ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
SCHOOL OF INFORMATION STUDIES
FOR AFRICA (SISA)

ASSESSMENT OF PADIS INFORMATION STORAGE AND RETRIEVAL SYSTEM:
The case of PADdev bibliographic database

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION SCIENCE

BY
Solomon Teferra

MAY 16, 1997

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

SCHOOL OF INFORMATION STUDIES FOR AFRICA

ASSESSMENT OF PADIS INFORMATION STORAGE AND RETRIEVAL SYSTEM: THE CASE OF PADDEV BIBLIOGRAPHIC DATABASE

By

Solomon Teferra Abate

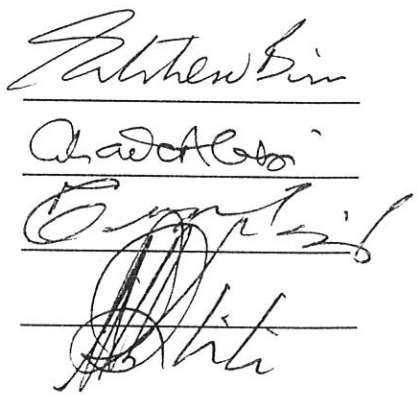
Name and Signature of Members of the Examining Board

Ato Getachew Birru, Chairman, Examining Board

Dr. G.A. Alabi, Advisor

Ato Tesfaye Biru, Internal Examiner

Prof. A.M. Kaniki, External Examiner



DEDICATED

To my
Grand mum Gorfinesh Gebre Mikael

ACKNOWLEDGMENTS

I am grateful to Dr. G.A. Alabi, my thesis advisor, for his continues assistance, comments and suggestion in the progress of this thesis. I am equally grateful to ato Gashaw Kebede for his invaluable comments.

My thanks also extend to all the staff of the Department of Library Science, specially to W/ro Ayneabeba Amenu, all the staff of SISA who help and encouraged me in the different works of the thesis.

I would like to extend my thank further to all the staff of PADIS who participated with me in the different data collection activities specially to Ato Lishan Adam, a Ph.D candidate, for his special concern and helpful advices.

Finally I wish to thank all those who stayed with me throughout this study especially my friend Abinet Sime who assisted me to overcome the king of the nigh, and my parents and all other relatives who gave me this moral support.

ABSTRACT

This thesis evaluates the performance of the different technical features of PADdev IR system from the point of view of professional system users, who are staff members of PADIS, requirements of development information handling, and the -state-of-the-art development in the IR technologies.

Features, merits and demerits of different types of databases and database development paradigms are reviewed. The development of IR technologies and IR features of ISIS packages are summarized, which highlights the deficiencies of the existing ISIS-based IR services. Search types and database formats are also discussed as factors of IR system's performance.

Establishment, objectives, services and databases of PADIS are also considered. The new IT-oriented plans of the Development Information Services Division are also discussed.

The data analysis and discussion of the existing features of PADdev IR system shows its present IR capabilities and what would have been considered to further enhance its performance. On the basis of the conclusions drawn from the different parts of the thesis, long-range and immediate recommendations are proposed.

TABLE OF CONTENTS

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

CHAPTER SEVEN

CHAPTER EIGHT

<< Table of Contents will generate here >>

CHAPTER ONE

INTRODUCTION

1.1. BACKGROUND

The term assessment is defined by English dictionaries as the act of judging the quality. With this understanding, the background of this study is given below.

This thesis assesses the capability and effectiveness/performance of information storage and retrieval system, that is to be referred to as IR throughout this thesis, implemented on the data base of development information on Africa (PADdev), created by pan-African Development Information System (PADIS).

PADIS is a cooperative regional development information system established in 1979 under the aegis of the United Nations Economic Commission for Africa (ECA)(Kangulu 1994). It is the result of the recognition of the poor state of numerical and non-numerical information and documentation resources and referral services in Africa by Resolution 360(XIV) adopted by the Economic Commission for Africa (ECA) conference of Ministers on the establishment of a Pan-African Information System.

The system became operational in 1980. Its main objective was to provide ready access to information required by policy makers, technicians, planners, and others engaged in the economic and social development of African countries (Kangulu 1994).

To achieve this main objective PADIS was aimed at, as indicated in its manual for document analysis (PADIS 1993b), establishing a regional information system in Africa which would serve as a conduit for information and data for development, centred around national, sub-

regional and regional networks to which all members would voluntarily contribute information and share their information management techniques. It has also the following immediate objectives.

1. enhancing the capability of member states to collect, store, retrieve and disseminate information on issues of African development
2. development and maintenance of a series of numerical and non-numerical databases for storage and dissemination of development information to be used by member states.
3. development and encouragement of acceptance of norms and standards for harmonizing development documentation and information
4. training personnel from member states, sub-regional and regional institutions in the introduction and utilization of up-to-date methods of development information processing and dissemination.

Structurally, PADIS is part of the UN system. It receives financial and technical support from the UN regular budget for the ECA, African member States of the ECA, the UN Development Programme (UNDP) and other multi-lateral and bi-lateral donors like, unesco and IDRC (PADIS 1993b). Recently, it is within the Development Information Service Division that includes: former PADIS, Statistics Division, Library, Cartography and remote sensing, and Publication and Reproduction unit.

In line with the above general and specific objectives PADIS is responsible to play the following major roles(Kangulu 1994:208)

- promote bibliographical control of development information in and about Africa using the territorial formula, each African country is expected to record development information produced in and about it and forward this information to PADIS Central Coordination Office (PADIS-CCO) in Addis Ababa, for inclusion in PADdev.
- Increase accessibility to development information in Africa and elsewhere by providing such services as retrospective bibliographical searches, selective dissemination of information(SDI), question and answer services, and
- provide training in modern information handling techniques and use of development information.

However, Kangulu(1994) concluded that:

- PADdev is not comprehensive in that there is lack of input from many national coordinating centres, there is delay in merging inputs received from national centres and

there is also deliberate exclusion of materials about Africa produced outside Africa covered by other databases.

- Some of the PADIS methodologies are outdated, i.e., the PADIS methodologies used in recording information are difficult and unnecessarily long.
- Recording of development information has not yet been standardised. This is to mean that standardisation of information recording has not been achieved at any level, even at PADIS.

He also recommended that:

- Core database, PADdev being one of the three databases that he has considered as core, should be given priority over others.
- Efforts should be directed at making PADdev as comprehensive as possible.
- access to development information in and about Africa should be improved via the improvement of PADdev.
- Newer technologies for the recording and dissemination of information should be explored.
- As priorities for PADIS, he suggested that PADIS should develop other methods for the recording and delivery of information.

