NATIONAL OPEN UNIVERSITY OF NIGERIA

SCHOOL OF SCIENCE AND TECHNOLOGY

COURSE CODE: DAM 207

COURSE TITLE: INDEXING AND CLASSIFICATION THEORY

DAM 207 COURSE GUIDE

INDEXING AND CLASSIFICATION THEORY

Course Writer: Edeama O. Onwuchekwa
National Open University of Nigeria
Lagos

Programme Leader Dr B. Abiola
National Open University of Nigeria
Lagos

Course Coordinator Vivian Nwaocha
National Open University of Nigeria
Lagos

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria Headquarters
14/16 Ahmadu Bello Way
Victoria Island
Lagos

Abuja Office
No. 5 Dar es Salaam Street
Off Aminu Kano Crescent
Wuse II, Abuja
Nigeria
E-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng

Published by

National Open University of Nigeria

Printed 2011

ISBN

CONTENTS PAGE

Introduction

The course guide provides you with the various topics on the introductory course in indexing and classification theory. In this course we will study the origin of classification and of the various classification schemes that are widely used. The theories of Indexing and classification process have also been discussed in this course. This course also considers the subject access system, general principles of subject headings, thesaurus and its construction. It goes beyond the traditional library access methods to discuss automatic indexing and classification.

The overall aims of this course are to introduce you to various classification schemes and Indexing system. Subject access system, general principles of subject headings and thesaurus construction are equally discussed.

There are four modules in this course; each module consists of 4 units of topics that you are expected to complete in 2 hours. The four modules and their units are listed below.

What You will Learn in this Course

The overall aims and objectives of this course is to provide guidance on what you should be achieving in the course of your studies. Each unit also has its own unit objectives which state specifically what you should be achieving in the corresponding unit. To evaluate your progress continuously, you are expected to refer to the overall course aims and objectives as well as the corresponding unit objectives upon the completion of each.

Course Aims

The overall aims and objectives of this course will help you to:

1. Develop your knowledge and understanding of the underlying concepts and theories of Indexing and classification.
2. To outline and critically analyze the different classification schemes
3. Build up your capacity to develop/ construct a subject access system(Thesaurus)
4. Develop your competence in understanding the concept of automatic indexing and classification.

Course Objectives

It is pertinent to note that each unit has precise objectives. Students should learn them carefully before proceeding to subsequent units. Therefore, it may be useful to refer to these objectives in the course of your study of the unit to assess your progress. You should always look at the unit objectives after completing a unit. In this way, you can be sure that you have done what is required of you by the end of the unit. However, below are overall objectives of this course. On successful completion of this course, you should be able to:

- Discuss the origin of classification
- Explain the term " Classification"
- Identify the purpose of classification
- Describe the theories of Indexing and classification
- Evaluate and Indexing and classification system
- Explain the basic concepts of the modern Classification Schemes
- Explain the criteria of a workable classification scheme
- Discuss the general principles of classifying a document
- Explain the general principles of subject headings
- Explain the similarities between the Sears and Library of Congress heading list.
- distinguish between a special classification scheme and a general classification
- outline the reasons for a special classification scheme
- Describe the process of thesaurus construction
- Explain the concept of Indexing
- Describe the techniques for Indexing
- identify the differences between the natural, free and controlled indexing language
- Discuss the concept and development in automatic Indexing and classification.

Working through this Course

We designed this course in a systematic way, so you need to work through it from Module one, Unit 1 through to Module four, Unit 15. This will enable you appreciate the course better.

Course Materials

For this course, you will require the following materials:

1) The course guide :
2) Study units which are fifteen (15)
3) Textbooks recommended at the end of the units and
4) The presentation schedule

Study Unit

There are fifteen study units in this course broken into four modules. They are as follows:

MODULE 1 BACKGROUND TO CLASSIFICATION

Unit 1 Early history, definition and purpose of classification
Unit 2 Theory of Classification
Unit 3 Classification Schemes and its basic components
Unit 4 Criteria for workable classification Scheme

MODULE 2 CLASSIFICATION SCHEMES

Unit 1 Dewey Decimal classification Scheme
Unit 2 Library of Congress Classification Scheme
Unit 3 Brief Summary of Other Classification Scheme
Unit 4 General Principles of classifying a document

MODULE 3 SUBJECT ACCESS SYSTEM

Unit 1 Subject Headings
Unit 2 General Principles of Subject Headings
Unit 3 Thesaurus and Practice in thesaurus construction

MODULE 4 INDEXING, AUTOMATIC INDEXING AND CLASSIFICATION

Unit 1 Indexing
Unit 2 Indexing System
Unit 3 Evaluation of an Index
Unit 4 Automatic Indexing and Classification

Each unit contains some exercises on the topics covered and you will be required to attempt the exercises. These will enable you evaluate your progress as well as reinforce what you have learnt so far. The exercises, together with the tutor-marked assignments will help you in achieving the stated learning objectives of the individual units and the course.

Assignment File

The assignment file will be given to you in due course. In this file, you will find all the details of the work you must submit to your tutor for marking. The marks you obtain for these assignments will count towards the final mark for the course. Altogether, there are 15 tutor marked assignments for this course.

Presentation Schedule

The presentation schedule included in this course guide provides you with important dates for completion of each tutor marked assignment. You should therefore endeavour to meet the deadlines.

Assessment

The course, Indexing and classification theory entails attending a two-hour final examination which contributes 50% to your final grading. The final examination covers materials from all parts of the course with a style similar to the Tutor-marked assignments.

The examination aims at testing your ability to apply the knowledge you have learned throughout the course, rather than your ability to memorize the materials. In preparing for the examination, it is essential that you receive the activities and Tutor-marked assignments you have completed in each unit. The other 50% will account for all the TMAs at the end of each unit.

Tutor-Marked Assignment

In this course, you will be required to study fifteen (15) units, and complete Tutor-marked Assignments, provided at the end of each unit. The assignments carry 10% marks each, the best four of your assignments will constitute 30% of your final mark. At the end of the course, you will be required to write a final examination, which counts for 70% of your final mark.

You may wish to consult other related materials apart from your course material to complete your Tutor-marked Assignments. Ensure that your assignments reaches your tutor on or before the stipulated deadline .If for any reason you are unable to complete your assignment in time, contact your tutor before the date due to discuss the possibility of an extension. Note that extensions will not be granted after the due date for submission unless under exceptional circumstances.

Final Examination and Grading

The final examination for DAM 207 will be of last for a period of 2 hours and have a value 60% of the total course grade. The examination will consist of questions which reflect the self assessment exercise and tutor marked assignments that you have previously encountered. Furthermore, all areas of the course will be examined. It would be better to use the time between finishing the last unit and sitting for the examination, to revise the entire course. You might find it useful to review your TMAs and comment on them before the examination. The final examination covers information from all parts of the course.

Course marking Scheme

The following table includes the course marking scheme

Table 1 Course Marking Scheme

ASSESSMENT	**MARKS**
Assignments 1-15	15Assignments, 40% of the best 4 Total=10%X 4 = 40
Final Examination	60% of the overall course marks
Total	100% of Course Marks

Course Overview

This table indicates the units, the number of weeks required to complete them and the assignments.

Table 2: Course Organizer

Unit	**Title of Work**	**Weeks Activity**	**Assessment (End of Unit)**
Course Guide		Week 1	
Module 1		**Background to Classification**	
Unit 1	Early History , definition and purpose of classification	Week 1	Assignment 1
Unit 2	Theory of Classification	Week 2	Assignment 2
Unit 3	Classification Schemes and its basic components	Week 3	Assignment 3
Unit 4	Criteria for workable classification Scheme	Week 3	Assignment 4
Module 2		**Classification Scheme**	
Unit 1	Dewey Decimal Classification Scheme	Week 4	Assignment 5

Unit 2	Library of Congress Classification Scheme	Week 4	Assignment 6
Unit 3	Brief Summary of other classification Scheme	Week 5	Assignment 7
Unit 4	General Principles of classifying a document	Week 5	Assignment 8
Module 3		**Subject Access System**	
Unit 1	Subject Headings	Week 6	Assignment 9
Unit 2	General Principles of Subject Headings	Week 6	Assignment 10
Unit 3	Thesaurus and practice in thesaurus construction	Week 7	Assignment 11
Module 4		Indexing ,Automatic Indexing and Classification	
Unit 1	Indexing	Week 8	Assignment 13
Unit 2	Indexing System	Week 9	Assignment 14
Unit 3	Evaluation of an Index	Week 10	Assignment 15
Unit 4	Automatic Indexing and Classification		Assignment 16

How to Get the Most from this Course

In order for you to learn various concepts in this course, it is essential to practice. Independent activities and case activities which are based on a particular scenario are presented in the units. The activities include open questions to promote discussion on the relevant topics, questions with standard answers and program demonstrations on the concepts. You may try to delve into each unit adopting the following steps:

1. Read the study unit
2. Read the textbook, printed or online references
3. Perform the activities

4. Participate in group discussions
5. Complete the tutor-marked assignments
6. Participate in online discussions

There are also optional readings in the units. You may wish to read these to extend your knowledge beyond the required materials. They will not be assessed.

Facilitators/Tutors and Tutorials

About 15 hours of tutorials will be provided in support of this course. You will be notified of the dates, time and location for these tutorials, together with the name and phone number of your tutor as soon as you are allotted a tutorial group.

Your tutor will mark and comment on your assignments, keep a close watch on your progress and on any difficulties you might encounter and provide assistance to you during the course. You must mail your TMAs to your tutor well before the due date (at least two working days are required). They will be marked by your tutor and returned to you as soon as possible.

Do not hesitate to contact your tutor by phone, e-mail if you need help. The following might be circumstances in which you would find help necessary. You can also contact your tutor if:

i. You do not understand any part of the study units or the assigned readings
ii. You have difficulty with the TMAs
iii. You have a question or problem with your tutor's comments on an assignment or with the grading of an assignment.

You should try your best to attend tutorials, since it is the only opportunity to have an interaction with your tutor and to ask questions which are answered instantly. You can raise any problem encountered in the course of your study. To gain maximum benefit from the course tutorials, you are advised to prepare a list of

questions before attending the tutorial. You will learn a lot from participating in discussions actively.

Summary

The course, Indexing and classification theory is intended to develop your understanding of the basic theories of Indexing and classification theories, thus enabling you understand the subject access processes. This course also provides you with an understanding of automatic indexing and classification processes. We hope that you will find the course enlightening and that you will find it both interesting and useful. In the longer term, we hope you will get acquainted with the National Open University of Nigeria and we wish you every success in your future.

