

INFS 324

INDEXING AND ABSTRACTING

SESSION 7– INDEXING SYSTEMS I: PRE-COORDINATE INDEXING SYSTEMS

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

- Indexing systems may be described as a set of procedures for managing the contents of documents for retrieval and dissemination. The indexing process as you learned in Session 4 involves familiarization, analysis, and translation of the selected concepts into the language of the index. At the end of the translation stage, the indexer has to decide on the type of index he wishes to create. This is where the issues of indexing systems come in. There are two broad groups of indexing systems namely pre-coordinate indexing systems and post-coordinate indexing systems. In this Session, I will be discussing the pre-coordinate systems with you.

Session Objectives

By the end of this section, you should be able to:

- Distinguish between the two categories of indexing systems.
- Explain how Pre-Coordinate indexing works.
- Provide examples of Pre-Coordinate indexes.
- Discuss the difficulties inherent in the use of pre-coordinate indexes.
- Create samples of Pre-Coordinate indexes

Session Outline

The key topics to be covered in the session are:

- Topic One: Introduction
- Topic Two: Issues Relevant to Pre-Coordinate Indexing
- Topic Three: PRECIS
- Topic Four: Other Pre-Coordinate Indexing Systems

Reading List



Topic One:

Introduction to Pre-Coordinate Indexing Systems

What is Pre-Coordinate Indexing?

Indexing systems are a set of prescribed procedures for organizing the contents of records of knowledge (documents) for the purpose of retrieval and dissemination.

- There are two broad groups of indexing systems.

- These are the

- Pre-Coordinate Indexing Systems and the
Post-Coordinate Indexing Systems.

- They have been categorized in this manner as a result of the way they handle the subject matter of documents for the purpose of retrieval.

What is Pre-Coordinate Indexing(Cont.)

Pre-coordinate Indexing Systems are called so because of the way the systems treat the contents of documents for the purpose of search and retrieval.

- A pre-coordinate indexing system is one in which the indexer analyses the contents of composite subjects in order to select relevant terms for the purpose of retrieving information.
- The coordination of the selected terms for the purpose of retrieval is done by the indexer at the indexing stage.
- In other words in the pre-coordinate system the indexer prescribes the terms that may be used to access the index

What is Pre-Coordinate Indexing(Cont.)

Examples of pre-coordinate indexing systems are:

- the normal back of the book index;
 - the normal library catalogue;
 - other manual or hardcopy indexes.
- Users of pre-coordinate indexing systems will look for the prescribed term and follow instructions until the document is retrieved.

Advantages of Pre-Coordinate Indexing Systems

- One advantage of pre-coordinate indexing is that it avoids complex search logic.
- Another advantage is that they require no special physical format – many of them are printed or hardcopy indexes.
- There are a few that are computer-based.
- A notable example is **PRECIS** which we will look at in detail later on in this discussion.

Disadvantages of Pre-Coordinate Indexing Systems

There are a number of disadvantages associated with pre-coordinate indexing. These are

- The acceptable length or bulk of the index heading restricts the environment within which pre-coordinate indexing is workable.
- If very exhaustive indexing is done it will create headings with more components than it is desirable.
- Headings that exceed a particular length become unpredictable, that is they are difficult to handle or remember.
- The number of references that may be needed to index a document may make the index bulky.

Topic Two:

ISSUES RELEVANT TO PRE-COORDINATE INDEXING SYSTEMS



Consistent Description of Subject Heading

There are two main issues relevant to pre-coordinate indexing. These are:

- Consistent description of subjects
- Referencing
- Pre-Coordinate indexing deal with many subject headings.
- Describing the heading must be done in a consistent manner.
- Equally, there must be consistency in the arrangement of terms (Citation Order).

To be able to ensure consistency in the description and arrangement of terms, controlled vocabulary and standard citation order have to be used.

Consistent Description of Subject Heading(Cont.)

- Without controlled vocabulary, an independent indexer may decide to use different description for the same concept;

For example:

- 'automated libraries in Ghana'
 - 'computerized information units in Ghana'.
- Once this happens inconsistency occurs. Inconsistency affects the efficiency of the index.

Principles and Guidelines

Principles and guidelines have been evolved over the years to ensure :

- consistency in the description of concepts and
- the arrangement of terms.