The theoretical background of evaluation research shows that IR can be evaluated against the following criteria point of view, as has been identified by Cleverdon (Chowdhury 1994).

1. recall
2. precision
3. time lag
4. the effort, intellectual as well as physical, required from the user in obtaining answers to the search requests
5. the form of presentation of the search output which affects the user's ability to make use of the retrieved items
6. the coverage of the collection

Rowley(1992) also forwarded the following checklist for the evaluation of databases and text retrieval software.

1. Inputting data that includes:

screen format

editor

updating facility

field type

2. Indexing

3. Information retrieval including the help and search facilities

4. Output facilities

5. Security

6. Displaying the index

7. Setting up databases

8. Relational databases

9. User-friendliness

Using these criteria an IR can be evaluated either from the management or from the users point of view (Cleverdon 1978). The following are some of the functional requirements listed by Chowdhury (1994) that a user-oriented IR system should meet.

1. Information should be provided in a form suitable for ready access.
2. The scope of the system should be broad enough to cover all kinds of requirements of the users' community.
3. The system should be capable of providing right information at the right time.
4. Information provided should be easily accessible.
5. The system should be adaptable to environmental change.
6. The system should maintain standards for protection of data.
7. The system should devise ways for facilitating oral communication.
8. The system as a whole should be easily accessible and convenient to use.

This thesis therefore, assesses the fulfilment of some of Chowdhury's (1994) functional requirements. To this end, it adopts some of the evaluation criteria put forwarded by Cleverdon, as cited by Chowdhury(1994), and most of the check lists offered by Rowley (1992).

By 'user' David Bawden, as cited by Gilchrist (1996), means primarily the information

professionals who are the 'users' of the system in the sense of being searcher, intermediaries, operators, providers, maintainers or whatever, rather than the 'information requestor' or direct user. For him, although evaluation considers the ultimate information users, it rests squarely with information professionals. This thesis also considers PADIS' professional staff members as users of PADdev IR system. With this background in mind, the statements of the problem are discussed below.

1.2. STATEMENT OF THE PROBLEM

Development information systems handle data including numerical and non-numerical, like bibliographic, referrals as well as factual information relevant for idea formatting, planning, implementation, monitoring, evaluation, and research (PADIS 1993a). Africa's needs for data/information for socio-economic development planning fall into two broad categories: (1) strategic data/information for planning development and execution; and (2) data/information in the area of science and technology (PADIS 1996). Therefore, software packages required to handle development information include DBMSs, Statistical database packages, Bibliographic information Management packages, or software for exchange of development information through electronic networks (PADIS 1994b). However, a paper presented at the ninth session of the joint conference of African Planners, Statisticians, Population and Information Scientists attributed Africa's lack of development to the information poverty. The cause of the poverty, as the paper pointed out, is not lack of data but lack of the appropriate information infrastructure (PADIS 1996), through which the available data/information can be accessed. To be accessed data/information should be handled with an effective information retrieval system that stores all

the available data/information and retrieved as required. This initiated the researcher to assess PADdev IR system.

Any definition of the phrase Information Storage and Retrieval shows that the first thing that an IR system should do is database creation in a structured way. To this end, PADIS developed PADdev using its own format of record structure. In planning an appropriate record structure for the database, PADIS aimed at keeping the data entry procedures of the format as simple as possible, avoiding the repetitive entry of identical material whilst at the same time providing the most comprehensive information possible (PADIS 1993b). In this regard, this thesis looks into the information recording capacity of PADdev IR system.

PADdev is a database of reference/bibliographic type. It is obvious that bibliographic databases can only indicate the availability of information sources for a given information need. Yet, traditionally, much attention for retrieval purposes has been focused on the beginning of an article (title, author, abstract) and to its end (list of references). However, attention has now been drawn to the body of the article by the full-text retrieval systems (Meadows 1992). Although, in his conclusion Meadows pointed out the complexity of full-text retrieval systems, he indicated that all the limitations of traditional information retrieval like use of keyword, boolean combinations and concentration on documents are being transcended. Milet(1989) also noted that the dream of multimedia information retrieval systems visualized by Vannevar in 1945, which was called Memex, provided the conceptual basis for more sophisticated interactive information retrieval systems, such as hypermedia and object-oriented programming systems. Although the hardware limitations delayed the realization of these technologies, presently both the hardware and software technologies have developed enough to make them a reality. They have, therefore, started to optimize the user-machine interaction by maximizing the human cognition process (Ozkarahan 1995). In this regard, this thesis identifies the need for the advancement of the storage

the available data/information and retrieved as required. This initiated the researcher to assess PADdev IR system.

Any definition of the phrase Information Storage and Retrieval shows that the first thing that an IR system should do is database creation in a structured way. To this end, PADIS developed PADdev using its own format of record structure. In planning an appropriate record structure for the database, PADIS aimed at keeping the data entry procedures of the format as simple as possible, avoiding the repetitive entry of identical material whilst at the same time providing the most comprehensive information possible (PADIS 1993b). In this regard, this thesis looks into the information recording capacity of PADdev IR system.

PADdev is a database of reference/bibliographic type. It is obvious that bibliographic databases can only indicate the availability of information sources for a given information need. Yet, traditionally, much attention for retrieval purposes has been focused on the beginning of an article (title, author, abstract) and to its end (list of references). However, attention has now been drawn to the body of the article by the full-text retrieval systems (Meadows 1992). Although, in his conclusion Meadows pointed out the complexity of full-text retrieval systems, he indicated that all the limitations of traditional information retrieval like use of keyword, boolean combinations and concentration on documents are being transcended. Milet(1989) also noted that the dream of multimedia information retrieval systems visualized by Vannevar in 1945, which was called Memex, provided the conceptual basis for more sophisticated interactive information retrieval systems, such as hypermedia and object-oriented programming systems. Although the hardware limitations delayed the realization of these technologies, presently both the hardware and software technologies have developed enough to make them a reality. They have, therefore, started to optimize the user-machine interaction by maximizing the human cognition process (Ozkarahan 1995). In this regard, this thesis identifies the need for the advancement of the storage

and retrieval system of the PADdev database to the level of the state-of-the-art development from the intermediaries' point of view.

Despite all the developments and complexities of user requirements mentioned above, PADIS maintains its database using bibliographic packages of ISIS family.