DAM 207

INDEXING AND CLASSIFICATION THEORY

Course Writer: Edeama O. Onwuchekwa
National Open University of Nigeria
Lagos

Programme Leader Dr B. Abiola
National Open University of Nigeria
Lagos

Course Coordinator Vivian Nwaocha
National Open University of Nigeria
Lagos

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria Headquarters
14/16 Ahmadu Bello Way
Victoria Island
Lagos

Abuja Office
No. 5 Dar es Salaam Street
Off Aminu Kano Crescent
Wuse II, Abuja
Nigeria
E-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng

Published by

National Open University of Nigeria

Printed 2011

ISBN

MODULE 1: BACKGROUND TO CLASSIFICATION

UNIT 1: HISTORY, DEFINITION AND PURPOSE OF CLASSIFICATION

CONTENTS

1.0 INTRODUCTION

Classification pervades all the various activities of our life. Those who are orderly in life make a greater use of classification, though unconsciously. Most of us are unaware of the fact that we classify to a large extent in our daily lives. Without classification, human progress could be impossible. This process helps convert unorganized thought and impressions into recognizable patterns.

Since inception, efforts have been made to classify information based on certain criteria such as order of acquisition, size of documents, title, and subject. The most successful attribute so far for classifying documents is by subject. This unit will discuss the history and origin of classification and shed light on the definition and purpose of classification in an information environment.

2.0 OBJECTIVES

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

- explain the historical antecedent of classification
- define and explain the term classification
- Identify the purposes of classification in an information organization.

3.0 MAIN CONTENT

3.1 THE HISTORY OF CLASSIFICATION

Earlier attempts at classification were really attempts to organize human thought: they were designed to aid the mental plotting of the universe of thought and objects, rather than serve as practical methods of document arrangement or library systems. These can be traced, somewhat tentatively, from Plato's *Republic* to the late 19th century classifications, via Aristotle, Pliny and others. Libraries, too, have a history of developing schemes. The clay tablets in the Assyrian library of Asur-ban-i-pal were divided into at least two main classes; those dealing with knowledge of the earth and those dealing with the heavens. Aristotle, too, is said to have 'taught

the Kings of Egypt how to arrange a library'. The earliest recorded scheme was that designed by Callimachus for the library of the Pharaohs at Alexandria (260-240BC).

The term classification is derived from the Latin word "classis" used to distinguish any one of the six groups into which the Romans were according to their wealth and social status. In the early times, Library materials were arranged in one or a combination of one of the following ways: size, title, broad subject, author, chronology, binding and color.

The traditional ideas of library classification were borrowed from the logical or philosophical principles of classification. Classification began with the universe of knowledge as whole and divided into successive stages of classes and subclasses, with a chosen characteristic as the basis for each stage. On the whole, the progression is from the general to the specific, forming a hierarchical, or "tree," structure, each class being a *species* of the class on the preceding level and a *genus* to the one below it. The classes on each level, usually mutually exclusive and totally exhaustive categories, form a coordinate relationship to one another and are collocated according to the affinity of their relationships. Classification according to hierarchical principles, with biologic taxonomy the prevailing model, was in a particularly active stage of development during the latter part of the nineteenth century.

The first modern scheme devised specifically for the arrangement of books in a library was one designed by Aldus Manutius in France in 1498. It developed into what was called *The French System* or *The System of the Paris Booksellers,* which became the most influential and widely used of all bibliographic schemes, especially in Europe.

Other schemes of interest appeared in the eighteen and nineteen centuries. The real beginnings of classification as we know it today took place in the nineteen century in America with the growth of the Congress Library.

SELF ASSESSMENT EXERCISE

Discuss the history and origin of classification.

3.2 DEFINITION OF CLASSIFICATION

Classification, broadly defined, is the act of organizing the universe of knowledge into some systematic order. It has been considered the most fundamental activity of the human mind. The essential act of classification is the multistage process of deciding on a property or characteristic

of interest, distinguishing things or objects that possess that property from those which lack it, and grouping things or objects that have the property or characteristic into a class. Other essential aspects of classification are establishing relationship among classes and making distinctions within classes to arrive at subclasses and finer divisions. The classification of library materials follows the same pattern; it is thus a special application of a much more general human intellectual activity.

Library classification in particular has been defined as "the systematic arrangement by subject of books and other material on shelves or of catalogue and index entries in the manner which is most useful to those who read or seek a definite piece of information". In other words, library classification serves a dual function: to arrange items in a logical order on library shelves and to provide a systematic display of bibliographic entries in printed catalogs, bibliographies, and indexes. Today, in some online catalogs, classification also serves a direct retrieval function.

Classification is the formal process by which a mechanism is established to translate these similarities and dissimilarities into a place in a physical sequence. Documents display a number of attributes which can be used to determine likeness. The four attributes that documents possess are: -

Author: the person or persons intellectually responsible for the creation of the work.

Title: the title of the individual work

Form: the physical form in which the document appears

Subject: by the content of the work, the subject matter which it contains.

In any collection, the most appropriate basis for determining groups varies according to the needs of the collection, for example, library materials may be grouped by author, physical form, size, date of publication, or subject. In modern library classification systems, subject is the predominant characteristic for grouping.

3.3 PURPOSE OF CLASSIFICATION

Classification in the sense of grouping things together (either literally or mentally) goes back to ancient civilizations. In this widest sense, it remains an activity which we all practice in everyday

life: we have a mental map or 'classification' in which we associate or dissociate the objects, ideas and impressions that are our experience of the world. Classification systems seek to provide a structure for the organization of materials so that an item may be retrieved according to some aspect of its character.

Classification is thus seen as organizing stock for effective service, a collection is an amorphous and unrevealing entity without the guiding light of classification. The ultimate aim of the classification sequence is to provide a physical arrangement where similar materials are closely located on the shelf and, within subject groupings of like subjects, an order of general to specific subjects is observed.

Classification is not only the grouping of things for location or identification purposes; it is also their display in some sort of rational, progressive (usually subject) order so that their chief relationships may be ascertained. Relative Location of subjects is a great time saver; it is the purpose of classification systems to collate subjects and allows the subject to be preserved.

Classification provides insight to the generally accepted divisions of knowledge. It splits entire knowledge and provides opportunities for limiting a subject and for proceeding from a general subject to the smaller specialized areas of that subject. A close look at any classification scheme reveals in considerable detail the components of that subject.

As a shelving device, library classification has two objectives: to help the user identify and locate a work through a call number and to group all works of a kind together. In order to fulfill the first objective, any method of numbering or marking would be sufficient as long as there is a correspondence between the number or mark on the document and that on the cataloging entry. The second objective, on the other hand, represents a collocating function and requires the grouping of like materials together on the basis of chosen characteristics. Thus, in its function as a retrieval tool, classification may help to identify and retrieve a group of related items as well as a specific known item.

SELF ASSESSMENT EXERCISE

Briefly explain the purposes of classification?

4.0 CONCLUSION

In this unit you have learnt about the historical perspective of classification of information and several views of definition of classification have been discussed. You should also have learned about the purposes and rationale of classification.

5.0 SUMMARY

Classification is a major device for organizing resources for effective use. What you have learned in this unit concerns the origin of classification and makes light on a proper understanding and purpose of the concept of classification. The units that follow shall build upon issues discussed in this unit.

6.0 TUTOR-MARKED ASSIGNMENT

1) Explain the term "classification"?

2) Discuss the purpose of classification in an information organization?

7.0 REFERENCES/FURTHER READINGS

- Aina, L.O. (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd
- Chan,Lois Mai (1994) Cataloguing and Classification: An Introduction. New York: Mc Graw Hill Inc
- Edoka, B.E. (2000) Introduction to Library Science. Nigeria: Palma Publishing & Links Company Ltd.
- Kumar, Krishan. (1998) Theory of Classification. New Delhi: Vikas Publishing House.
- Richardson, E.C (1991) Classification, theoretical Practical. New York: Gower.

UNIT 2 THEORY OF CLASSIFICATION

CONTENTS

1.0 INTRODUCTION

A theory refers to an organized body of principles. These principles provide guidance to practitioners of the concerned subject. Any theory, like any subject, goes through a process of growth and development. Therefore, a theory might be an elementary one or an advanced one, depending upon its stage of growth. In this unit, you will learn about the theory of classification and its basic concepts.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- Describe the theories surrounding classification
- Explain the systematic study of classification as it concerns its growth and development.

3.0 MAIN CONTENT

3.1 THE THEORY OF CLASSIFICATION

Any theory of a subject undergoes a process of evolution and the same goes for theory of library classification. The theories of classification which existed till the early fifties are referred to as the descriptive theories of library classification because they more or less described as the practices in use at that time.

In the earliest applications of classification when for the first time the shelves were open to the public, classification practice forged ahead of theory. In many respects, perhaps of necessity, it did so for a long time. The description of theory which follows is a concise overview of a process which developed at first very gradually. During the period of descriptive theory, theory followed practice. Thus theory had little power to influence practice; it was manipulated to fit into practice.

During that period, schemes were largely designed with the flair or natural aptitude of the classificationist. The classificationist received occasional help from intuition. This period lacked to a large extent guidance from a theory of library classification. Often there was lack of scientific approach; as a result there was no objectivity in the approach.

The beginning of another stage in the evolution of library classification theory should be considered as an important landmark. This has led to tremendous development in the field of library classification. From about the middle of the early century there was acceleration; Ranganathan in particular, and later his disciples, began to perceive how a fully developed theory could be put together.

The dynamic theory was able to provide a sound methodology for the designing of a scheme for library classification. Various special schemes were produced, some clearly intended for use and based on user needs, others more experimental in outline. In the United Kingdom, a Classification Research Group was set up. By the time one British enthusiast, Jack Mills, produced a textbook in 1960, it was possible to present principles which had been thought through. This later theory is quite different from the early years when some enthusiasts had looked to the history of science, philosophical principles and logic for inspirational guidance.

In more recent years, the pace of theoretical innovation has slowed; much of what could be argued and demonstrated had so been done by the late 1960s. The theories developed have been accepted much more enthusiastically by some people than by others and, generally speaking, much more enthusiastically in some countries than in others.

SELF ASSESSMENT EXERCISE

Explain the developmental growth of the theory of classification.

4.0 CONCLUSION

Since the existence of libraries in ancient times, effort have been made to classify books based on certain criteria such as order of acquisition, size of documents, title, color chronology and subject. The most successful attribute for classifying documents is by subject. This has been proven by the different theories of classification that has emerged.

5.0 SUMMARY

In this unit you have learned about the theory of classification. You have also learned about the progression of the classification theory from the descriptive theory to the dynamic theory. Finally, you have been able to learn about the developmental phases of the classification theory.

6.0 TUTOR-MARKED ASSIGNMENT

Discuss the theory of classification

7.0 REFERENCES/FURTHER READINGS

- Aina, L.O (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd
- Foskett, A.C (1996) the Subject Approach to Information 5th Edition. London: Library Association.
- Kumar, Krishan. (1998) Theory of Classification. New Delhi: Vikas Publishing House.
- Rowley J.E. and Farrow J, (2000) Organizing Knowledge: An introduction to managing Access to Information. 3rd Edition. London: Gower.
- Wynar, Bohdan S. (1992) Introduction to Cataloging & Classification. 8th Edition. Englewood , Colorado: Libraries Unlimited
- Richardson, E.C (1991) Classification, theoretical Practical. New York: Gower

UNIT 3 CLASSIFICATION SCHEMES AND ITS BASIC COMPONENTS

CONTENTS

1.0 INTRODUCTION

Several general classification schemes are in use. There are also specialized schemes that are used for special collections with in-depth stock but in a specific subject area. It is therefore needful to have a general overview of major classification schemes in order to know the different context to which they can be applied.