- Three of the most notable of these are:
 - Cutter's Rules for a Dictionary Catalogue;
 - Kaiser's Systematic Indexing; and
 - Coates's rules for British Technology Index.

Cutter's Rules for a Dictionary Catalogue

The rules were evolved by Charles Ammy Cutter in 1876.

- The rules say that where a subject and a place are both elements of a topic:

the subject should precede place in scientific related areas.

For example: **“Ghana Institution of Engineers”** would be realized as **“Institution of Engineers: Ghana”**.

Cutter's Rules for a Dictionary Catalogue(Cont.)

On the other hand

- place should precede the subject in other areas like commerce, history, government, etc.

For example, **“Government of Ghana”** will be **“Ghana: Government”**.

- There have been improvements on the rules which are far more superior but
- Cutter's rules still remain important especially as they are used in the Library of Congress List of Subject Headings.

Kaiser's Systematic Indexing

This was also evolved in 1911.

- Kaiser's approach to citation was based on the significance of the concept or Heading.
- He believed that concepts that are significant must be cited first.
- To him composite subjects may be analyzed into:

“Concrete”

“Process”

He believed that if subjects with the two components were cited in the order of 'concrete' first and 'process' last, it would correspond to natural language.

Kaiser's Systematic Indexing(Cont.)

For example

“Boat building in Ghana”

according to Kaiser's guidelines will be cited as

“Ghana : Boat building”.

Another example is

“Servicing of Motor vehicles”.

This will be cited as

“Motor vehicles: Servicing”

Coates and British Technology Index

Coates also established rules to cater for the British Technology Index now called Current Technology Index.

- His rules covered a wide range of composite subjects.
- He based his rules on Kaiser's Systematic Indexing
- except that he changed Kaiser's
 “Concrete” and **“Process”** to
 “Thing” and **“Action”**.

Referencing

The second issue relevant to pre-coordinate indexing is Referencing.

Pre-coordinate indexing provides one term in primary position at a time using citation order.

It therefore makes it necessary to provide references or added entries to cater for other approaches in accessing the index.

For example

- "Education of information management personnel in Ghana" may have a 'see' or 'see also' reference to 'Nigeria', 'Kenya' etc and 'librarians'.
- 'See' refers to the authorized term that may be used to search the index;
- for example, 'murder' **see** 'homicide.' It means that documents that treat the topic of murder have been indexed under the preferred term.

Referencing(Cont.)

In effect references are used in two ways.

- First, a cataloguer uses references to help locate the correct headings.
References to headings in a catalogue or bibliography are interfiled with the headings.
- The second use of references is the information seeker who uses it to locate the correct heading for a topic as well as a means of moving between related topics.
- The use of references as pointed out by Rowley and Hartley (2008), is becoming outmoded.
When card and printed catalogues and bibliographies were the order of the day, it was normal to use references.
- It would be totally out of place to include references in computerized catalogues and bibliographies.

- TOPIC THREE

PRECIS



Introduction

Pre-Coordinate Indexing Systems are mainly manual indexes, however, there are a few which are computerized.

PRECIS is a notable pre-coordinate indexing system which is computer-based.

- PRECIS is the acronym for PREserved Context Indexing System.
- It is primarily a set of procedures for generating index entries.
- It is an alphabetical subject indexing system that gives 'précis' or summary of the subject content of a document at each entry point while displaying the terms in context.
- That is the subject of a document or item of information is provided under each term that the indexer considers to be important enough to be used as a sought term or a search word.

Introduction(Cont.)

- A controlled vocabulary of the terms is compiled and maintained in order to ensure that the same terms are used to describe the same topics or concepts whenever they appear in any subject.

By so doing the context within which the terms are used is preserved.

Hence the name PREserved Context Indexing System.

- PRECIS generate entries and references from an indexing string. The string establishes citation order for syntactic relationships while generating references for semantic relationships.
- To see clearly the citation order that is achieved in PRECIS index entries, it is important to note how an indexing string is created. A PRECIS indexing string cannot be created without PRECIS operators.
- Hence I will describe the operators before we look at how PRECIS works in practice.