In trying to answer the question most frequently being asked by computer specialists when implementing a documental database "Why micro ISIS and not a relational database management systems?", Deco and his co-authors (1995) enlisted a number of features of documental/textual databases that the relational DBMSs cannot handle. Moreover, a database software evaluation report(PADIS 1994a) presented a number of facilities available in ISIS as compared to xBase.

Despite their good features, due to the fact that packages of ISIS family do not use what is called the 'relational' model to structure databases, it would be cumbersome to handle a database of complex nature (PADIS 1994a) by them. According to a database software evaluation report (PADIS 1994a), the interface design tools and the field, variable, and record control techniques available in CDS/ISIS are limited, and would make complex data entry or analysis schemes very difficult to implement. The report also shows that ISIS has also the limitation of handling non-text data types like dates, images, sounds, floats, memos because it is specialized for bibliographic databases. Moreover, windowing, mouse-driven interface techniques standard to MS windows, Macintosh, and xwindows operating shells are not currently available in CDS/ISIS (PADIS 1994a). In this view the thesis tries to assess if these limitations have impeded the storage and retrieval performance of PADdev database and their impact on the futurity of the database.

In this line, showing the development of complication and user requirement even in the bibliographic world, Heaney (1995) proposes the revision of MARC and AACR2 into their

object-oriented version so that they can accommodate the natural generalizations required and the dynamic nature of the bibliographic databases. Moreover, according to Chen (1994), the benefit of an object-oriented IR approach is twofold:

1. IR systems are mostly long-life software, owing to the massive amount of stored information which is difficult to reorganize. After several years, the requirements of the software may be quite different from those in the initial stage. In this regard, persistence is more in object-oriented approach than the conventional ones.
   In this view the thesis opts to analyse the existing PADdev retrieval system's responsiveness to change.
2. The query-answering machinery of the deductive object-oriented database provides a good framework for retrieving information needed by the user.
   With the above, the need to study the retrieval capability and user friendliness of the system is justifiable.

On the basis of the analysis that will be made, the researcher attempts to recommend possibilities by which PADIS gains the benefits of the technological development in the area of IR to alleviate the problems that may hinder the performance of its IR service, if any.

In view of the above-stated problems this study was set out to achieve the following objectives.

1.3. OBJECTIVES

The general objective of the study is to analyse the technical features of the existing information retrieval system of PADdev database in view of input, storage and retrieval and output performance; level of development, i.e., the state-of-the-art-development of the retrieval technology, and to make suggestions that may further enhance the useability of the system if found necessary.

Specifically the study has the following objectives

1. To analyse the effect of PADIS methodology on the content of PADdev in the type of information(bibliographic, full-text, sound, image), format of information sources, and the development of the PADIS information retrieval system implemented on PADdev database.
2. To explore the effect of the inherent limitations of ISIS family retrieval packages on the performance of PADIS information retrieval system and how to overcome these effects, if any.
3. To analyse the user-friendliness of PADdev IR system.
4. To measure the effectiveness from the professional intermediaries point of view.
5. To analyse the display/output format of the system.
6. To explore the need to develop the system to the level of full text, hypermedia/hypertext and/or object-oriented information retrieval system.
7. To make suggestions that may further enhance the performance of the IR system as per requirements that may come out of this research and the-state-of-the-art development in the area of information retrieval technology.

1.4. METHODOLOGY

Smithson(1994) noted that evaluation of an information retrieval system essentially involves a twofold choice between carefully controlled, essentially repeatable experiments resembling those of the natural sciences, as opposed to an interpretive approach. That is, the

choice is between the two paradigm: the traditional system-oriented view and the more recent user-oriented one. Robertson and Hancock-Beaulieu (1992) also noted that the conflict between laboratory and operational (user-oriented) experiments is essentially conflict, on the one hand, control over experimental variables, observability, and repeatability, and on the other hand, realism.

In the light of the necessity for practical evaluation, Smithson(1994) preferred the user-oriented paradigm of evaluation research. This study also adopts user-oriented paradigm to measure IR effectiveness; to evaluate features like user-friendliness, display format, search type; and to explore future development of the system under investigation. With such an evaluation research paradigm and van Rijsbergen's (1975) factors of retrieval effectiveness - the response time of the system, the presentation of the output, and the effort the user of the IR system has to invest - in mind, in addition to the input capability/flexibility, this research adopts survey methods of data collection. The techniques utilized include questionnaire, interview, observation, discussion and document analysis.

These data collection techniques are to be implemented on the professional users of the system at PADIS, as system operators, maintainers etc, who are considered as a whole.

The analysis of the collected data is undertaken using the Statistical Packages for Social Sciences(SPSS) as required. The graphical and tabular results are accompanied with the required interpretation. The expert judgement of the researcher will be presented as a discussion.

1.5. SIGNIFICANCE OF THE STUDY

We can make a reasonable cause for the importance of evaluation, under three headings (Gilchrist 1996) two of which are applicable to this thesis. These two are:

1. Provision of better products and services.

According to Gilchrist, without some formal, or semi-formal, evaluation there could never be any clear understanding of the deficiencies of existing products and services, and hence of means for their improvement.

2. Improved understanding of systems.

Gilchrist explains this as follows: Evaluation, by both operators and users, is the only means of ensuring a reliable and in-depth knowledge of what the system actually does, and how it is best used.

This research is significant in this sense that it can uncover the deficiencies of the system under investigation, for no system can totally be perfect, and hence the means of their improvement. It is also believed to uncover what the system actually does, what it is supposed to do and how it can best be used and make it more understood by both the operators and users.

More significantly, findings of such a research made on such a regional IR system is believed to be applicable to the other IR systems inside and outside the immediate environment. Its improvement will, therefore, be beneficial to any institution working towards developing an information retrieval system.

Deficiency improvements of development IR systems, if any, directly improves development research productivity of the area which is the determinant factor of actual development. 'Information anxiety' of decision makers can also be reduced through improvements

of development IR systems. This research is, therefore, believed to contribute towards the development of the African information system for decision making at policy level.

Moreover Information Professionals who want to further explore the improvement of African information retrieval systems will have a reference as a guide.

1.6. SCOPE AND LIMITATION OF THE STUDY

The scope of this study is limited to evaluate the retrieval packages implemented on the PADdev database without any regard to the collection exhaustivness, of the database. This issue has been investigated by Kangulu(1994). It is planned to evaluate the system's input capability, search facility, retrieval effectiveness, and output capability.

