Broad Term Cross reference in main part of thesaurus indicating descriptor to which the reference is superordinated} \]
Thesaurus (cont’d)

- NT = Narrow Term A cross reference in the main part of thesaurus indicating that the DE to which the reference is directed is subordinated to the DE from which the reference is made.

- UF = Used For A synonym; a term having a different form but exactly the same meaning as another term.

- RT = Related Term A cross reference in the main part of the thesaurus indicating that two DE are somewhat related.
Thesaurus (cont’d)

• SN = Scope NoteDefinition; how the term is used in the thesaurus

By looking in the LCSH you can often avoid frustration, save time, and locate useful terms that might not have occurred to you.

On the left is a sample from the LCSH.

• Headings in boldface may be used for subject searching. For example, “Models and modelmaking” may be used as a subject heading.

• May Subd Geo. Place names may follow the heading.

• Call number given (TT154) indicates where most books on this subject are found. Rather than search further, you might just go directly to this call number.

• UF (Used For) These are unauthorized headings. Don’t use these headings. Refer to related subject headings.

• BT (Broader Topic) refers to more general subject headings. These headings would be useful if you need to broaden your topic.

• RT (Related Topic). These terms can provide ideas of other topics to investigate.

• SA (See Also) refers you to other ways of looking up the same topic.

• NT (Narrower Topic) refers to more specific headings than the boldface heading.

• – (a dash) refers to a subdivision of the boldface subject heading. These listed subdivisions are useful for narrowing a topic.
Actual Indexing

Steps in Indexing

• Document analysis
• Subject analysis
• Evaluation and representation of subjects
• Description
• Review
Document analysis

- The process of familiarization with the document and its contents by: reading the title; reading the abstract, if one is provided; and, scanning the entire article and noting:
  a) the source of the document
  b) its purpose
  c) its target audience
  d) its authors
  e) any appended materials
  f) the title and how it fits the content

Subject analysis/content analysis

- The process of identifying indexable matter and indexable ideas or concepts. The indexable matter includes words, names, bibliographic citations, subjects, concepts, tables of contents, etc.
Evaluation and representation of subjects

• The process of deciding whether to index and with what depth the documents is to be indexed. The index terms or subject headings are chosen from the indexing language. The decision to index will be based on inclusion criteria, e.g. relevance, permanent references value, signed article and policies on subject and period scope.

Evaluation and representation of subjects...

• Indexing depth, or the number of index terms used to represent the subjects, depends on the numbers of indexable subjects and concepts, economic and space considerations and the level or class of information. Index terms are selected from an indexing language which would have been determined during the system designing.
Evaluation and representation of subjects...

- Types of Indexing Language
  a) free or uncontrolled or natural language
  b) controlled vocabulary - formalized indexing language which provides control and consistent use of index terms at any or all phase of system operation

Description

- The process of writing down on input sheet the bibliographic entry in its prescribed format together with index terms to describe indexable matter
Review

• Reviewing the indexing accomplishments with the following questions in mind:
  a) Do the major terms represent the chief emphasis of the document?
  b) Are the major terms reflected in the abstract so that there is no question about their relevance?
  c) Do the descriptors reflect all of the important concepts in the document?

Periodicals to be Indexed (For Filipiniana collection)

• Scholarly journals of colleges & universities, organizations, scientific & learned societies
• Weekly magazines supplements of the metropolitan dailies
• Publications issued by government and private agencies
• Foreign publications which contain articles on the Philippines or about well-known Filipinos
PROCEDURE

• Use a 3X5 slip or card for easy handling and location
• Read through the articles before assigning subject heading(s). Provide cross references for subject heading(s) that cannot be used

PROCEDURE...

• On the slip, the following data are written with the proper indentions:
  a) Author of the article in an inverted form
  b) Title of the article
  c) Periodical title and underscored
  d) Volume number &/or issue number
  e) Page(s) where the article appears
  f) Date of the publication preferably---month, day and year
PROCEDURE...

- Check the article indexed with the red pencil
- Indexed periodicals are marked Indexed on the upper right hand corner
- Slips are set aside for typing
- Issues of periodicals indexed are recorded
- Statistics of issues indexed are noted down, by number of titles and issues

Example

Cifra, Herminia, Melanie Sancho.

Proposed index terms:
Free text: touch research
Controlled: massage therapy (?)
Hands-on Exercise

• Divide the participants in groups of 15. Each group assigns a Coordinator, Scribe and a Reporter.

• Choose the type of indexing you want to perform: Free text, KWIC, KWOC, KWAC and Controlled vocabulary.

• Following the steps in indexing, each member of the group should be able to accomplish at least 5 sheets of 3x5 slips = 3 articles. Jot down all pertinent details.

• If time permits make a summary of the terms you have selected and counter checked with the LCSH or Sear’s list and share your output with the rest of the participants.
Terima kasih!
Arigato gozaimasu!