PRECIS Operators

PRECIS operators are an important factor in PRECIS indexing.

- You need to be conversant with them in order to be able to create a PRECIS indexing string from which you will generate PRECIS index entries.
- PRECIS operators are called **role operators** also.
- There are over thirty role operators and other manipulation codes.
- They have three main functions.

Firstly they indicate the role of each term in the subject statement or string;

secondly, they determine the citation order; and

finally, they pass instructions to the computer for the precise pattern of rotation of the index entries under each lead term.

The instructions include the typography, punctuation and capitalization of the terms.

Primary Operators

There are two types of operators in PRECIS.

These are Primary operators formerly called Mainline operators and Secondary operators also formerly called Interposed operators.

- There are seven of the Primary operators (0-6) which stand for three sets of concepts.
 - The first operator, '0' stands for the **environment of core concepts**.
 - Operators 1, 2, and 3 are used for the **core concepts** while 4, 5, and 6 are used to represent **extra-core concepts**.
- Every string must start with either one of the operators 0 – 3.

Primary Operators(Cont.)

- The primary operators determine the citation order for the string and they are cited ordinally.
-
- 0 - Stands for specific geographical location e.g., Ghana, Africa
- 1 - Represents the key system or the object on which an action may be taken or its effect may be experienced.
- 2 - Shows action or phenomena, that is action or the effect of an action on the key system or the object.
- 3 - Indicates the performer of an action

Primary Operators(Cont.)

Thus the composite subject

- “Trailers for the conveyance of timber from Kumasi” will be represented as:
 - (3) Trailers (performers of the action)
 - (2) Conveyance (action)
 - (1) Timber (key system or object of the action)
 - (0) Kumasi (geographical location)

Primary Operators(Cont.)

- 4 - Introduces a point of view

For example,

“Political implications of students’ demonstrations in Ghana”
would be represented as:

(0) Ghana

(4) Political implications

(1) Students’ Demonstrations

- 5 - Introduces terms like ‘study regions’, ‘study samples’

Primary Operators(Cont.)

- For example, “African Women in Agricultural Development: a
- Case Study of Sierra Leone” will have the following concepts and operators:
 - (3)African Women
 - (1)Agricultural Development
 - (5)Case Study
 - (q)Sierra Leone

Primary Operators(Cont'd)

- 6 - Shows terms which describe the form of a document

For example:

bibliography;
annual report;
dictionary etc.

It also refers to target audience of a document

For Example:

“Manual of Indexing and Abstracting for Students”.

(6) Indexing and Abstracting Manual

(1) Students

Secondary Operators

The secondary operators perform secondary functions.

- They are also used for three sets of concepts
- They are represented by letters.
- Thus operators:
 - (f) and (g) are for co-ordinate concepts;
 - (p), (q) and (r) are for dependent elements;
 - (s), (t) and (u) are for special classes of action.

Secondary Operators(Cont'd)

- P is used to indicate a part of a thing or property of a thing or action

For example

“Bicycle wheels” would be:

(1) Bicycle

(p) Wheels

- q represents a member of a quasi generic group(refer to the example of Operator 5 above)
- r represents assembly

Secondary Operators(Cont'd)

- s indicates role definer
- t shows author-attributed association
- u represents two-way interaction.

However, we will look at only p and q in detail.

Steps in Creating PRECIS Indexing String

Having described what précis operators are let us now look at how the string is created.

Creating a PRECIS indexing string follows a series of steps.

Step One

- The first step is to identify the concepts in the composite subject that are to be reflected in the index entries. In PRECIS, a concept is defined as a term or topic that matches a précis operator.
- Using the composite subject
“Movement of labour from Mozambique to the mines of South Africa” as an example:

Step One(Cont.)

The following concepts may be identified:

Movement

Labour

Mozambique

Mines

South Africa

Economic Relations

Step Two

Express the identified concepts in the controlled vocabulary that will be used in the index.

- “Movement” may be changed to “migration” as the accepted term or the authorized term. Thus step two would be expressed as:

Migration

Labour

Mozambique

Mines

South Africa

Economic Relations

Step Three

Assign an “operator” or “code” to each term that has been identified.