Although it is a limitation that should be admitted, the study could not consider information requestors' view in the evaluation because most information requestor are scattered all over Africa.

1.7. ORGANIZATION OF THE STUDY

In line with its objectives the study is organized in the following way. Its first and introductory chapter deals with the background, statements of the problem, the methodology it adopted and its significance. This part also states the scope and limitation of the study. The last section of this part is the organizational layout of the study.

The next four chapters deal with literature review of the background in the context of which the study evaluates the system. To this end, these chapters present reviews on different types of databases and their design approaches, IR systems in general and ISIS packages in particular, evaluation of IR systems, search types and database formats. The sixth chapter presents

a review on PADIS.

The analysis and discussion of the collected data are dealt with in the seventh chapter.

The eighth chapter presents the conclusions and recommendations that come out of the other parts of the study.

CHAPTER TWO

DATABASES AND DATABASE DESIGN APPROACHES

2.1. INTRODUCTION

This chapter of the thesis reviews different types of databases and the approaches commonly adopted to design their structure. As databases are the basis of information to be retrieved, and as the type of information they contain, , a variety of them are presented. Databases also differ according to the structure of their records or files. Approach to the structure of a record or a file of a database evolved at different stages. Some of these approaches are presented in general terms along with their relative merits and demerits.

2.2. DATABASE TYPES

Townley (1978) sees a database or an information bank as a collection of entries each of which is about an entity, be it book or person, commodity or statistic, and about its essential characteristics. Townley, further pointed out that at the very least each entity must contain a means of identifying it uniquely and an organized description of it which may be used for searching. Representing an entity of a database for an IR system, the information to be moved in the system is said to include documents; information about documents; information contained in documents; and information about the information in documents; specialist skills, and experience or sources of information (Townley 1978). Based on which of these elements a database contain, it may be either a reference or a source database (Rowley 1993: 68; Convey 1989:8). Rowley (1993: 68-69) pointed out that reference databases refer or point the user to other sources such as

a document, an organization or an individual for additional information, or for the full text of a document. This category comprises: bibliographic databases, catalogue databases and referral databases. On the other hand, Rowley(1993:9) sees source databases as databases of the original source data which may be either numeric databases, full-text databases or text-numeric databases. Some of these are discussed below.

2.2.1 BIBLIOGRAPHIC DATABASES

Convey (1989:3) pointed out that bibliographic databases give the main bibliographic details to uniquely identify the paper, book, etc. referred to and will normally give additional subject information like keyword, classification numbers and in many cases an abstract or summary. Describing bibliographic databases as that contain a series of linked bibliographic records, Rowley (1993:69) pointed out some of the typical components, including full text, to be contained by a record. The list is appended at the end of this thesis.

2.2.2 FULL TEXT DATABASES

Full text database is the major type of source databases. Defining source databases as electronic documents, Rowley (1993:73) said that their content may be as varied as the contents of the printed book, and many include text, numbers, tables, figures and graphics. As opposed to the printed book they take advantages of the fact that they are not constrained by physical limitations, and can be multimedia, images, sound, maps and charts with a flexible structure of their contents. On the basis of their content, Sievert (1996) mentioned three types of full text

databases.

1. Factual or directory databases in which each entry is divided into fields but mirror online the entire text.
2. Hybrids of full text and bibliographic citations plus abstracts.
3. Hybrids but every word in the portion of the text is searchable and can function as keys for retrieval.

The second and third types are good alternatives to developing a bibliographic database, (like PADdev) to a full text one, because, PADdev is already of the third type, except the fact that its abstract is smaller in size. Developing PADdev into the fourth type avoids the need to manage the full text files separately by enabling simply append the full text to the existing bibliographic records. It is also made easier by the development in the media especially CD-ROM, and text processing algorithms.

As a solution for the tedious document input process of full text databases Spinello (1991) suggests the proliferating optical scanners and Optical Character Recognition (OCR) devices. For the hardware limitations in text scanning Rowley (1992: 442-443) proposed the consideration of the three main options:

1. Database machines special purpose computers.
2. Array processors that rely upon parallel processing
3. Transputer networks that may be either Multiple-Instruction stream or Multiple-Data stream machines.

2.2.3 STATISTICAL/NUMERIC DATABASES

The other type of source databases is the numeric or statistical database. Convey (1989:10) noted that numeric databases consist of information expressed primarily as numbers presented in the form of tables or 'time series'. They usually comprise a measurement (population or quantity) a time-series (statistics collected over a period of time) and a variable (production, imports, exports etc. of a specific commodity). Convey (1989:11) further noted that numeric databases often process some computational features where by the retrieved data can be analysed and manipulated using some statistical techniques.

2.2.5 CENTRALIZED AND DECENTRALIZED OR DISTRIBUTED DATABASE SYSTEMS

Rowley (1993:66) noted that a database system may comprise a number of linked databases. The databases may be at different geographical location being managed centrally, located centrally and manipulated by distributed workstations, or located at different locations and managed by distributed processors but made accessible to all the networked workstations.

Data that is physically dispersed but centrally managed is the essence of distributed database technology (Krasowski 1991). Distributed DBMSs (DDBMSs) can improve the performance of such databases by allowing the data to be replicated in an orderly way and accessible by each database servers as if it is stored locally.

Krasowski listed about 13 rules that should be supported by a DDBMS. Most of the rules provide autonomy to the distinct DBMS making them independent of the central site, location,

fragmentation, replication, hardware, operating system, network, other DBMS of the system and require the work stations' database servers work continuously allowing them to send query to different sites.

With regard to handling different types of databases by an information retrieval system Neelameghan, as cited by Chowdhury (1996), has advocated an integrated database structure that would, within one structure, be able to hold the different data types. Chowdhury (1996) also pointed out that one of the major objectives of an integrated database is that users can, with one query, search for and retrieve different types of records (bibliographic, non-bibliographic) on the same subject. Furthermore, he noted the different ways of database integration. Some of them are: integrating separate databases containing different types of records into a single database; functional integration, whereby the output of one process is used as the input to another; searching separate databases as if they were one through the use of capable software to do so. Neelameghan's way of integration, which is the same as the first option given above requires to have a bibliographic format like that of ABNCD that allows the system to handle/create integrated databases.