This unit describes classification schemes. It discusses the main component of a classification scheme. You will learn about Schedule, Notation and the index and its features and qualities.

In the subsequent unit, you will be taught about special classification schemes.

2.0 OBJECTIVES

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

- Understand the general classification schemes
- Identify the components of a classification scheme
- Discuss the schedule as a main component of a classification scheme.
- Explain the features and qualities of a notation

- Describe the importance of an index to a classification scheme

3.0 MAIN CONTENT

3.1 CLASSIFICATION SCHEMES

The general classification scheme is one which is designed to cover all the subjects in the universe of subjects. Classification schemes must first list and arrange the principal disciplines of knowledge. It will be useful to consider the manner that subjects are treated in various classification schemes. Three types of subjects exist: simple, compound and complex subjects

A **simple subject** is one that deals with the whole main class or with a single aspect of a class. 'Surgery' as a subject treated as part of the 'medical sciences' would be a simple subject. Other examples might be 'politics' or 'the Democratic Party' or 'political campaigns'

A **compound subject** is one where two or more simple subjects are combined, or treated equally in a text; an example might be 'cats and dogs' or 'painting and drawing'. Compound subjects are often described as interactions of two or more simple subjects from the same main class. Although in the examples above the simple subjects which are combined come from the same subdivision of their class, compound subjects may also be multi-faceted, such as 'monetary economics in France' or 'campaigning in the Democratic Party'.

A **complex subject** is one where the combined simply subjects emanate from two or more main classes. Again there are limitless possibilities, but examples might include 'market research in book publishing' or 'surveys and their impact on political campaigning'. With complex subjects the two or more subjects which are brought together may not be equal in treatment; they may rather have a relationship to each other, as in 'the impact of income tax changes on the life style of old age pensioners'.

3.2 COMPONENTS OF A CLASSIFICATION SCHEME

A classification scheme consists of three main components. These are:

3.2.1 The Schedule

The schedule is perhaps the most important part of a classification scheme. In a schedule, the terms representing the subject content of documents are arranged systematically showing their relationships. A schedule must meet certain criteria and some of them are:

- All major disciplines should be represented if the classification is supposed to cover all subject knowledge.
- The space allocated to each discipline should be approximately proportional to the literature of the field.
- The order of classes should bring all related subjects to close proximity. This is why botany and agriculture should be together because they are related. Similarly, language and literature are in close proximity.
- The schedule must provide space for new knowledge, especially since the frontiers of knowledge are being continually extended.
- In order for the schedule to cover all subject terms, it is necessary to provide a place for each subject term whether simple or complex.

The two main methods for constructing a classification schedule are the enumerative and faceted methods.

3.2.2 Notation

In order to retrieve documents from the shelves, there is always a notation which is assigned to the subject headings of the document. The notation is a shorthand code representing the various subject headings used in describing the subject content of documents in a collection. The notation is usually made up of letters, numerals, or a combination of both. This is the code assigned to subject terms, which is used in an index or catalogue. It helps in achieving the orderly arrangement as expressed in the schedules. A notation is an important requirement in a classification scheme, as a poor notation would lead to a complete breakdown of the arrangement of documents. A good notation must be able to accommodate new subjects, be flexible and simple to use.

There are two types of notation – pure and mixed. Generally, a notation uses alphabetical letters and Arabic numerals. Pure notation uses only one type of symbols, either Arabic numerals or alphabetic letters. Dewey decimal classification schedule uses only Arabic numerals for

example. Mixed notation on the other hand uses both Arabic numerals and alphabetic letters. An example of mixed notation is the Library of Congress Classification scheme.

The notation is the link between the subject terms in the schedules, which are arranged systematically, and subject terms in the index, which are arranged alphabetically.

Features of a Good Notation

There are certain features a notation must possess in order to perform its functions. These are that:

- A notation must be constructed in such a way that it will be easy to use. A good notation must be: (i) easy for users to write, remember, or type (ii) simple (iii) brief and (iv) mnemonic, that is there should be aids that will make it easy for users to remember, e.g. in Dewey Decimal Classification – 6 always stands for Africa – 03 for dictionaries. In LC, T stands for technology and G for geography.
- A notation must be hospitable; it should be able to accommodate new subjects.
  A notation must leave gaps that are unassigned, which can be used in future, or it could be decimal which would ensure that the notation can be expanded by the use of decimal sub-division.
- A notation must possess a device for synthesis as this enables the classifier to specify a document.

3.2.3 Index

This is the last component of a classification scheme which lists all the subject terms that have been arranged systematically in the schedules in an alphabetical order. The index enables a user to locate topics, which have been systematically arranged, and also displays the related aspects of a subject, which have been scattered in the schedule. There are two main types of index. These are specific index and relative index.

Specific index lists all subject terms in a straightforward alphabetical order disregarding the various aspect of a subject term. This type of index is useful when there is a specific entry for each subject term in the classification.

Relative Index on the other hand, gathers together all the aspects of a subject term which are likely to have been scattered in the schedules.

Every classification scheme needs an index as a guide to the place of subjects in the schedules themselves. This is in addition to any subject index which might be provided in a library as a

guide to its collection on the shelves. The index to the classification scheme is essentially a guide for the classifier to find the appropriate section(s) of the schedules where the subject concerned may be found. It lists topics, locates them and includes all necessary synonyms. Essentially it should do two things:

1. Locate subjects within the systematically arranged classification, and
2. show related aspects of a subject which are distributed beyond a single class, often due to the problems, noted earlier, caused by classification by discipline.

The index is an essential tool for the classifier, showing these relationships and guiding the classifier's choice of possible locations for a work. The published index to a classification scheme is not intended as an index for library users or client; it is intended to be a key or guide to the schedules, rather than to the works on the shelf of the library. It will of course contain entries for all subjects dealt with by the classification scheme; possibly all knowledge will be represented in the schedules to a general scheme and in its index and this is likely to include a significant amount of material that would not be held in any but the very largest of libraries.

SELF ASSESSMENT EXERCISE

List and explain the types of subjects that exist in a classification scheme.

4.0 CONCLUSION

In this unit you have learnt about the classification scheme and the main components of a classification scheme which are:

Schedule - the subject terms are listed systematically in the schedule.

Notation - this contains the codes assigned to the different subject terms listed in the schedules.

Index - this is an alphabetical arrangement of all the subject terms listed in the schedule. The notation assigned to each subject term is listed against the subject term in the index.

5.0 SUMMARY

What you have learned in this unit concerns the different subject in classification schemes and the components of a classification scheme.

6.0 TUTOR-MARKED ASSIGNMENT

List and explain the three main components of a classification scheme.

7.0 REFERENCES/FURTHER READINGS

- Aina, L.O. (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd
- Chan,Lois Mai (1994) Cataloguing and Classification: An Introduction. New York: Mc Graw Hill Inc
- Edoka, B.E. (2000) Introduction to Library Science. Nigeria: Palma Publishing & Links Company Ltd.
- Kumar, Krishan. (1998) Theory of Classification. New Delhi: Vikas Publishing House.
- Richardson, E.C (1991) Classification, theoretical Practical. New York: Gower.

UNIT 4 CRITERIA FOR A WORKABLE CLASSIFICATION SCHEME

CONTENTS

1.0 INTRODUCTION

All classification schemes are made up of *classes* or categories. If we browse through any of the widely used general schemes, they appear to be gigantic maps of knowledge. They begin with general works in what is called *a generalia class.* This class is designed to accommodate general collections of essays, multi-topical encyclopedias and other works too wide in scope to be linked with any one discipline; these are works which deal with, or attempt to deal with, all knowledge.

Not all classification schemes can be agreed to be workable or efficient. The ultimate test of all schemes lies in their practical application and in their responsiveness to the most common approach of users. In this unit, we shall therefore discuss the criteria for a workable classification scheme

2.0 OBJECTIVES

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

- Identify the features of a workable classification scheme
- List and explain the criteria of a workable classification scheme

3.0 MAIN CONTENT

3.1 CRITERIA FOR A WORKABLE CLASSIFICATION SCHEME

A good universal classification scheme, apart from having a schedule, notation and index, must have certain features that would make it usable. A good classification scheme must:

- Cover the whole knowledge as reflected in the literature. Thus, single-concept and multi-concept documents must be taken care of.
- Be systematic, that is related subjects must be brought together as close as possible. All aspects of a subject must be brought together in a systematic manner.
- Be regularly revised. Thus, it must have an organizational support that will ensure constant revision. The need to be up to date is very important as new subjects appear and existing subjects sometimes need to be expanded because of the growth of the literature. The scheme must be able to accommodate such subjects. The accommodation of new subjects and expansion of existing subjects, however, should not disrupt the entire arrangement of the scheme. In between revisions. The organization might produce updates which are called additional changes.
- Ensure that the terminology used in the scheme is unambiguous. It must be clear and precise to the users and the classifiers.

4.0 CONCLUSION

The nature of the universe of subjects is multi-dimensional, with the development of knowledge; these subjects have to be incorporated in the classification scheme.

Thus a scheme has to keep pace with the developments in the universe to remain relevant and workable.

5.0 SUMMARY

What you have learned in this unit is focused on the different criterion of a workable classification scheme.

6.0 TUTOR-MARKED ASSIGNMENT

List and explain the criteria of a classification scheme.

7.0 REFERENCES/FURTHER READINGS

- Aina, L.O. (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd
- Edoka, B.E. (2000) Introduction to Library Science. Nigeria: Palma Publishing & Links Company Ltd.

.

MODULE 2 CLASSIFICATION SCHEMES

INTRODUCTION

Unit 1 Dewey Decimal classification Scheme

Unit 2 Library of Congress Classification Scheme

Unit 3 Other Classification Scheme

Unit 4 General Principles of classifying a document

UNIT 1: DEWEY DECIMAL CLASSIFACTION SCHEME (DDC)

CONTENTS

1.0 INTRODUCTION

The publication in 1876 of a pamphlet entitled a *classification and subject Index for cataloguing and arranging the books and pamphlets of a library* marked the beginning of the Dewey decimal classification which was soon adopted by many libraries in the United States and later by Libraries around the world. This unit provides an introduction to the Dewey Decimal Scheme, its main classes and its strengths and weaknesses.

2.0 OBJECTIVES

After going through this unit, you should be able to:

- describe the origin and beginnings of the DDC
- discuss the basic concepts of the scheme
- explain the subjects in the main class of the DDC
- List the weaknesses and strengths of the Dewey Decimal Classification Scheme.

3.0 MAIN CONTENT

3.1 BEGININGS OF DDC

The Dewey decimal classification scheme was the brain child of Melvil Dewey. It is the oldest of the classification schemes in use. Born in 1851 in the USA, Dewey graduated from Amherst College in 1874.As a student; he began to work in the college library and stayed on after graduation.