- Operators are numbers or letters that are used to indicate the role of each concept in the subject.
- The operators have specific filing values.
- This is to ensure that terms appear in the indexing string in an order that will produce a meaningful set of index entries.

Step Three(Cont.)

Step Three would be expressed in the following manner:

- (p) Labour
- (2) Migration
- (0) Mozambique
- (1) Mines
- (0) South Africa
- (4) Economic Relations

Step Four

Arrange the index terms according to the filing value of the operators that have been assigned to the concepts.

- By the end of this step, an indexing string would have been created.
- The string would be made up of both terms and operators. Step Four therefore would be:
 - (0) Mozambique
 - (0) South Africa
 - (1) Mines (p) Labour
 - (2) Migration
 - (4) Economic Relations

Step Five

The indexer explores the possible entries that may be generated from the string.

- He may make whatever adjustments that may be necessary in terms of which terms should take lead position.
- He could also include or exclude terms at this stage.
 - (0) Mozambique
 - (0) South Africa
 - (1) Mines (p) Mine Workers
 - (2) Migration
 - (4) Economic Relations

Step Six

- Computer instruction codes or commands
 - replace the operators in the string and convert the operators into machine readable codes
 - that will show which terms to be used as entry terms.
- The indexer's task ends here.

Step Seven

The computer creates a series of entries based on the indexing string that was generated.

The next series of slides demonstrates the generation of index entries in PRECIS.

Generating Index Entries in PRECIS

- In Précis each entry has 3 basic positions. These are
- The **LEAD**, which appears on the first line to the left and in block letters.
- The **Qualifier**, which appears on the same line with the LEAD but to the right and in small letters.
- The **DISPLAY**, which appears on the 2nd line below the LEAD. The Display also is in small letters. A sample entry will look like the following:
 -
 - **LEAD** **Qualifier**
 - **Display**
 - The **Lead** is the term to be used as the access point. The **Qualifier** is the context establishing term and the **Display** indicates narrower terms.

Generating Index Entries in PRECIS(Cont.)

The first component of the string is used for the first index entry. The remainder of the string is left in the display position.

- A hypothetical index entry using **ABCD** would appear as the following:
- By a process of rotation, the previous **LEAD** term goes to **Qualifier** position.
- The next term at the head of the **Display** will move to **LEAD** position.
- Thus our hypothetical index entries will look like these:

A	
b c d	
B	a
c d	
C	b a
d	
D	c b a

Generating Index Entries in PRECIS(Cont.)

As can be seen all the concepts have been displayed in context. The process of rotation is called

-SHUNTING.

It is not always the case that you will have a standard pattern. Adjustments are sometimes made for the sake of clarity or comprehension,

For Example

it might be necessary to omit one or two terms;
sometimes adjustments of terms may be
necessary;

at other times too it might be necessary to insert
more terms.

Generating Index Entries in PRECIS(Cont.)

- Whatever interventions are made is done through computer instructions

For Example

the instruction '**LO**' to the computer.

means '**Lead Only**', therefore the computer is being asked to print the term when it appears in **Lead** position only.

- The standard lay out for the entries; punctuation; and typography have all been pre-programmed into the computer.
- Thus it automatically performs those functions.

PRECIS Index Entries: Step I

- Let us use the composite subject
“The diagnosis of heart disease in man by electro–
cardiography” (Foskett...)
to illustrate how PRECIS works.

Step I

Diagnosis

Heart

Disease

Man

Electro–cardiography

PRECIS Index Entries: Step II

Step II

Diagnosis

Heart

Acute Diseases

Man

Electro-cardiography

PRECIS Index Entries: Step III

Step III

(2) Diagnosis

(p) Heart

(p) Acute diseases

(1) Man

(3) Electro-cardiography

PRECIS Index Entries: Step IV

Step IV

- (1) Man (p) Heart (p) Acute diseases
- (2) Diagnosis
- (3) Electro-cardiography

PRECIS Index Entries: Step V

Step V

- (1) Man (p) Heart (p) Acute diseases
- (2) Diagnosis
- (3) Electro-cardiography

PRECIS Index Entries: Step VI

Step VI

- Computer instruction codes or commands
 - replace the operators in the string and convert the operators into machine readable codes
 - that will show which terms to be used as entry terms.