2.3. DATABASE DESIGN APPROACHES

The other factor that differentiates one database from another is the approach followed to design the structure of the information to be kept. Chowdhury (1994:1) pointed out that the concept of information presupposes that there are some items of information which have been kept organized in any suitable order for easy retrieval. Different information items need to be ordered differently to be retrieved easily. In reality there are information items that can be kept in a very structured way and in the same order for a long period of time; there are also information

items that cannot be given any structured order. Very importantly most of the real information items need to be kept in a very dynamically changing order. In searching for a suitable way of keeping information, during the first three decades of database technology, DBMS software has evolved through four generations: file, hierarchical, network, and relational (Lin 1991). Because of its ability to handle different relations between information items without redundancy and the ease of systems development from the data access, data share and programming perspective the relational approach took over the market of non-relational approaches. But its inability to handle the non-structured information items such as primary data as documents or manuscripts resulted textual or non-structured approaches to database design. However, both these approaches are not able to handle the dynamically changing relationships that exist between information items as it exists in the real world. As a solution to this and other limitations, the object-oriented approach has come into the picture. These three competing approaches in the world of database design are discussed below.

2.3.1 RELATIONAL APPROACH TO DATABASE DESIGN

Motivated by the need for a more declarative data model, Dr. E.F. Codd in the early 1970s came up with the relational data model (Khoshafian 1993: 121). The concept of a relational approach to IR was, of course, put forward by Vannever Bush in 1945(Ellis 1990). In relational systems, information items are held in a set of relations or tables, each table is composed of rows which is equivalent to a record of non-relational database structure. Similarly the columns in a table are equivalent to what we call fields in the non-relational approaches. Citing Brackett, Neelameghan and Rahel (1994:18), pointed out that a relational database consists of three parts:

the structural part that defines relations of data and their interrelations; the integrity part to assess the uniqueness of each occurrence of a relation; the manipulative part which provides operators for processing relations. In a relational database environment it is the DBMS, not the programmer, that interfaces directly with the operating system's access methods and therefore, the programmer does not need to know the different file organization and access techniques. Relational DBMSs also embed SQL (Structured Query Language) statements in their programs to perform the I/O (Input and Output) tasks.

Regarding their popularity, Khoshafian (1993: 121) noted that in terms of new purchases and installation of DBMSs, relational databases dominate all other DBMSs combined. He also expects the trend to continue well into the 21st century.

2.3.2 TEXTUAL DATABASE DESIGN APPROACH

Not all information items in the real world can be persistently put into a set of tables with rows and columns. Factual information items, which consist the largest proportion of information required in any organization (Chowdhury 1996) are good examples. The second main type of information required is bibliographic or textual in nature (Chowdhury 1996) which is also the other non-structured information. The textual database design approach came to handle such non-structured information items. The textual database management systems allow the construction and administration of databases which are basically built on texts, and their main characteristic is that they are specifically designed to manage variable-length fields, and thus records of variable length (Deco et al. 1995). Besides, according to Rowley (1992: 399-400), it is their sophisticated searching facilities that distinguish text based systems from other database systems. But we should note that the sophistication came out of the need to search and retrieve information from

non-structured databases. Sin (1995: 73-74) also listed the basic distinguishing characteristics of information retrieval software, which is based on textual database design approach:

- Variable-length fields
- Access to records through an inverted file index keys or text terms which are drawn from the records that lie in the database.
- A range of retrieval facilities which support retrieval based on words in records, where there is limited control over the form of the search key in the record.
- Emphasis on the management of one or more distinct databases, where the ability to draw data from a number of related databases is not critical.
- Fixed applications which require relatively limited programming or system development facilities.

Both of the above-discussed approaches do have their own limitations and relative strengths. With this understanding the designers of the TINman database and some other databases adopted a hybrid of the inverted file (textual) approach and that of the relational database. To utilize the advantages of the two, without their disadvantages, TINman is structured in a multi-linked way (Ellis 1990:98). Yet there remained problems of representing all the information items as they exist in reality into the database world. These problems called upon a new approach, that is, Object-oriented.

2.3.3 OBJECT-ORIENTED DATABASE DESIGN APPROACH

Object orientation is a result of users' (database users) need to store more complex information element types as they are stored in the real world. Reid (1996) pointed out that databases are increasingly becoming Object-Oriented (OO) as users need to store more complex data types. He also noted that the past few years have witnessed the proliferation of object-oriented technologies into every field of computer science: programming languages, simulation, graphical user interfaces, and of course, databases. As that of the relational (structured approach)

the object-oriented approach is introduced by programming language designers. Khoshafian (1993:9) noted that it is the designers of the language Simula 67 who introduced the concept object that conceptually contained both data and the operations that manipulate the data. Moreover simula incorporated the notion of classes and supported inheritance. The concept of data abstraction with the aspects of clustering and information hiding was also introduced by language designers who needed to manage large programs (Khoshafian 1993:9). These important concepts of object orientation have been reaped by the small talk programming language development.

All these attractive outlooks of language designers have been adopted by database designers. Object-oriented databases, therefore, integrated object orientation with database capabilities (Khoshafian 1993:3). Defining object orientation loosely as the software modelling and development disciplines that make it easy to construct complex systems out of individual components, Khoshafian (1993:3) related it to the summation of abstract data typing, inheritance and object identity. On the other hand database capabilities are the summation of persistence, concurrency, transactions, recovery, querying, versionning, integrity, security and performance. This in effect means that Object-oriented databases have got all these features, the former from object orientation and the latter from database capability.

Along with the research development, the software industry accepted the trend growing towards object orientation, with a widely varying implementation approaches (Reid 1996). Reid (1996) further noted that some companies develop pure object-oriented databases (OODBMS) from the ground up, while others prefer to gradually enhance existing relational architectures with object capabilities. From their experience Neelameghan and Rahel (1994) also said that object-oriented Databases (OODBs) can be developed with some of the textual DBMSs, such as MicroIsis.

Pointing out the fact that more than 10 emerging companies characterize their products as object-oriented database systems in the first mid of 1990s, Khoshafian (1993) indicated that a number of relational database vendors started incorporating various object-oriented database features into their next generation products.

A buyer's guide (object-oriented and object relational DBMSs 1996)also presented a brief description of 15 object relational and object-oriented DBMSs.

2.3.4 MERITS AND DEMERITS OF DATABASE DESIGN APPROACHES

The basic objective of a database designer is to give ways of creating a database that consumes/requires the smallest storage and processing space possible and can be maintained with the least effort and time, more importantly, that allows/facilitates the fastest retrieval of the required information item.