The first edition of his scheme was published in 1876 under the title: *A classification and subject index for cataloguing and arranging the books and pamphlets of a library*. The name Dewey did not appear on the title page but in the copyright notice of the verso of the title page.

Melvil introduced the idea of "relative location" as opposed to "fixed location". He assigned decimal numbers (Arabic numerals with decimal fraction notation) to books and not to shelves. Dewey was the first to popularize the idea of mechanization.

DDC has been translated, with or without abridgement, expansion or adaptation into many languages such as Spanish, Danish, Turkish, Japanese, Portuguese etc. This is perhaps the most popular classification scheme in libraries all over the world. It is universal and very popular in Africa.

3.2 INTRODUCTION AND BASIC CONCEPTS

The Dewey decimal classification divides the whole spectrum of knowledge as contained in information materials into ten broad categories. Each of this which is called a "class" is given a three-digit number. Each main class can be further divided into 10 sub-classes; each subclass can still be further divided into 10 divisions and each division into 10 subdivisions until all the subject terms have been specified. The primary arrangement of classes is based on disciplines rather than subjects.

000 Generalities
100 Philosophy
200 Religion
300 Social Sciences
400 Language
500 Natural Sciences and Mathematics
600 Technology (Applied Sciences)
700 The Arts
800 Literature
900 Geography & History

- It uses decimal to specify subject terms that are specific which will probably result in digits of long numbers. The use of Arabic numerals and decimal allows it to be expanded indefinitely.
- It uses Arabic numerals only as notation, therefore it has pure notation.
- It has many mnemonic devices, especially in the use of auxiliary tables that can be used throughout the scheme e.g. 03 for dictionaries and encyclopedias. Since the same numbers can be used in the schedule it is very easy to remember.
- Dewey decimal classification scheme provides a relative index to the diverse materials in the schedules.

3.3 EVALUATION OF DEWY DECIMAL CLASSIFICATION SCHEME

A great deal has been written about the merits and weakness of the Dewey decimal classification. Following is a brief summary of some of the opinions.

Merits

1. It is a practical system. The fact that it has survived many storms in the past hundred and twenty years and is still the most widely used classification scheme in the world today attests to its practical value.
2. Relative location was an innovation introduced by Dewey, even though it is now taken for granted.
3. The relative indeed brings together different aspects of the same subject scattered in different disciplines.

4. The pure notation of Arabic numerals is universally recognizable. People from any cultural or language background can adapt to the system easily.
5. The self-evident numerical sequence facilitates filing and shelving.
6. The hierarchical nature of the notation expresses the relationships between and among the class numbers. This characteristic particularly facilitates online searching. The searcher can broaden or narrow a search by reducing or adding a digit to the class number.
7. Use of the decimal system enables infinite expansion and subdivision.
8. The mnemonic nature of the notation helps library users to navigate within the system.
9. The continuous revision and publication of the schedules at regular intervals ensure the currency of the scheme.

Weaknesses

1. The Anglo-American bias is obvious, particularly in 900 (Geography and history) and 800 (Literature). A heavy bias toward American Protestantism is especially evident in 200 (Religion).
2. Related disciplines are often separated, e.g., 300 (Social sciences) from 900 (Geography and history) and 400 (Languages) from 800 (Literature).
3. The proper placements of certain subjects also have been questioned, e.g., Library science in general works (000s), Psychology as a subdivision under Philosophy (100s), and Sports and Amusements in The arts (700s).
4. In 800, literary works by the same author are scattered according to literary form when most scholars would prefer to have them grouped together.
5. The base of ten limits the hospitality of the notational system by restricting the capacity for accommodating subjects on the same level of a hierarchy to nine divisions.
6. The different rate of growth in various disciplines has resulted in an uneven structure. Some classes, such as 300 (Social sciences), 500 (Natural sciences), and 600 (Technology), have become overcrowded.
7. Even though an existing subject can be expanded indefinitely by virtue of the decimal system, no new numbers can be inserted between coordinate numbers e.g., between 610 and 620 – even when required for the accommodation of new subjects. The present method of introducing a new subject is to include it as a subdivision under an existing subject.

8. While the capacity for expansion is infinite, it also results in lengthy numbers for specific and minute subjects. The long numbers have been found inconvenient, particularly when the system is used as a shelving device.
9. Relocations and completely revised (i.e., phoenix) schedules, while necessary to keep up with knowledge, create practical problems in terms of reclassification for libraries using the scheme.

SELF ASSESSMENT EXERCISE

Explain the origin and beginning of the Dewy Decimal Classification Scheme

4.0 CONCLUSION

In this unit you have learned about early beginnings of the Dewey Decimal Classification Scheme. You have also learned about its main Classes, the strengths that has kept the scheme throughout the years and the also about the issues that has been considered the Scheme's weakness.

5.0 SUMMARY

DDC is the oldest and most widely used scheme of classification .Despite its severe criticism it has been adopted by a majority of libraries in English speaking and British Commonwealth countries. It has been adopted mainly due to its simple notation, its case in application, the adaptability of its notation to the requirements of libraries of different sizes and nature(and for the fact that it can be expanded with ease), availability in a variety of editions and its use in bibliographies and catalogue. In this unit you have learnt about the Dewey Decimal Scheme, the main classes and its strengths and weakness.

6.0 TUTOR-MARKED ASSIGNMENT

Discuss the strength and weaknesses of the Dewey Decimal Classification Scheme.

7.0 REFERENCES/FURTHER READINGS

- Aina, L.O. (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd
- Chan,Lois Mai (1994) Cataloguing and Classification: An Introduction. New York: Mc Graw Hill Inc
- Edoka, B.E. (2000) Introduction to Library Science. Nigeria: Palma Publishing & Links Company Ltd.
- Foskett, A.C (1996) The Subject Approach to Information 5th Edition. London: Library Association.
- Kumar, Krishan. (1998) Theory of Classification. New Delhi: Vikas Publishing House.
- Marcella Rita & Newton Robert. (1994) A New Manual of Classification. Gower Publishing limited, Aldershot.
- Richardson, E.C (1991) Classification, theoretical Practical. New York: Gower.
- Wynar, Bohdan S. (1992) Introduction to Cataloging & Classification. 8th Edition. Englewood, Colorado: Libraries Unlimited

UNIT 2: LIBRARY OF CONGRESS CLASSIFICATION SCHEME (LC)

CONTENTS

1.0 INTRODUCTION

In the nineteenth century, the Library of Congress collection was organized according to a system devised by Thomas Jefferson. When the library moved into its new building in 1897, the Jeffersonian system was found to be inadequate for a collection that had grown to over one and a half million pieces. Two other classification systems, the Dewey Decimal Classification (DDC) and Charles A. Cutter's Expansive Classification (EC), had emerged during the last few decades of the century and were in use in many other libraries in the nation, but neither was considered suitable for the Library of Congress. It was decided to construct a new system, to be called the *Library of Congress Classification* (LCC), and work began on its development.

In this unit we will look at the History and Early beginnings of the Library of Congress Classification Scheme. We will also consider its basic concepts and the Strengths and weaknesses of the scheme.

2.0 OBJECTIVES

- By the end of this unit, the student should be able to:
- describe the evolutionary development of the library of Congress Classification Scheme
- outline the main classes of the scheme
- explain the basic concepts of the L.C scheme
- State the strengths and weakness of the L.C Scheme

3.0 MAIN CONTENT

3.1 BEGININGS OF LIBRARY CONGRESS CLASSIFICATION

Herbert Putnam initiated the Library of Congress (LC) Scheme. It evolved between 1899 and 1920 when the scheme was first published. From the beginning, individual classes were developed by different groups of specialists under the direction of J.C.M. Hanson and Charles Martel; the schedules, each of which contains and entire class, a subclass, or a group of subclasses, were published separately. Thus, unlike most other classification systems, LCC was not the product of one master mind. Indeed, it has been called "a coordinated series of special classes."

Today, the Library of Congress Classification consists of twenty-one classes displayed in over forty-seven separately published schedules. Its provisions are continually updated, and information on additions and changes is made widely available to the library community.

Although the scheme was originally designed for the Library of Congress, many other libraries all over the world are using it. It is useful in large libraries because it is a detailed classification scheme

During the 1960s in particular, there was a trend among academic libraries previously using DDC or other systems to switch to LCC. There were several reasons for the trend: (1) the basic orientation of LCC toward research libraries; (2) the economic advantage offered by LC cataloging services – libraries can simply adopt whole call numbers as they appear on LC

cataloging records; and (3) the increasing ease with many libraries can bring up full LC records online and add them to their own catalog databases.

SELF ASSESSMENT

Describe the early beginnings of the Library of Congress Classification Scheme.

3.2 BASIC CONCEPTS

- The scheme is based on literary warrant, which is the collection of the library of Congress.
- Since it is based on existing collection, the scheme is minutely detailed.
- It is an enumerative Scheme and covers all knowledge.
- The notation is mixed; it uses single capital letters for its main classes and double letters for sub-classes. It uses Arabic numerals for further sub-divisions. It also uses Cutter numbers to further specify a document.
- The format of schedules is the same for each schedule consisting of the following: a preface or prefatory note, a brief synopsis which covers the primary sub-divisions of each volume; an outline and an index which is very detailed.
- The field has been divided into 20 classes, with an additional class for general works. The outline is given below

A. General works
B. Philosophy, Psychology, Religion
C. Auxiliary sciences of history
D. History: General and Old World
E-F History: America
G. Geography, Maps, Anthropology, Recreation
H. Social sciences
J. Political science
K. Law
L. Education
M. Music and books on music
N. Fine arts
P. Language and Literature

Q Science
R. Medicine
S. Agriculture
T. Technology
U. Military science
V. Naval science
Z. Bibliography and Library science

3.3 EVALUATION OF THE LIBRARY OF CONGRESS

The library of Congress Classification has both strong and weak points.

Merits

1. It is a practical system that has proved to be satisfactory.
2. It is based on the literary warrant of the materials in the library of Congress classification collection, the nature and the content of which are parallel to those in academic and research libraries.
3. It is largely an enumerative system that requires minimal notational synthesis.
4. Each schedule was developed by subject specialist rather than by a generalist who cannot be an expert in a field.
5. Its notation is compact and hospitable.
6. There are frequent additions and changes, and these are made readily available to the cataloging community.
7. The need for reclassification of large blocks of material is kept to a minimum because, to ensure stability of class members, few structural changes have been made over the year.

Weakness

1. Its scope notes are inferior to those of DDC
2. There is much national bias in emphasis and terminology.
3. Too few subjects are seen as compounds. Alphabetical arrangements are often used in place of logical hierarchies.
4. There is no clear and predictable theoretical basis for subject analysis.

5. As a result of maintaining stability, parts of the classification are obsolete in the sense that structure and collocation do not reflect current conditions.
6. It is expensive to keep an up-to-date working collection of schedules, supplements, new announcements of changes, and accumulations of additions and changes.

4.0 CONCLUSION

In this unit you have learned about the Library of Congress Classification Scheme. You have also been able to identify the strengths and weaknesses of the L.C Scheme.