PRECIS Index Entries: Step VII

Step VII

Computer instruction codes to generate the following index entries:

1. MAN

Heart. Acute diseases. Diagnosis. Electro-cardiography

2. HEART Man

Acute diseases. Diagnosis. Electro-cardiography

PRECIS Index Entries: Step VII(Cont.)

3. ACUTE DISEASES Heart. Man
Diagnosis. Electro-cardiography

4. DIAGNOSIS Acute diseases. Heart. Man
Electro-cardiography

5. ELECTRO-CARDIOGRAPHY Diagnosis. Acute diseases. Heart. Man

Topic Four

Other Pre-Coordinate Indexing Systems



Other Pre-Coordinate Indexing Systems

PRECIS is by no means the only notable Pre-coordinate Indexing System.

- Other notable systems are:
 - POPSI,
 - Book Index, and
 - Keyword indexes.

POPSI

POPSI is another example of a pre-coordinate indexing system.

- It is the acronym for Postulate-based Permuted Subject Indexing.
- POPSI uses analytico – synthetic method for
 - string formulation and
 - permutation of the constituent termsto allow access from different points.
- Analytico – Synthetic method is a system that allows the indexer maximum leverage in adding new terms to the string even though they may not be listed in the controlled vocabulary.

POPSI(Cont.)

- A POPSI index entry is made up of two parts, namely
 - Lead Heading** which contains the access term
 - Context Heading** which normally appears on the second line after the lead heading.

The context heading contains the subject words with auxiliary words which show the context in which the lead term has been discussed.

- There are four basic categories of subjects under POPSI called **DEPA** for short. These are:
 - Discipline
 - Entity
 - Property
 - Action

POPSI(Cont.)

- **Discipline** is an elementary category. It includes conventional fields of study or any aggregate of such fields, e.g. mathematics, physics, library science, computer science, etc.
- **Entity** includes manifestations having perceptual correlates or conceptual existence, e.g. patent, lung, plant, etc.
- **Property** deals with manifestations that show qualitative or quantitative attributes, e.g. specific gravity, disease, efficiency, etc.
- **Action** deals with manifestations that indicate the concept of doing either as self action or external action, e.g. evaluation, examination, migration, etc.

Book Indexing

Book indexes, also called Back of the Book indexes are another type of Pre-coordinate Indexing System.

They are the most common indexes.

They assist a user in locating concepts (topics) in the text of a book.

Book indexes may be compiled by

- authors of the books themselves
- professional indexers.

Book Indexing(Cont.)

Entries in book indexes may be alphabetical by names of persons, corporate bodies, places or subject.

Most of the time, however, arrangement of entries follows a dictionary type of arrangement.

Each entry is followed by a page number or number of the pages on which the information is to be found.

Book Indexing(Cont.)

The same principles proposed by British Standards (BS3700: 1988) as the purposes of an index may apply to Book indexes. These are:

- Identify and locate relevant information within the material being indexed
- Discriminate between information on a subject and passing mention of a subject
- Exclude passing mention of subjects that offers significant to the potential user
- Analyse concepts treated in the document so as to produce a series of headings based on its terminology

Book Indexing(Cont.)

- Indicate relationships between concepts
- Group together information on subjects that is scattered by the arrangement of the document
- Synthesize headings and subheadings into entries
- Direct the user seeking information under terms not chosen for the index entries to the headings that have been chosen by means of cross-references
- Arrange entries into a systematic and helpful order

Book Indexing(Cont.)

- To create a book index, the indexer creates an informal controlled vocabulary based on the language used by the author.
- AACR2 (Anglo-American Cataloguing Rules 2nd edition) may be used in selecting headings for persons, places and corporate bodies.
- AACR2 is the rules that govern cataloguing of library materials.
- You can see copies of the rules at the Balme Library of the University of Ghana.

Book Indexing(Cont.)

- Book indexes may be done manually using 5” x 3” cards for ease of filing and inter- filing but
- There are dedicated indexing software available for use.
- Features of such software include:

Editing and display features which involve

- copying a previous entry;
- transposing main heading and subheading;
- searching and find-and-replace entries;
- creating of the index for the indexer to work on;
- verifying cross-references;
- on screen editing.

Book Indexing(Cont.)