A database once designed should also be responsive to the rapidly changing technology and user requirement in terms of the kind of information items users want to store, the way the users want to access what has been stored and generally the way they want to interact with the system. The security of a database should also be as much reliable as possible, allowing, at the same time, several users to share it. In these and some other regards, each of the above discussed approaches to database design has its own merits and demerits.

Comparing, technically, the ISIS-based databases that represent textual approach with xBase as representatives of the relational approach form a technical point of view, a report on evaluation of Microcomputer Database Software (PADIS 1994a) discussed the following relative limitations and strengths of these approaches.

1. Very complex database structures and interfaces.

In ISIS it would be very cumbersome to manage record of very complex nature, because any data/information item of the database resides in a single file. Whereas xBase databases typically are made up of many database files. Each file contains some particular list of entities which then relate to the other entities in complex ways. Rowley (1992) indicated the ability to work with several linked databases, drawing data from or adding data to several databases simultaneously as an important feature of relational DBMSs. This is, in effect, what we call database integration.

2. Non-text data types

Handling dates, images, floats and the like is difficult in CDS/ISIS and most of the text retrieval packages, based on textual database design approaches. Whereas for xBase handling such data is a simple operation, the images can be handled through different third party utility packages, and floats can be stored in their number forms and can be manipulated; there are also xBase third party packages to compute linear, non-linear, and nonparametric statistics, complex financial calculations, simultaneous equation solution, and a wide variety of other analytical operations. The report also noted some features that are native to ISIS but can be possible only through third party packages in xBases. These are: Delimiters, ANY terms, Hit lists, Boolean algebra, Stop words, formatting language.

Comparing the relational databases with the textual databases Deco et. al(1995) explained the reasons that make textual database management systems better than the relational database management systems in the following points for the implementation of textual databases:

- Previous knowledge of the structure of information, i.e. to search relational databases, knowledge of the structure of information is mandatory, as opposed to textual databases.
- The search for a word in a complete document i.e. the search in relational databases is

field based, but a word can be searched in textual databases in the whole document.

- Document qualification

Document ranking is not possible in relational databases but can be made in the textual ones

- Predetermined queries

Textual databases can handle/respond to ad hoc queries which is not the case in the relational ones

- Field repetition

Repeated fields are not allowed in relational databases.

- Words

Some concepts which the textual databases can manage are very complicated for the relational databases to implement.

- Synonyms

relational databases cannot manage synonyms whereas it is implemented in many textual databases.

- Adjacency

the concept of adjacency search is not possible in the relational databases while it is possible to search requests that may be separated up to five words in textual databases.

In conclusion Deco et. al,(1995) observed that there are characteristics, such as the words, synonyms, and adjacencies concepts, as well as inverted files and field repeatability, which cannot be implemented in a relational systems but are fundamental for the retrieval of textual information. On the other hand, textual systems associate or join different records by means of a common field, which is an essential characteristic in the relational systems and is of great

importance so as to be able to manage more than one databases.

Although it allows fast access, a limitation to an index-based search-the fundamental feature of textual databases - is its high consumption of disk space by the indexes (Sievert 1996). On the other hand, without extensions, relational technology lacks the fundamental components needed to support integrated applications that will increasingly use such multimedia data types as long text fields, cluster and vector screen images, voice data, and Video. They cannot also provide direct, natural representation of graphic-structured object spaces (Khoshafian 1993:Vi) which is also the case of textual databases.

In this regard object-oriented approach has the following merits.

- removes the semantic gap between an application domain and its representation in persistent storage. Since the real world is modelled as closely as possible, the links and relationships among entities in it are represented and manipulated directly (Khoshafian 1993:Vi);
- Alleviate the impedance mismatch between programming languages and database management systems.
- Allows, new object-oriented databases programmers to define new data types with their own defined rules as to how they will work. These are just as valid as the ones that come with the databases (Reid 1996);
- Store and manipulate multimedia objects;
- Has open architecture.

As the other two approaches/databases object-oriented databases also have some deficiencies that include:

- The basic mechanics of pure object-oriented databases, such as traversal of objects through object IDs (OIDs) and some inability to create new objects on the fly (such as the

result of a SQL query), have negatively impacted ad hoc query capability.

- Most security models have been designed for relational databases. Because the database model itself is more complex than the relational model, and It is not based on a formal (mathematical) model (Oliver and Salms 1994).

PADdev is a bibliographic database that is developed and being handled according to the paradigm of textual database development. Therefore, it is limited to have only the good features of its category and suffer from the deficiency of the other good features associated with the other categories discussed above.

CHAPTER THREE

INFORMATION STORAGE AND RETRIEVAL SYSTEMS

Information Storage and Retrieval (IR) system is one that stores information to databases and retrieves information from them. Basically research foundations and developments in the area of IR are discussed with special emphasis to those dealing with textual databases. To be specific to an IR family adopted by the research area of the thesis, ISIS family IR system is specifically discussed with a special emphasis to CDS/ISIS.

The ultimate goal of collecting related items of information into a database is retrieving them back when required. Information retrieval using a computer is the searching, and the retrieving of, selected information from the database held on a computer (Convey 1989:3). Recognizing the fact that information retrieval systems have become almost synonymous with computers, Rowley (1993:113) argues for the existence of manual IR systems. Taking up the recognized fact, this thesis considers only the computerized IR systems and reviews what it is along with the historical development.

3.1. WHAT IS AN IR SYSTEM ?

With no disrespect to Calvin Mooers who coined the term information retrieval, Gilchrist (1996) preferred the term computer assisted bibliographic reference retrieval (with no guarantee of being able to find and acquire the full document) for the sense of using computers to store and retrieve only bibliographic information. Recognizing the fast development of IR technology, Gilchrist (1996)also pointed out that today's IR includes surrogate retrieval, text retrieval, document retrieval and, if only in a rudimentary way, picture retrieval. In all of these cases, the

major tasks that an IR system is designed to perform are database creation, analysis of user's queries, preparation of a strategy to search the database, actual searching or matching of the user's queries with the database, and retrieving items which fully or partially match with the search statement (Chowdhury 1994:2). As any other systems the functions of an IR system are accomplished by its different components. Three of the six components of IR identified by Lancaster, as cited by Chowdhury (1994:2) are:

1. The searching

2. the user-system interface and

3. The matching subsystems

To these we can add the database creating and output presenting components.

Rowley (1993:117) defined a searching language as the terms that are used by a searcher when specifying a search requirement. The searching component of an IR system facilitates all or more of the following search logics (Simui 1995: 76-78).