5.0 SUMMARY

What you have learned in this unit concerns the Library of Congress Classification Scheme, its history, classes, strengths and weakness. In the next unit you shall learn about other Classification Schemes and the special classification scheme.

6.0 TUTOUR-MARKED ASSIGNMENT

Discuss the concept of the Library of Congress classification scheme.

State the merits and weakness of the LC scheme

7.0 REFERENCES/ FUTHER READINGS

- Aina, L.O. (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd
- Chan,Lois Mai (1994) Cataloguing and Classification: An Introduction. New York: Mc Graw Hill Inc
- Edoka, B.E. (2000) Introduction to Library Science. Nigeria: Palma Publishing & Links Company Ltd.
- Foskett, A.C (1996) The Subject Approach to Information 5th Edition. London: Library Association.
- Kumar, Krishan. (1998) Theory of Classification. New Delhi: Vikas Publishing House.
- Marcella Rita & Newton Robert. (1994) A New Manual of Classification. Gower Publishing limited, Aldershot.
- Richardson, E.C (1991) Classification, theoretical Practical. New York: Gower.
- Wynar, Bohdan S. (1992) Introduction to Cataloging & Classification. 8th Edition. Englewood , Colorado: Libraries Unlimited

UNIT 3: OTHER MODERN CLASSIFICATION SCHEME

CONTENTS

1.0 INTRODUCTION

In this unit we will look at other classification schemes. We will also consider the special classification scheme and the reasons for making a special classification scheme.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- describe the feature of the other modern classification scheme
- distinguish between a special classification scheme and a general classification
- outline the reasons for a special classification scheme
- List and explain the other types of classification scheme.

3.0 MAIN CONTENT

3.1 THE UNIVERSAL DECIMAL CLASSIFICATION

The origins of UDC lie in Europe. The UDC was produced by two Belgian Lawyers, Paul Otlet Paul Henry La Fontaine, who were much more interested in analysis of ideas rather than on general theory of knowledge(DDC) or Literary ideas of knowledge (Library of congress classification Scheme). The first edition was published in French in 1905.Subsequent editions, some of which have been abridged versions have appeared in various languages.

Its creation was a conscious attempt to develop with permission and redesign Dewey classification in order to meet the needs of precise classification for highly specific themes, such as might be the subject matter of periodical articles and general literature. It developed from DC, but in a very different way, and in some respects pointed the way forward to Colon Classification and synthetic classification.

UDC uses ten Arabic numerals with decimal fraction notation. The decimal point in the beginning is omitted. A decimal point is usually places after every three digits, merely to serve as a visual aid.

It is generally used by special Libraries. In view of the high level of precision which can be achieved by its use, specialized information centres in industrial and technological establishment apply the scheme.

3.2 THE COLON CLASSIFICATION

The Colon Classification (CC) was developed by S.R. Ranganathan, a prominent librarian from India who is considered by many to be the foremost theorist in the field of classification because of his contributions to the theory of facet analysis and synthesis. The Colon Classification is a manifestation of Ranganathan's theory, which has had a major influence on all currently used classification and indexing systems.

In the colon classification, knowledge is divided into more or less traditional main classes. Each class is broken down into its basic concepts or elements according to certain characteristics, called facets. In isolating these component elements, Ranganathan has identified five fundamental categories, often referred to as PMEST: Personality (entity in question), Matter

(materials, substances, properties, etc.), Energy (operations, processes, activities, etc.), Space (geographic areas and features), and Time (periods, dates, seasons, etc.). When classifying a document, the classifier identifies component parts that reflect every aspect and element of the subject content and puts them together according to a structural procedure, called a facet formula, which has been individually designed for each main class.

Thus, unlike enumerative classification schemes, CC does not list complete ready-made numbers in its schedules. A combination, or synthesis, of notation is tailored for each work in hand. Notation for the Colon Classification is extremely mixed and complex. It combines Arabic numerals, capital and lowercase letters, some Greek letters, brackets, and certain punctuation marks.

The Generalia classes are represented by Arabic numerals. Main classes are shown by capital letters of the Roman alphabet and certain Greek letters. Basic concepts and elements under each main class are represented mainly by Arabic numerals. CC itself, however, has not been widely used.

3.3 THE BIBLOGRAPHIC CLASSIFICATION

Henry Evelyn Bliss (1870-1955) believed for many years that libraries needed a more erudite system than Decimal classification to win more intellectual respectability of subject specialist and of education.

From the beginning, several principles guided Bliss's work. These are consensus, collocation of related subjects, subordination of special to general, graduation in specialty, and the opportunity for alternative locations and treatments.

BC uses the decimal fraction notation. It uses mixed notation, consisting of 26 roman capital letters, 26 Roman smalls alphabet and nine Arabic numerals (excluding zero).It also uses punctuation marks, mathematical symbols and improved digits. Its notation is largely non-hierarchical. Bliss was obsessed with achieving the possible shortest class number

BC special features include alternative location for certain themes where expert views might differ, short notation, and some selective thinking of pure and applied sciences. BC has been mainly used in Britain and the commonwealth countries.

Self Assessment: From the above description, state the differences between the UDC, CC and BC Scheme.

3.4 SPECIAL CLASSIFICATION

The special classification scheme is the one designed to cover the field of specialization which may be astronomy, astrophysics, microbiology, forestry, Philosophy, international relations, English poetry, plant anatomy, human nervous system, political thought etc. A In this case the field of specialization may be referred to as a host class.

There are many reasons claimed for making a special classification. Some of these reasons are:

(i) Lack of co-extensiveness: most general schemes do not provide enough details required for dealing with micro-documents in documentation.

(ii) Lengthy class numbers: most general schemes provide lengthy class numbers for complex subjects to be dealt with.

(iii) Special requirements or special point of view: general schemes are designed to take into consideration the majority point of view. Therefore, these are not able to fully meet the special requirements of a particular special library or information centre.

(iv) Lack of flexibility (provision for new subjects without disturbance of the preferred sequence): very often, general schemes lack flexibility to a certain extent.

(v) Lack of helpful sequence: very often, a general scheme may not be able to achieve optimum helpfulness in the arrangement of documents or entries.

Broadly speaking, we will identify three different approaches to special classification schemes. These are given below:

(i) The approach to a fully autonomous special classification. Such a scheme may be independent of any general scheme.

(ii) The special classification scheme may be prepared so that it is autonomous for special subject(s), and dependent upon general classification for allied subjects. Such a scheme is not fully autonomous.

(iii) The special classification scheme may be designed in such a way that it is dependent upon a general classification. It may be an extension of the general scheme. Such an approach may be called the do-all classification approach.

4.0 CONCLUSION

In this unit you have been introduced to other modern classification scheme. You have also learnt about the special classification scheme. Finally, you have been able to learn about the reasons for a special classification scheme in an information unit.

5.0 SUMMARY

What you have learned in this unit is focused on other types of modern classification scheme, the special classification scheme and the reason for its establishment.

6.0 TUTOR-MARKED ASSIGNMENT

List and explain the UDC, CC and BC classification Schemes

What are special classification scheme? Explain the reasons for making a special classification scheme.

7.0 REFERENCES/FURTHER READINGS

- Chan, Lois Mai (1994) Cataloguing and Classification: An Introduction. New York: Mc Graw Hill Inc
- Edoka, B.E. (2000) Introduction to Library Science. Nigeria: Palma Publishing & Links Company Ltd.
- Kumar, Krishan (1998) Theory of Classification. New Delhi: Vikas Publishing House.
- Marcella Rita & Newton Robert. (1994) A New Manual of Classification. Gower Publishing limited, Aldershot.
- Richardson, E.C (1991) Classification, theoretical Practical. New York: Gower.
- Wynar, Bohdan S. (1992) Introduction to Cataloging & Classification. 8th Edition. Englewood , Colorado: Libraries Unlimited

UNIT 4: GENERAL PRINCIPLES OF CLASSIFYING A DOCUMENT

CONTENTS

1.0 INTRODUCTION

The first important step to classifying a document is to have an overview of the classification scheme in use. The mastery of the scheme will normally come with ease and the summary of the classes can easily be learnt. This will enable the student appreciate how the overall knowledge is grouped by the scheme.

2.0 OBJECTIVES

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

- explain what it means to classify a document?
- list and explain the guidelines involved when classifying a document

3.0 MAIN CONTENT

3.1 General Principles of Classifying Documents

The process of classification involves establishing the main subjects treated in a document. Having identified the relevant subject terms and using a subject heading list or a thesaurus to assign the appropriate subject heading, the classifier would check the index to enable him/her have an idea of the classification number. This number can then be confirmed from the schedule.

- A document must be classified first, according to the subject content and then by the form in which the subject is represented except in the general class and literature where the forms are predominant over the subject.
- A document must be classified where it is most useful. A dictionary of physics would be most useful with physic collection rather than in the general class collection.

A class assigned to a document must be specific as much as possible. Thus a document should be placed in the specific class that will contain it rather than with a general topic .For example, classify dog under notation assigned to dog rather than domestic animals.

- A document that deals with two or more subjects should be placed in the class of the most predominant subject or the subject that is named first. But when a document covers more than three subjects than a class that covers all the subject should be chosen.

4.0 CONCLUSION

The classification of a document involves translating it from natural language into a classificatory language of ordinal numbers. This unit highlighted the guidelines to follow when classifying a document.

5.0 SUMMARY

The Universe of knowledge is dynamic and new subjects in various forms are always cropping up. What you have learned in this unit is focused on the general principles of classifying a document.

6.0 TUTOR-MARKED ASSIGNMENT

Discuss the guidelines to consider when classifying a document.

7.0 REFERENCES/FURTHER READINGS

- Aina, L.O. (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd
- Edoka, B.E. (2000) Introduction to Library Science. Nigeria: Palma Publishing & Links Company Ltd.

MODULE 3: SUBJECT ACCESS SYSTEM

Unit 1 Subject Headings

Unit 2 General Principles of Subject Headings

Unit 3 Thesaurus and Practice in thesaurus construction

UNIT 1 SUBJECT HEADINGS LIST

CONTENTS

1.0 INTRODUCTION

The subject heading is a word or phrase that describe the subjects treated in the book. As much as possible, a subject heading must represent the common usage among the readers of that subject and also the subject heading chosen must be specific to the content of the document. However, only one subject term can be chosen to index all materials on the same subject in the collection.

This unit describes the Sears Subject Heading list and the Library of Congress subject heading list.

2.0 OBJECTIVES

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

- describe a subject heading
- give examples of the two types of subject headings
- describe the Library of Congress subject heading
- explain the similarities between the Sears and Library of Congress heading list.

3.0 MAIN CONTENT

3.1 SEARS LIST OF SUBJECT HEADINGS

The list is primarily used in small public libraries and school libraries. It was first published in 1923 by Minnie Sears. The first edition was known as the list of Subject Headings for Small Libraries. It was based on the subject terms used by nine small libraries in the United States of America. *Sears List of Subject Headings* is narrower in scope than the *Library of Congress Subject Headings* (*LCSH*). The headings of sears are based on *LCSH* but with modifications to cater for small libraries. Many terms from *LCSH* for children's Literature are included. *Sears List of Subject Headings* appears both in electronic version and print version. The CD-ROM version is updated annually while the print version is updated every three years.