- Sorting features comprising
 - word-by-word or letter-by-letter alphabetization,
 - immediate sorting of entries;
 - page number sorting;
 - merging of index files.
- Formatting and printing features:
 - removal of duplicate entries or page references;
 - automatic formatting and printing of the index in a range of styles including user-selected formats;
 - creation of user-defined style sheets.

Book Indexing(Cont.)

- The indexer's responsibility is to judge which items of information are important to be included and which should be excluded.
- The depth of indexing required (exhaustivity) will depend on the nature and length of the book.
- The average ratio of index pages to text pages is three percent (3%).

Book Indexing(Cont.)

- This may be much higher for reference books.
 - You will learn about reference books in your Information Sources course.
- A well outlined book may be easier to index than a less well structured one.
- The specificity of index headings may also be informed by the content of the book.

Keyword Indexes

They are called Keyword-in-Context (KWIC) indexes.

- They used to be popular because they were straightforward to create.
- In KWIC indexes, words in the title of documents are compared with a stop word list in order to avoid useless index entries.
- A stop word list contains words under which entries are not required.

These include

- prepositions (on, in, for etc),
- conjunctions (and, but yet, etc) ;
- pronouns (them, this, he, us etc).

Keyword Indexes(Cont.)

- Each word in the title is checked against the stop word list.
- If a match occurs the word is rejected, but
- if no match occurs the term is designated a keyword.
- The selected keywords are used as entry words (sought term) with one entry relating to the document for each word.

Keyword Indexes(Cont.)

- The word is printed in context with the rest of the title including stop words.
- The entry words are arranged alphabetically and aligned in the centre or left column.
- A single line entry, including title and source reference is produced for each significant word in the title. Below is an example of a KWIC index layout from the Biological Abstracts database :

Keyword Indexes(Cont.)

Subject Context	Keyword	Subject Context	Ref. No.
CONTROL WEED CONTROL/ LIMIT TEMPERATURE/ ECOLOZONATION SYNONYMY/A SU	ALTITUDINAL	DISTRIBUTION OF THE LANTANA DOWNSTREAM REPLACEMENT RAS- SPLENDENS VEGETATION	205 3128 1458
WEED CONTROL/ALTITUDINAL IVORY COAST CAMEROON/ BRAIN DEVELOPMENT/ ...	DISTRIBUTION	OF THE LANTANA LACE BUG OF THE MATING TYPE ALLE OF THE NOVEL DEVELOPMENT	205 9684 7907
WEED CONTROL BIOLOGICAL CONTROL SEEDING AND FREQUENCY SEED HUS-RETROFLEXUS GROWTH	WEED	CONTROL/ALTITUDINAL DIST CONTROL/COTTON GOSSYPI CONTROL/PETROLEUM OIL A\	205 172 9759

Keyword Indexes(Cont.)

- This index is described as more user friendly than some, however many titles are drastically truncated on any given index line.
- A slash appears at the end of a title to separate it from added subject terms.
- The significance of title indexes is mainly seen in the low human input in creating the index.

Advantages of KWIC

- Single KWIC indexes are completely computer-generated. As a result:
 - a large number of titles can be processed quickly and at low cost.
 - again the absence of personal intervention ensures consistency and predictability.
 - additionally, indexing based on words in titles reflects current usage.
 - also it makes for easy creation of cumulative indexes (covering about five annual volumes) because it does not require additional intellectual input.

Disadvantages of KWIC

In spite of the above advantages, title indexes have received some criticisms. For example:

- Some titles do not give accurate description of the context of a document. Titles can be misleading or eye-catching rather than informative.
-For instance, what do you make of a title like “On the care and Construction of White Elephants”?
It is the title of an article on the values of library catalogues.
- Basic KWIC indexes are unattractive and tedious to read because of their physical arrangement and type face and the limited information that they provide
- Vocabulary control is absent leading to irrelevant and redundant entries as well as scattering under different terms.

Solutions to KWIC Problems

To circumvent some of these criticisms, various terms of title indexes have been developed to enhance the basic KWIC index concept.

- These include :
 - KWOC (keyword-out-of-context),
 - KWAC (keyword-and-context) and
 - Double-KWIC (based on pairs of words).

References



References