- Boolean search
- Adjacency /proximity searching
- Search refinement
- Transaction and searching on word stems
- Range searching . . .
- Free text scan
- Browse on index . .
- Hypertext
- Text string searching
- Artificial intelligence
- Concept-based search

Greethananda and Perera, cited by Simui (1995:76), defined the user interface of IR system as the surface on which the user interacts with the system. Perera, again cited by Simui (1995:76), gives an illustration of a search interface as consisting of search language, screen layouts, on screen instructions and printed manuals of the particular software package, with which

the user has to interact in order to retrieve information. Some of the interface options mentioned by Rowley (1993:117) are:

- Menu selection
- Command language
- WIMP interfaces
- Form filling
- Question and answer

The first two have been given emphasis in IR systems.

User-friendly interfaces have been developed in order to make a system easier for the user both to input data to the system and to use the system output. Facets of user-friendly interface development include the following(Rowley 1993:6-7):

- options on interface so that both novice and expert modes are available to the user
- the use of a graphical user interface including windows and direct manipulation (this makes some processes easier to conduct, and allows additional information such as authority files or help systems to be displayed on the screen alongside the data being entered
- special system that adapt user-friendly interfaces in specific environments, such as OPACs
- library-definable interfaces, such as library-definable OPACs
- report generators, which are becoming increasingly flexible.

The matching subsystem is the component that actually determines the similarity between the users' inquiry and the required information elements in the database. The information elements to be matched with users' inquiry may be either indexed or exist freely in the text. Rowley (1993:117-118) defined an indexing language as the terms or codes that might be used as

access points in an index. She further noted that it may be one of the following:

- Controlled indexing languages or assigned term system

- Natural indexing languages or derived term systems

- Free indexing languages

The terms or codes of the indexing language may be arranged in different indexing approaches which may be:

- inverted file structure in which every list contains only one record;

- a ring which is a linear list that closes upon itself;

- a multi-list in which there is only one list per keyword and chained together to form a list, the start of which is given in the directory;

- links and role indicators that may be used to formulate a more discriminating query; or

- hierarchy that provides a framework for the indexer and the searcher.

The output component of IR system provides facilities for (Simui 1995; 78):

- Screen display
- Saving, including formatting
- Printing, including formatting and content limiting.
- Current Awareness services
- Selective Dissemination of Information services
- Index or catalogue printing etc.

The features that we see in any one of IR system components discussed above are the results of developments that happened through time. The following part reviews the historical development of IR system.

3.2. HISTORICAL DEVELOPMENT

Chowdhury (1994: 193-196) dates the history of computerized IR back to the birth of modern computers. He summarized the history of IR as follows:

1950s (early) - Robert Fairthorne investigated the punched card equipment of retrieval of bibliographic information

1960s (early) - The coming up of proposals of techniques, systems and models. But they were not fruitful because of less development of computational power.

1970s - Mathematically sophisticated IR models like vector space and probabilistic.

1980s - String, keyword and keyword frequency searching were proved to be efficient

- Natural language processing started to be considered to alleviate problems that arise by textual databases.

The development in the 1990s will be discussed in the next part together with their contribution towards the improvement in IR performance.

3.3. RECENT DEVELOPMENTS OF IR SYSTEMS

IR software are continuously being revised to provide for improved functions that enhance user friendliness, fast access and retrieval of information (Simui 1995:87). In this regard, the emphasis in the development of IR system in 1990s has been given to full text database handling; natural language advancement; improvement of user interfaces; application of artificial intelligence, hypertext/hypermedia, best-match searching, and object orientation.

With the fulfilment of the required hardware facilities IR has moved beyond such limited early functionality on homogeneous collections of such textual objects as bibliographic references (Smeaton 1994). Even at the beginning of 1990s a growing number of organizations have successfully implemented full text retrieval systems, (Spinello 1991). It has been further enhanced by the rapidly growing use of text retrieval applications in the microcomputer environment which is supported by the advancement of compression techniques to store a large sized document in a small storage space.

Sievert (1996) evidenced the growth of full text retrieval systems by the increasing number of articles about full text databases making a CD-ROM database search on 'full text' and 'information retrieval' on five databases.

The need to search a full text database not by a single word or words artificially combined by boolean logic or else, but by meaningful phrases has given way to the development of natural language application to IR. In this regard Strzalkolkowski (1995) expressed his feeling that 'bag-of-words' representations, prevalent among today's information retrieval systems, can hardly do justice to the complexities of free, unprocessed text with which we have to deal.

Natural language IR system is aimed at enabling the searcher to express his/her information need in human way of expression. Smeaton (1994)indicated the levels at which Natural language can be developed and applied to IR, as : Lexical level, syntactic level, semantic level and pragmatic or Discourse-level analysis. Taking Mauljdin's work of 1991, Strzalkolkowski (1995) pointed out that more advanced techniques aimed at overcoming the limitations of shallow linguistic processing included semantic and discourse-level processing, although proved to be unacceptably expensive and difficult to evaluate.

The other struggle in the line parallel to the development of natural language processing

efforts was that made on the development of a cognitive approach to represent searcher's queries. Ellis (1990:67; 1994) noted that two features characterize cognitive approaches to IR system design: the construction of a model of the user in the system, and the derivation of these model from cognitive characteristics of the user. He cited Oddy's THOMAS program as the early work in the approach and Belkin's proposals for designing a retrieval system with reference to the concept of an Anomalous State of Knowledge (ASK).

According to Oddy, cited by Ellis (1990), rather than issuing orders to the computer system, retrieval was to be facilitated by means of a conversation or dialogue between the user and the system which was meant to resemble the personal communication between human minds through conversation. Because through conversation the user can get the information that he is likely to want to use for the reformulation of his query.

In this line, Rowley (1993:115) sees the capacity of dialogue design styles to influence the effectiveness of retrieval in the same rank as that of indexing and searching languages, search logic, and search facilities.

After a survey of about six systems that use different approaches to helping end users perform their own search, Gauch (1992) came to some conclusion. To mention a few

- They assist during the query formulation stage and/or the query reformulation stage
- They all provide user interfaces which shield the users from the mechanics of the search allowing them to concentrate on conceptual issues.

The highest wish to assist users of IR systems is the application of Artificial Intelligence. Spinello (1991) exemplified the approach by Information Text Management System (ITMS) Produced by Information Access Systems Incorporation (Boulder Co). He pointed out that the system incorporates human judgement into the query process and allows natural language queries that are based on ideas and concepts. Users can ask to find all the documents relevant to an idea

even though these documents might not contain those exact words. ITMS is able to accept conversational language requests, analyse the semantic content of those requests, and retrieve documents in order of their relevance to the request.