The philosophy of sears List is based on two principles:

1. Theory of Specific Entry
2. Theory of Unique Entry

In the Theory of Specific Entry, Sears emphasizes the need to choose the most specific entry to describe a document. Thus, soccer will be used to describe a book on football rather than sports.

The second theory, *Theory of Unique Entry* emphasizes the need to choose only one subject heading to describe all documents on the same subject. For example, while the terms and phrases-librarianship, library science and library studies refer to the same subject, only one on these terms and phrases can be used to describe all documents on the same subject at all times. Other terms and phrases would be referred to the chosen term and phrase. The chosen subject term must be logical, consistent and unique.

SELF ASSESSMENT

Discuss the features of a Sears heading List

3.2 LIBRARY OF CONGRESS HEADINGS (LCSH)

Library of Congress Headings (LCSH) is a list of subject headings originally developed by the Library of Congress for use on its cataloging records. The list was begun toward the end of the nineteenth century and first published in 1914. Since then, it has become the standard list used by most large general libraries in the United States, as well as by many special libraries and some smaller libraries; it is also used in many libraries abroad. The machine-readable version is also available on CD-ROM, called CDMARC Subjects. LCSH is revised weekly, with new and changed headings incorporated into the Subject Authority file. The weekly updates are published every month in print; the microfiche and CD-ROM versions are issued every three months and represent an accumulation of recent additions and changes into the main list.

LCSH is essentially a subject authority list; in other words, it is a list of terms authorized by the Library of Congress for use in its own subject cataloging. Libraries using LCSH for subject authority control have relied on the list and follow Library of Congress policies and practices as de facto standards. To use LCSH effectively, it is important to realize its scope: what it contains and what it does not contain. Its most prominent feature is the set of headings authorized for use as subject access points in bibliographic records.

4.0 CONCLUSION

Subject headings list require ongoing maintenance. This could be accomplished through a control system called subject authority system which for each term documents the basis for decisions on the term and on what links connect it with other terms.

5.0 SUMMARY

In this unit we have learnt that a subject heading list is a term (a word or a group of words) denoting a subject under which all material on that subject is entered in a catalog and that the Sears list of subject heading and the Library of Congress subject headings list are very similar in format and structure.

6.0 TUTOR-MARKED ASSIGNMENT

Differentiate between the Sears heading list and the Library of Congress Heading List.

7.0 REFERNCES/FUTHER READINGS

- Foskett, A.C (1996) the Subject Approach to Information 5th Edition. London: Library Association.
- Rowley J.E. and Farrow J, (2000) Organizing Knowledge: An introduction to Managing Access to Information. 3rd Edition. London: Gower.
- Rowley J (1992) Organizing Knowledge: An introduction to information Retrieval 2nd Edition London. Gower
- Taylor Arlene G. (1999) Organization of Information. Englewood, Colorado: Libraries Unlimited
- Wynar, Bohdan S. (1992) Introduction to Cataloging & Classification.8th Edition. Englewood, Colorado: Libraries Unlimited

UNIT 2: GENERAL PRINCIPLES OF SUBJECT HEADINGS

CONTENTS

1.0 INTRODUCTION

On the concept of subject headings, the public usage's becomes an important determining factor in selecting the terms and the forms of subject headings. The effective approach to effective subject headings is to attempt to develop a system that adheres to strictly formed principles. This unit will list and describe some of these general principles of subject headings

2.0 OBJECTIVES

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

- Discuss the essence of the principles of a subject heading list
- list and explain the general principles of subject headings
- understand how these principles apply to different subject headings

3.0 MAIN CONTENT

In current subject heading systems, the most important factors, evolved over the years. These are the general principles of subject headings (1) uniform and unique headings, (2) specific and direct entry, (3) consistency and currency in terminology, and (4) provision of cross-references.

3.1 UNIFORM AND UNIQUE HEADINGS

In order to show what a collection or a database has on a given subject, it must adopt a principle of *uniform headings;* that is, it must bring under one heading all the materials dealing principally or exclusively with that particular subject. This principle is similar to that requiring a uniform heading for a given personal author. If a subject has more than one name *(ascorbic acid and vitamin C,* for instance), one must be chosen as the heading. In general, it is hoped that the term chosen is unambiguous and familiar to all users of the catalog. Similarly, if there are variant spellings of the same terms *(e.g., marihuana and marijuana)* or different possible forms of the same heading, only one is used as the heading. Examples of variant heading forms might be *Air quality* versus *Air-Quality* or *Quality of air.* One must be chosen, with the others listed as lead-in terms.

The converse of the principle of uniform headings is *unique headings;* that is, the same term should not be used for more than one subject. If the same term must be used in more than one sense, as is often the case when different disciplines or fields of knowledge are involved, some qualification or classification must be added so that it will be clear to the user which meaning is intended, e.g., *Cold and Cold (Disease).*

3.2 SPECIFIC AND DIRECT ENTRY

The principle of *specific entry* governs both how subject heading are formed (thesaurus construction and maintenance) and how they are assigned to documents (indexing or subject cataloging). Regarding formulation, the principle requires that a heading be as specific as (in other words no broader than) the topic it is intended to cover. In application, it requires that a work be assigned the most specific heading that represents its subject content. Ideally, the heading should be coextensive with (no broader or narrower than) the subject content of the work.

3.3 CONSISTENT AND CURRENT TERMINOLOGY

It follows from what has been said above, particularly regarding the justifications for uniform headings, that the terminology in headings should be both consistent and current. Two elements are particularly important here: synonymy and changing usage.

Choices among synonymous terms may require difficult decisions. By principle, common usage prevails when it can be determined. For example, a popular tem is preferred to a scientific one in a general library and in standard lists of headings designed for general collections. Of course, the more specialized a library's collection and clientele, the more specialized its indexing terminology should be; special libraries, therefore, often develop their own thesauri or make extensive modifications of standard lists.

3.4 CROSS-REFERENCES

Three types of cross-references are used in the subject in the subject headings structure: (1) the *see* (or USE) reference, (2) the *see also* (or BT (broader term), NT (narrower term), or RT (related term) reference, and (3) the general reference.

3.4.1 See (or USE) references

To make sure that users who happen to consult the catalog under different names for (or different forms of the name of) a given subject will be able to locate material on it, see or USE references are provided to lead them form the terms they have looked under to the authorized heading for the subject in question. These references guide users *from* terms that are not used as headings to the authorized headings.

3.4.2 See also (including BT, NT, and RT) references

This type of reference connects headings that are related in some way, either hierarchically or otherwise. Unlike the see reference, a see also reference relates headings that are all used as entries in the catalog. The headings involved may overlap in meaning but are not fully synonymous – if they were, they would not both be used in the catalog. By connecting related headings, the see also (RT, for related term) reference calls the user's attention to material related to his or her interest. By linking hierarchically related headings, see also (BT, for broader term, and NT, for narrower term) references restore some of the advantages of the classed catalog in an alphabetical specific catalog, in that the user is guided to specific branches or aspects of a subject.

3.4.3 General reference

While a specific reference directs the user from the term being consulted to another individual heading, a general reference directs the user to a group or category to headings instead of to individual members of the group or category. An obvious advantage of using general reference is economy of space; they obviate the need to make long lists of specific references.

SELF ASSESSMENT

What are the types of cross –references used in the subject headings structure?

4.0 CONCLUSION

In this unit you have learned about the general principles of subject headings.

5.0 SUMMARY

You have learnt that the general principles of subject headings are

- Uniform and Unique Headings
- Specific and Direct Entry
- Consistency and Current terminology
- Cross- References

6.0 TUTOR-MARKED ASSIGNMENT

List and discuss the general principles of subject headings.

7.0 REFERNCES/FUTHER READINGS

- Chan,Lois Mai (1994) Cataloguing and Classification: An Introduction. New York: Mc Graw Hill Inc
- Foskett, A.C (1996) the Subject Approach to Information 5th Edition. London: Library Association.
- Rowley J.E. and Farrow J, (2000) Organizing Knowledge: An introduction to Managing Access to Information. 3rd Edition. London: Gower.

- Rowley J (1992) Organizing Knowledge: An introduction to information Retrieval 2nd Edition London. Gower
- Taylor Arlene G. (1999) Organization of Information. Englewood, Colorado: Libraries Unlimited
- Wynar, Bohdan S. (1992) Introduction to Cataloging & Classification.8th Edition. Englewood, Colorado: Libraries Unlimited

UNIT 3: THESAURUS

CONTENTS

1.0 INTRODUCTION

In this unit we will learn a few basic concepts of a thesaurus as one of the methods of subject access system. This unit will introduce you to the practice of a thesaurus construction

2.0 OBJECTIVES

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

- explain what a thesaurus is
- explain terms “descriptors” and “non-descriptors”
- discuss the relationships in thesaurus
- Differentiate between a thesaurus and subject heading list
- Describe the process of a thesaurus construction.

3.0 MAIN CONTENT

3.1 THESAURUS

A thesaurus is compilation of words, It also a form of an authority list. It is similar to a subject heading list because it is a 'compilation' of words and phrases showing synonyms, hierarchical and other relationships and dependencies, the function of which is to provide a standardized vocabulary for information storage and retrieval.

Just like subject headings list, it exercises control over subject terms and displays the relationship amongst terms which will inform the reader of other search terms that can be used in searching. Unlike the subject headings list, which are generally aimed at all types of searches; the thesaurus is aimed at a specialized collection. Thus it is always limited in subject scope.

The thesaurus consists of both descriptors and non-descriptors. The descriptors are terms that that can be used to describe the subjects or concepts in a document while the non-descriptors cannot be used as search terms, but helps the searcher to have a broader search, nevertheless as the searcher would be directed to the appropriate descriptor.

For each descriptor listed, there is a sort of relationship between the descriptor and other words. Generally the descriptors are single terms or what is called uni-terms but occasionally they could be multi-terms. Each term that is used as a descriptor or an index term is unambiguous.

3.2 RELATIONSHIP IN A THESAURUS

3.2.1 Hierarchical Relationship

There are some descriptors that are use to index subject catalogues but they have hierarchical relationships.

a) BT (Broader Term). This involves directing the user to a broader term of the descriptor being searched e.g.

   Journalism

   BT Press

   Media

b) NT (Narrower Term).This is used to direct the searcher to other narrower terms of the descriptor chosen.

Press

NT Journalism

Broadcasting

Broadcasting" and "journalism" are narrower terms to Press.

These terms (BT and NT) are hierarchical relationships to the descriptor under which they are indicated. This enables the searcher to look for additional information that might aid the search. The use of (BT and NT) allows a searcher to expand his/her own search.

3.2.2 Preferential Relationships

This type of relationship is more or less a preferential one. The searcher is directed from a term that cannot be used to the one that is allowed to be used. Examples of these relationships are USE and UF (Used for).

USE precedes word that can be used as a descriptor. There are terms that are listed which cannot be used but can only use the descriptor under which they are listed.

