INFS 324
INDEXING AND ABSTRACTING

Session 4 – SUBJECT INDEXING II

Lecturer: S. Nii Bekoe Tackie, School of Information and Communication Studies, Department of Information Studies
Contact Information: snbtackie@ug.edu.gh

UNIVERSITY OF GHANA
College of Education
School of Continuing and Distance Education
2014/2015 – 2016/2017

Session Overview

This Session is a continuation of the first session on subject indexing. In this Session, I will focus on the processes involved in arriving at the point where the indexer decides what type of indexing system to create. Thus I will take you through the indexing process and some problems encountered in the process; the searching process and the process of manual and automatic indexing.

Session Objectives

At the end of the session, the students would be :

- Conversant with the stages of the indexing process.
- Able to describe the searching process.
- Able to identify problems encountered at the translation stage of the indexing process.
- Able to describe manual and automatic indexing.

Session Outline

- Topic One: The Indexing Process
- Topic Two: Problems Encountered with Translation of Concepts
- Topic Three: The Searching Process
- Topic Four: Manual Indexing vrs. Automatic Indexing

Topic One:

THE INDEXING PROCESS

Familiarization

The creation of indexes is identified by Rowley and Hartley (2008) as a three stage process namely:

- Familiarization, Analysis and Translation stages.

Familiarization

- This is the first step in the indexing process.
- At this level, the indexer wants to have an understanding of the overall purpose of the author of the document.
- He will have that understanding by making a survey of the entire contents of the document.

Analysis

This stage requires a close examination of the subject matter of the document in order to decide which topics or concepts are important enough to be used in the index.

- In doing this, he has to look at the physical form of the document.
- Documents come in two forms namely:
  - Textual
  - Non-textual.

Analysis(Cont.)

Textual documents are print-based materials like books, journals, reports etc.

- For textual documents the key parts that needs to be examined closely includes:
  - the title
    the abstract if it is present
  - the introduction, opening phrases of chapters, sections or paragraphs
  - illustrations, tables, diagrams and their captions
  - the conclusion if there is one
  - Words or groups of words that have been underlined or italicized.

Analysis(Cont.)

Non–textual documents include

- pictures,
- paintings,
- photographs,
- sound recordings,
- CDs,
- slides,
- films,
- videos,
- microforms etc.

Some of these materials cannot be read with the naked eye.

Analysis(Cont.)

- To analyze non- textual documents not in eye – readable medium, the appropriate equipment must be used.
- When the special equipment is not available then the indexer will have to derive the concepts from the title or the synopsis of the document.

Translation of concepts

The third stage of the indexing process is a very critical one.

- This is because it is at this stage that the indexer must select terms which must agree with the concepts that have been identified at the analysis stage.
- It is at this stage that the indexing language is applied.
- This stage is critical also because in assigning labels a number of problems arise.

Topic Two:

PROBLEMS ASSOCIATED WITH THE TRANSLATIONS OF CONCEPTS

Introduction

In translating the concepts to index terms a number of problems arise.

- These include issues about:
  - synonyms,
  - homographs,
  - differing levels of specificity,
  - subjects that can only be expressed with more than
  - one word,
  - complex subjects.

Synonyms

The problem of synonyms has to do with how to treat:

- Words with similar meaning,

  For example:

  Salaries,

  Wages,

  Income

- In a general index treated as the same in meaning,
- In an index on taxation they will be treated as different terms.

Synonyms(Cont.)

Subjects with common and technical names

- For example:

  salt (Sodium Chloride),

  Water (H20)

- Pose problems because the indexer may not know which term to use.

Synonyms(Cont.)

Names that have undergone changes in usage over time are also problematic

- For example:

  - transistor, wireless, radio.

Also one has to contend with American and British usage of language

For example:

- British - aerial, American – antenna;
- British – luggage, American – baggage;
- British – railway, American railroad.

Synonyms(Cont.)

Subjects with the same stem may be recognized as synonyms

- For example:

  - nation, nationality, nationalism, nationalization.

- If the indexer is not careful and merges synonyms, it will result in a defective index.

Homographs

- Another problem is the occurrence of homographs. These are words that are spelt the same but have different meanings
- For example.
  - ‘minute’ which may mean time as in 30 minutes or small as in amount or size
  - ‘score’ as in the results of a game or as in music which is the written notes
  - ‘Crane’ as in the equipment used to lift heavy objects or the bird.

Nouns

These inflect for number

- Problem with which form to use

  - singular or

  - plural form.

The rules say that the plural form should be used for count nouns

For example:

  - pencils instead of pencil.

Subjects that can only be expressed with more than one word

Some concepts can only be expressed with more than one word

- For example:
  - War ship,
  - Pressure lamp,
  - Gas mask.
- Words like these create problems as to which word group to place them in.
- In such instances inversion may be used,
  - For example
    - ship, war;
    - lamp, pressure;
    - mask, gas.

Composite Subjects

These may have the same components or concepts.

- For example:
  - ‘assessment of lecturers by students’
  - ‘the assessment of students by lecturers’
- Same components or concepts.
- In such circumstance citation order is used to show differences in meaning.
- This involves the syntactic relationship that exists between words;
  - that is the position or occurrence of the words in the sentence may provide the meaning.