Librarianship

USE Library Science

Library Science

UF Librarianship

3.2.3 Affinitive Relationship

There are terms that have no hierarchical relationship but the relationships between such terms are coordinate. These terms are designated as related terms.

Related Terms (RT) are used to connect terms that are related (the two terms are descriptors) for e.g.

Media

RT Press

Press

RT Media

These terms are only indicated in the thesaurus, they will not be found in the in the catalogue or in the index.

Besides these three common relationships, there are some others that are regularly used in a thesaurus depending on the complexity of the subject or the users of the system. These include:

GT Generic To

SA See Also

TT Top Term in Hierarchy

XT Overlapping Term

AT Associated Term

CT Coordinate Term

ST Synonymous Term

SU See Under.

The use of scope notes (SN) is also prevalent in a thesaurus as it provides explanations on the proper use of the terms in the thesaurus.

SELF ASSESSMENT

Describe the relationships stated in a thesaurus

3.2 PRACTICES IN THESAURUS CONSTRUCTION

The process of thesaurus construction involves the following steps:

- The coverage of the subject field must be determined, so also is the depth of the subject to be covered. It is advisable to consult an encyclopedia in order to gauge the breadth of the subject.
- The coverage of the subject field can be categorized into two: these are the core subject field and the peripheral subject field. A thesaurus compilation should be restricted to the core area of the subject field.
- There is need to identify the main subject areas on the field. It is important to go through indexes, abstracts, dictionaries encyclopedias etc to ensure that all relevant terms have been selected.
- This stage involves selecting the synonyms, related terms and other variants of the terms selected earlier. This might be obtained from existing thesauri or classification schemes that cover the subject area or subject fields.
- The term can be recorded on a 3 by 5 slip. Each slip would show the index term BT, NT, RT, USE and UF (whichever is applicable) and the source from which the term was located. This slip is made in duplicate copies; one set would be filed in a straightforward

alphabetical order. The order sets of slips would be filed according to the major subject categories.

- The thesaurus should be tested to see if it works .Checks should be made on relationships. A this stage, the thesaurus may be edited as appropriate. The thesaurus should be revised regularly as to keep pace with new terminologies on the subject.

4.0 CONCLUSION

This unit highlighted that a thesaurus is a compilation of words and also a form of an authority list. Unlike the subject headings list, which are generally aimed at all types of searches; the thesaurus is aimed at a specialized collection and is limited in subject scope.

5.0 SUMMARY

In this unit you have been taught to understand the term “thesaurus” and the relationships in it. We have also considered the practice of thesaurus construction.

6.0 TUTOR-MARKED ASSIGNMENT

- What is a thesaurus?
- Describe the process of a thesaurus construction

7.0 REFERNCES/FUTHER READINGS

- Aina, L. O. (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd
- Chowdhury (1999) Introduction to Modern Information Retrieval. London : Facet Publishing.

MODULE 4: INDEXING, AUTOMATIC INDEXING AND CLASSIFICATION

UNIT 1 INDEXING

CONTENTS

1.0 INTRODUCTION

In this unit, you will learn about Indexing and the different types indexing languages.

2.0 OBJECTIVES

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

- explain the concept of indexing
- Discuss the levels in the process of indexing
- Explain the theories of Indexing process.
- identify the differences between the natural, free and controlled indexing language

3.0 MAIN CONTENT

3.1 INDEXING PROCESS

The technique of producing an index is called indexing. Indexing is the process of providing a guide to the intellectual content of a document or a collection of documents. The result of this process is an index, which will serve as a pointer to the intellectual content in a document. It is able to perform this role through the descriptors that are used in describing the intellectual content of documents. The reader who is interested in a document will use the descriptors assigned to the document by the indexer. The ultimate objective of the index is to reduce the efforts a user expends in accessing a topic of interest in a particular document or a set of documents which have been stored in a collection.

An index is an important tool for retrieving information contained in documents stored in the library, documentation or information centre. It provides a means of locating the information relevant to a request. Some researchers have postulated that there are five levels in the process of indexing:

The first level is known as the concordance, which consists of all references to all words in the original text arranged in alphabetical order.

The second level is the information theoretic level which calculates the likelihood of a word being chosen for indexing based on its frequency of occurrence in a given text document.

The third level is the linguistic level which attempts to explain how meaningful words are extracted from large units of text. (Some Indexers have proposed that opening paragraphs, chapters etc are good sources for choosing indexing terms).

The textual or the skeletal framework is the **fourth level**. Here the text is prepared by the author in an organized manner and held together by a skeletal structure. The onus therefore lies on the indexer to identify the skeleton and markers that will determine the content of the given text.

The fifth level of indexing theory is the inferential level. An indexer should be able to make inferences about the relationships between words and phrases by understanding the sentence structure.

3.2 INDEXING LANGUAGE

Indexing language is made up of words or descriptors that are used in the intellectual contents of documents. These terms are expected to be used by the searcher in order to search for documents in collection.

There is always a need for an artificial language to be used by the indexer and the searcher to describe a document since the terms or concepts identified in a book are represented by words or phrases. The function of this type of language is to ensure that the indexer and the searcher operate at the same level by using the same language. This is to facilitate the retrieval of relevant information from the collection of the library. This language appears in a variety of forms. It could be **Natural Indexing Language** in which the indexer uses the exact words and phrases used by the author of the document. This is very easy to use by the indexer and the searcher but the major problem is that there is no discrimination between synonyms, semantics, homographs, singular and plurals. This type of indexing tends to scatter documents on the same subject, where the authors have used different terms. Natural indexing language is used mainly in the back of book index and computerized indexes such as Keyword in Context (KWIC) and Key Word out of Context (KWOC) indexes.

Another type of indexing language is the **Free Indexing Language.** In this type of indexing language, there is no restriction as to the words or phrases used by the authors or some other words. Both types of indexing languages are very suitable for computer produced indexes. Perhaps the most important types of indexing language is the **Controlled Indexing Language**. In this type of language, the indexer exercises some control over the terms that are to be used as index terms because the indexer assigns only terms that have been listed as possible index terms. There is generally a preconceived standard list of terms to be used for a particular system. Thus, when and indexer has identified terms that represent the document, he/she will consult this standard list to ensure that the terms used are consistent. There are two types of this standard list. This list is sometimes called an authority list. The first type is the alphabetical controlled list in which the terms are arranged alphabetically. The two common examples are subject headings list and thesauri. The second type is the classification scheme which assigns notation to subject terms.

3.3 THEORIES OF INDEXING

Some theories for explaining the process of indexing does exist although information scientists differ in accepting some of these views. Different researches have but Fugmann (1993) proposed a theory of indexing based on five general axioms

1) The axiom of definability proposes that compiling information relevant to a topic can only be accomplished to the degree to which a topic can be defined.
2) The axiom of order suggests that any compilation of information relevant to a topic is an order creation process.
3) The axiom of sufficient degree of order posits that the demands made on the degree of order increases as the size of a collection and frequency of searches increase.
4) The axiom of predictability says that the success of any directed search for relevant information hinges on how readily predictable are the modes of expression for concepts and statements in the search file.
5) The axiom of fidelity equates the success of any directed search for relevant information with the fidelity with which concepts and statements are expressed in the search file.

SELF ASSESSMENT

What are the theories of Indexing?

4.0 CONCLUSION

In this unit you have learned about the concept of Indexing, levels and theories of the indexing processes and the different types of Indexing Languages.

5.0 SUMMARY

What you have learnt in this unit is that

Indexing is the process of providing a guide to the intellectual content of a document or a collection of documents and the theories of indexing are:

- The axiom of definability
- The axiom of order
- The axiom of sufficient degree of order
- The axiom of predictability
- The axiom of fidelity

Indexing language is made up of words or descriptors that are used in the intellectual contents of documents. These terms are expected to be used by the searcher in order to search for documents in collection. Examples of these indexing languages are Natural, free and controlled Languages.

6.0 TUTOR-MARKED ASSIGNMENT

Define the concept of Indexing and describe the different types of indexing Languages.

7.0 REFERNCES/FUTHER READINGS

- Aina, L.O. (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd
- Chowdhury G.G (1999) Introduction to Modern Information Retrieval. Facet Publishing. London
- Lancaster, F. W. (1991) Indexing and Abstracting: Theory & Practice. London: Library Association
- Fugmann, R. (1993) Subject analysis and indexing: theoretical foundation and Practical advice, Frankfurt : Indeks Verlag.
- Langridge, D.W. (1989) Subject Analysis: Principles & Procedures. London: Bowker-Saur.
- Wynar, Bohdan S. (1992) Introduction to Cataloging & Classification. 8th Edition. Englewood

UNIT 2: INDEXING SYSTEM

CONTENTS

1.0 INTRODUCTION

What you will learn in this unit concerns the types of indexing system, the techniques for indexing a document will equally be discussed.

2.0 OBJECTIVES

By the end of this unit, you should be able to

- describe the types of indexing system
- Discuss the subject indexes and its type
- explain techniques of indexing a document.

3.0 MAIN CONTENT

3.1 PRE-COORDINATE AND POST-COORDINATE SYSTEMS

There are two major types of indexing systems. These are pre-coordinate indexing and post-coordinate indexing.

In pre-coordinate indexing, a subject terms is chosen to represent a document which will serve as the lead term to that document. The document may contain one or more subject terms. One of the terms will be the lead term and the others will be coordinated with the lead term. Because the

coordination is done before searching by the user, such type of indexing is called “pre-coordinate indexing.” Post-coordinate indexing, on the other hand, involves breaking down a multi-concept subject into single concepts and the searcher would then combine the terms that represent the document required. There is no lead term. Each term is independent and can be combined to suit the interest of the searcher. The co-ordination is done at the time of searching.

A document entitled “The Influence of Britain in the Education of Librarians in Africa” is a multi-concept subject. It can be broken down into the following single concepts:

Librarians

Education

Britain

Africa

In pre-coordinate indexing, the lead term would be librarians because that is the main focus of the study. Thus pre-coordinate indexing can be done as follows:

Librarians – Education, Africa: Britain

But in post-coordinate indexing, the various terms can be combined independently as:

Librarians and Education

Africa and Britain

Any document in which these four terms are present would be retrieved. In pre-coordinate indexing, there is always a citation order, that is a prescribed order must be followed, whereas in post-coordinate indexing there is no need for a citation order: Printed indexes such as book indexes, printed indexes and abstracts, national bibliographies, subject catalogues of libraries, etc, are examples of pre-coordinate indexing.

SELF ASSESSMENT

Differentiate between a pre-coordinate indexing system and a post-coordinate indexing system.

3.2 TECHNIQUES FOR INDEXING DOCUMENTS

Indexing is an art that involves a number of stages. The first stage in indexing a document is to have a general idea of the document by going though the title, preface, foreword, content pages and possibly introduction. One can also flip through the text and make some spot reading. This will give the indexer sufficient familiarization with the document; hence this stage is called the ***familiarization*** stage. The indexer wants to know what the document is about by identifying concepts that are conveyed by words and phrases in the document, examining the title, abstract,

preface, introduction, chapter headings, major headings, sub-headings, etc. It is important that the indexer takes into account the needs of the users.