Topic Three:

THE SEARCHING PROCESS

Introduction

I am sure you would recollect that I noted that indexing and abstracting are interwoven with searching.

- Let me reiterate here that indeed indexing, abstracting and searching cannot exist in isolation of each other.
- They are linked with each other and together form the axis of any information retrieval system.
- In other words, it would be useless to create indexes and abstracting if they are not going to be used for searching. Searching, let me say, is equally important as indexing and just as indexing follows a process, so does searching.

Introduction(Cont.)

The information seeker uses the document surrogates created through indexing as the starting point for a search.

Searchers generally do not conduct searches with any particular document surrogate in mind.

Thus for a successful search to be conducted a process must be followed.

This is what Rowley and Hartley (2008) call the searching process.

Introduction(Cont.)

The searching process follows the same steps as the indexing process.

- The first stage of the process is the familiarization stage.
- This stage is followed by the analysis stage and then come
- the translation stage where the query is matched with the language of the index.

Familiarization

First step of the searching process

- It is important for the searcher to have a clear view of the objective of the search otherwise the search may end in dissatisfaction with the search results.
- Indeed later in the course you will learn that user orientation of the document and the index affects the effectiveness of an index.
- Achieved through an interview with the information seeker.
- The aim of the interview is to get a clear subject profile as well as other characteristics of the information that is needed.

Familiarization (Cont.)

These may include:

- language constraints,
- time within which the information may be considered still relevant, and
- the intellectual level of the information seeker.

Analysis

The Analysis stage is the Step 2 of the Searching Process

Once the objective of the search has been clearly established, the next stage in the searching process is to analyze the concepts in the search query (reference question, information need).

- Sometimes the query may be a straight forward one which may require a search in a printed index.
- In such circumstance the concepts may be established in the searcher's mind.
- For example, if information is needed on "database management" and this matches a search term in a database then the search profile will simply involve the term "database management".
- On the other hand, a query may have a number of interacting concepts and other parameters making the search profile much more difficult.

Analysis(Cont.)

In such instances it will be necessary to write down the concepts.

- For example, if the information required is on
- "using mercury to extract gold in galamsey operations in Ghana" and the information seeker is interested in only

  reports, books or periodical articles that give a review of the subject since 1990 and are written only in English,
- The search profile will consist of a series of search terms representing the subject and other characteristics of the search requirements.
- These together will indicate the scope and nature of the search.

Translation

This is the third and final Step in the searching process.

- Means to match the concepts identified in the analysis of the search query with

  - the thesaurus,
  - classification scheme or
  - list of subject headings (indexing language)

that has been used to index documents in the collection to be searched.

- If the terms march, then good results may be produced.
- Successful translation depends a lot on the support that has been provided in the system being searched.

Translation(Cont.)

- In a printed index, there may be guidance on indexing practices.
- This will help the searcher in the search decisions that are made.
- In a computer-based system, there may be a number of supporting facilities to help in searching different parts of the record.
- These may include:
  - assigned subject terms,
  - abstracts and
  - natural language text of the document.

Topic Four:

MANUAL INDEXING/AUTOMATIC INDEXING

Manual Indexing

The end result of subject analysis is the creation of an indexing system. The indexing system may be manually created or it may be automatic.

- **Manual Indexing**
- The basic steps of the manual subject indexing **process** are:
  - Analysis of subject
  - Identification of keywords
  - Standardization of keywords

Manual Indexing(Cont.)

Choice of an indexing system

- if the chosen system is a post-coordinate one then
  - preparation of entries under each term with reference to the document identification number
  - preparation of reference entries

- if the chosen system is a pre-coordinate one then
  - preparation of an entry (main entry) using all the keywords organized in a way prescribed by the system
  - preparation of index entries by using each significant term as an entry element and the full entry (main entry) as the context, or by rotation/permutation of the significant terms in the main entry according to the rules prescribed by the system chosen
  - preparation of reference entries
  - filing of entries

Automatic Indexing

- Salton (1989)
  - Assignment of content identifiers with the aid of modern computing equipment.
- Borko and Bernier (1978)
  - subject of a document can be derived by a mechanical analysis of the words in a document and by their arrangement in a text.
- In fact, all attempts at automatic indexing depend in one way or the other on the original document texts, or document surrogates.
- The words occurring in each document are listed and certain statistical measurements are made,
  - like word frequency calculation,
  - total collection frequency, or
  - frequency distribution across the documents of the collection (Chowdhury, 2004}.

Steps in Preparing an Automatic Index

- The first step as summed up by Chowdury (2004)
  - identification of all words occurring in all the documents in a given collection
- The second step is
  - deletion of function words by consulting a stop-word list
- The third step is
  - preparation of word stems by suffix stripping.
- This helps to limit the number of words by producing a common stripped form for all the words that have the same root.

For example, **Nation** will stand for **National**, **Nationalism**, **Nationality**, **Nationalized**, **Nationalization**, etc.

- The final step is the computation of the value of each term in each document.

Advantages of Automatic Indexing

These include:

- level of consistency in indexing can be maintained
- index entries can be produced at a lower csost in the long run
- indexing time can be reduced, and
- better retrieval effectiveness can be achieved