The next stage, which is the ***analysis***, involves the indexer using his intellectual judgment by identifying the concepts the book has treated. Sometimes the indexer may use the exact term used by the author or he might formulate an appropriate term. These terms are intended to accurately describe the whole document. The indexer at this stage is doing what is referred to as subject analysis or concept analysis.

This is where the subject background of the indexer comes into play, especially if he/she has a sufficient subject background of that document. At this stage, the terms identified by the indexer are what he/she judges to be the terms that represent the totality of the document. In a situation where the use of terminology is controlled, the indexer cannot use these terms directly as index terms or access points. Rather terms identified have to be translated into an indexing language used by the system which is the language used by both the indexer and the searcher in an information storage and retrieval process. This language exercises some control over what terms to be used as index terms.

During this stage, the indexer assigns subject descriptors chosen from the controlled language that the users of the discipline are familiar with. This stage is called the ***translation*** stage. However, in a setting where there is no need to exercise control over the terminology of the system, such as the bock of a book index or computerized indexes, this last stage may not be necessary.

4.0 CONCLUSION

In this unit you have learned about Pre- and Post-coordinate Systems of Indexing. You have also been introduced to the techniques for indexing a document.

5.0 SUMMARY

What you have learned in this unit concerns the types of indexing system and the techniques for indexing. In the next unit you shall learn about the evaluation of an index.

6.0 TUTOR-MARKED ASSIGNMENT

- Explain the pre- and post –coordinate systems of indexing.
- Describe the techniques in indexing a document.

7.0 REFERENCES/FUTHER READINGS

- Aina, L.O. (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd
- Lancaster, F. W. (1991) Indexing and Abstracting: Theory & Practice. London: Library Association
- Langridge, D.W. (1989) Subject Analysis: Principles & Procedures. London: Bowker-Saur.
- Wynar, Bohdan S. (1992) Introduction to Cataloging & Classification. 8th Edition. Englewood

UNIT 3: EVALUATION OF AN INDEX

CONTENTS

1.0 INTRODUCTION

Whatever the type of index that is produced, it needs to be evaluated in order to determine how effective the indexing has been in relation to how many documents that contain a particular term can be retrieved from the system. Also to be determined is how many of the documents retrieved from the system can be said to be relevant to the user who is interested in that term. Thus, a good index has a number of parameters by which it can be judged if it is good or not.

The unit describes the quality of an index and discusses the parameters for evaluating an index.

2.0 OBJECTIVES

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

- state the qualities of an index
- explain the parameters of evaluating an index.

3.0 MAIN CONTENT

3.1 QUALITY OF AN INDEX

If the indexer has done a good job during the indexing process, it would be expected that only relevant and specific documents would be retrieved within the shortest possible time, in which case the index could be said to be good. If, however, in the process of searching for relevant documents, a lot of difficulties are encountered, then we say the index is bad.

However, in determining the policies, certain features of indexing have to be explained. These include depth of indexing, specificity, exhaustively and weighting, etc. **Depth of Indexing** involves selecting as a large number of topics from a document, that is making as many important topics as are treated in a document as index terms for the document. **Specificity** involves selecting only terms that are specific to the document, which is a term that entirely covers the document.

Exhaustively on the other hand, is related to the number of concepts covered in a document that would be selected for indexing a document. Thus, the more topics selected from a document the more exhaustive the indexing is. **Weighting** is another important device employed by indexing agencies. This involves assigning weights to the various terms selected from the document, thereby showing their relative importance and then ranking the terms. Thus, if there are 10 terms selected from a document, and the indexing agency as a policy does not include more than five terms, it would be easy to select the five terms based on a weighting scale. The frequency of occurrence of words in a title or text of a document is a good way of weighting terms.

SELF ASSESSMENT

What are the qualities that make a good index?

3.2 EVALUATION OF AN INDEX

An index has to be evaluated in terms of its efficiency and effectiveness. Also, there are principal measures for evaluating the effectiveness of an index. These measures are:

Recall Ratio is a quantitative ratio of the number of relevant documents retrieved to the total number of relevant documents present in a collection. This is a quantitative measure used to determine the ability of an index as an aid to retrieving documents containing information on a particular request from a collection of documents present in a library of an information centre.

Precision Ratio is a quantitative ratio of the number of relevant documents retrieved to the total number of documents retrieved.

Precision Ration= No. of relevant documents retrieved x 100

No. of documents retrieved

Thus, if out of the 100 documents retrieved in the system, using the index prepared by Indexer D, only 35 are relevant to the user, the precision ratio of index is:

35 x 100 = 75% precision.

20

It is therefore obvious that there is an inverse relationship between recall ratio and precision ratio. The higher the recall ratio the lower the precision ratio and vice versa. The more documents that are recalled, the less precise the indexing system would be, and the less documents that are recalled, the more precise the indexing system is. Thus, the indexer must ensure a fair balance of recall ratio and precision ratio. We therefore expect about 70% recall ratio and 60% precision ratio.

It should be noted that specificity and exhaustively have influence on recall and precision ratios. When the indexing policy of a library or an indexing agency is to support exhaustively, then it would result in a high recall of documents and a low precision that is most of the documents recalled would not be relevant. On the other hand, when an indexing agency supports specificity, then the recall of documents would be low, but the precision would be high as only documents that are relevant to the user would have been recalled.

Specificity promotes low recall and high precision while exhaustively promotes high recall and low precision.

Time The main function of an index is to reduce the time it would take a use to retrieve documents in a collection. Thus, a good index is that which takes a minimum time to retrieve documents that are relevant and precise to the query. However, the time it takes a user to retrieve relevant documents does not depend solely on the index alone, the ability of the user to precisely select terms that appropriately describe the query is a factor in quickly retrieving relevant documents in a collection. Thus, if an index is good but the user has not used the appropriate descriptors, the user would take a longer time to retrieve relevant documents; but all things being

equal, a good index should enable a reader to use a short time to get relevant documents from the collection.

Cost A good index should be able to serve its purpose with minimum cost. Thus, a good index should be affordable to an average library. No matter how efficient an index, if it is costly, then it might not be available to an average library. No matter how efficient an index, if it is costly, then it might not be available to an average library, archives or information centre. When it is available a reader might have to subsidize the cost, which many readers might not be able to afford.

4.0 CONCLUSION

The ultimate aim of any index is to enable the user to use the index with minimum difficulties. In addition, a good index should be able to retrieve all documents needed for a particular query within the shortest possible period. If the indexer has chosen appropriate index terms to describe the documents in a collection, it would be possible for a user to retrieve all the relevant documents needed instantly.

5.0 SUMMARY

In this unit, you have learnt about the evaluation of an indexing system.

6.0 TUTOR-MARKED ASSIGNMENT

- What are the qualities of an effective index?
- Explain the parameters to be considered when evaluating an index.

7.0 REFERNCES/FUTHER READINGS

Aina, L.O. (2004) Library and Information Science Text for Africa. Nigeria: Third World Information Services Ltd

Lancaster, F. W. (1991) Indexing and Abstracting: Theory & Practice. London: Library Association

UNIT 4: AUTOMATIC INDEXING AND CLASSIFICATION

CONTENTS

1.0 INTRODUCTION

Now that you understand the concept of classification and indexing in the conventional way, it's time to learn about automatic classification and indexing processes. You will equally learn about classification and indexing beyond the traditional Library.

2.0 OBJECTIVES

By the end of this unit, the student should be able to:

- explain what is meant by automatic system of classification and indexing?
- describe the automated method of classifying information
- understand the concept of classification and indexing beyond the conventional setting.

3.0 MAIN CONTENT

3.1 AUTOMATIC CLASSIFICATION AND INDEXING METHODS

Subject access tools have been developed to organize and manage electronic resources. They are known by various names such as subject guides, Web guides, subject categories, subject directories, subject hierarchies, pathfinders, and so on. What many of these systems have in common is that they manifest the traditional classification principles of hierarchical structure, domain partition, subordination of the specific to the general, and array of related subjects.

Indexing and classification processes have been performed intellectually by human for quite a long time now. Automatic systems have been developed comparatively recently where classification and indexing are performed with the assistance of computers. Although the methods for representing the contents of document differ from system to system, the first task which is the analysis of the subject is the same in each case.

When the assignment of the content identifiers is carried out with the aid of modern computing equipment, the operation becomes Automatic Indexing. While term weighting lies in the heart of information retrieval many techniques have continue to be of interest to the process of information retrieval. Automatic classification is a multivariate statistical technique that groups together similar objects in a multidimensional space.

Although classification was mainly designed for organizing bibliographic items on shelves, in some cases attempts have been made to adapt existing schemes such as the Dewey Decimal Classification (DDC), Library of Congress Classification (LCC), and the Universal Decimal Classification (UDC) to the Web environment.

Many researchers have used library classification schemes for organizing information on the web. Typical examples are:

BUBL LINK: The subject terms used in BUBL LINK were originally based on LCSH; its users can gain access to the digital resources by a classified list or through an alphabetical list of subjects.

CYBERDEWEY: This is an example in the use of the DDC in organizing digital information resources. Users can select a subclass or topic to get access to the list the general information resource

CYBERSTACKS: is a centralized integrated and unified collection of selected digital resources using the LC class scheme. It allows users to browse through virtual library stacks containing monograph or serial works, files, databases or search services to identify relevant information. Resources are categorized first within a broad classification then within narrower subclasses, and resources are listed under a specific class. Each document record comprises a number of field and subfields, each one of which contains a particular unit of information which could be authors' name, publisher's name, title, keywords, number class, ISBN etc. The document record

may also contain an abstract or a full text of the document concerned. A text retrieval system is designed to provide fast access to records through any of the sought keys or access points.

In a way, the use of hierarchical or classificatory structure electronically is still relatively new. As information resources continue to grow, one may expect corresponding growth and refinement in ways to organize them. At this point in time, it is perhaps not too early to consider some of the functional requirements of information organizers. The desirable characteristics may be summarized as follows; a scheme designed for organizing digital resources should be:

- Intuitive, logical, and easy to use, with hierarchies and cross-references clearly displayed and with current and expressive captions;
- Flexible, adjustable, and expandable, to reflect rapidly changing and diverse environments;
- Useful in a wide range of settings, and applicable over a wide range of the number of sites to which it applies; and,
- Relatively easy to maintain and to revise.

4.0 CONCLUSION

In this unit you have learned about the automatic system of indexing and classification. You have also been able to identify the different methods of classifying information electronically.

5.0 SUMMARY

What you have learned in this unit concerns automatic indexing and classification

6.0 TUTOR-MARKED ASSIGNMENT

Discuss the concept and development in automatic indexing and classification

7.0 REFERNCES/FUTHER READINGS

- Chowdhury (1999) Introduction to Modern Information Retrieval. Facet Publishing. London
- Marcella Rita & Newton Robert. 1994 A New Manual of Classification. Gower Publishing limited, Aldershot.
- Juris Dilevko & Lisa Gottlieb (*2009)* .The relevance of classification theory to textual analysis Library *& Information Science Research, Volume 31(2), pp92-100*