The other area of development in IR that is getting greater attention is hypertext. It had been coined by Ted Nelson to refer to non-linear text. The hypertexts, therefore, consist of the network of items with their association trails. Because retrieval by association that reflects the way in which the human mind follows upon connections between ideas is fundamental to the design philosophy of any hypertext system (Ellis 1990).

Spinello (1991) pointed out that hypertext is often combined with the inverted index method. So that users can perform a query for a particular search term from where they can follow the links to other nodes in the hyperdocument, which provide further information about the search term. Because these nodes might not contain explicit reference to the words used in the query, this information would not have been retrieved if the user were relying exclusively on the index. In this line, Rowley (1992:448) listed the following advantages of hypertext:

1. references can be followed forwards and backwards.
2. new references, annotations, comments or links can be created so that the user can develop a personalized network of links.
3. hierarchical and non-hierarchical structuring of information is possible.
4. documents can be customized to suit different applications.
5. there is no need to duplicate points in different places.

To extend these advantages in a multimedia environment, Croft and Turtle (1993) were working on integrating hypertext approach with object-oriented databases. Their hypertext would enable those nodes to be retrieved based on their relationships to nodes that do have text content. They aimed to support the retrieval of multimedia documents with complex structure.

As developments in the matching component of IR system Best-matching, concept-based retrieval and the free text scan methods are worth noting. Chowdhury (1994:76) pointed out that

a best match searching is designed to produce ranked output. It therefore, requires a method to measure the relative importance of the retrieved items which again requires some method of weighting of the search terms. The major attractions of the best-match approach as compared to boolean approach, listed by Rowley (1992:445-6)are:

1. There is no need to specify boolean inter-connections
2. No problems with the volume of outputs
3. Weighting information is possible

Prior to the best-match, there were attempts to address problems attached to the boolean and inverted index retrieval methods like the concept-based and free text scan methods. Queries in the concept-based retrieval system, are formulated according to topics or search objects that can be constructed by knowledge experts or the users themselves (Spinello 1991). According to spinello (1991), this method can be used to retrieve any document relevant to the concept or search object of a strategy. He further pointed out that this novel approach to text retrieval had been developed by Verity Incorporation and marketed under the name Topic.

Similarly, GESCAN, a product of GESCAN International Incorporation, is the best representative of the free text scan method of text retrieval (Spinello 1991). It relies on its own text array processor to accelerate the search process. The processor consists of 12 individual query processors, each of which is capable of performing sequential text searches at the rate of 5 million comparisons per second. Spinello (1991) argues that, depending on document size and complexity of queries there is little performance degradation in GESCAN as compared with the inverted index method.

In the recent days, along with the move from databases to the development of knowledge-based systems, object orientation of IR systems is taking the centre of attention. In this line, Yogesh Gupta, the senior president for product strategy at Computer Associates (CA)

International Inc. recalls how, in the late 1980s, the vendors of flatfile databases responded to the growing popularity of relational technology by extending their products in that direction and their failure including the biggest players. He sees, then, equally profound changes happening now, including moves towards object-oriented, multimedia and Internet computing (Hillary 1996).

As a response to object orientation's lack of mathematical basis, Bourdeau and Cheng (1995) defined a formal method as a mathematical approach to software development that begins with the construction of a formal specification describing the system under development. They presented a formal semantics for the Object Modelling Technique (OMT) object model notations, where an object model provides the basis for the architecture of an object-oriented system believing that formalization of object models contribute to a mathematical basis for deriving system designs.

The development of object orientation is also expected to pull the attention of IR systems designers from the system, as per, to the users. Neelameghan and Rahel (1994:47-48) have given the different reasons for this fact. Some of these reasons are:

- The data elements (usually attributes of objects and methods of manipulating them) that differ from one type of object to another, as opposed to the bibliographic data elements that can be normalized from cataloguing rules and others, need to be contributed by the users.
- Input data for a record in a non-bibliographic object-oriented database may be taken from different sources including remarks of individual users but not limited to documents.

3.4. NETWORKED IR SYSTEMS

Rowley, (1993: 115) noted that all types of information databases and many of the products derived from them may be maintained and produced either in-house or externally. In their functional aspects Chowdhury (1994:5) also identifies two types of IR systems: in-house that serves mainly the users of the parent organization of the system and online that gives access to remotely located databases to a wide variety of users. Rowley (1993:116) provided the following comparative summary of these two alternatives of access to databases.

Features	local system	External System
Coverage	less extensive of published literature but may include local documents.	probably cover greater numbers of documents in a wider discipline perspective
Cost	Costs are set up and maintenance costs, as such, may be quite high and are more likely to relate to size of database than to extent of use	costs are incurred per search, occasional searchers incur low total costs
Retrieval facilities	will suit local requirements, in terms of index terms, searchable fields and form of records.	may have similar range to local database, but more likely to be a greater range of more sophisticated facilities; user needs more skill in selection
Training for users	will vary depending on local environment-can be important role for information professionals	Training schemes are well established, but they, may be regarded as expensive .
Products	usually a more limited range of products since user requirements may be less diverse.	As full a range of products as can be marketed.
Users	smaller, more coherent group with less diversity in interests.	Larger group-discipline grouped, but employers will be various.

Of course Rowley emphasized the features of commercial online database for that of external systems compared above.

3.5. ISIS FAMILY INFORMATION RETRIEVAL PACKAGES

3.5.1. HISTORICAL DEVELOPMENT

MicroIsis is a program developed by International Labour Organization (ILO) in 1965 for use on IBM mainframe computers. On the basis of this program, passed to it, IDRC developed a minicomputer version of ISIS, known as MINISIS, for use on HP 3000 series of minicomputers.

In the early 1970s Unesco developed the mainframe version of CDS/ISIS software package by combining features of their Computerized Documentation System (CDS) and the Integrated Set of Information Systems (ISIS) of the ILO.

In 1985 Unesco produced the first version of micro CDS/ISIS IR software for IBM micro computers and compatibles. Unesco did very fast improvement on it and came up with two versions, version 2.3 and 2.32, in 1989 followed by the 1992 version 3.0 of the software. After a year in 1993-versions 3.03 and 3.07 were released. Presently the windows version of the package is on the way to be released.

3.5.2. Micro CDS/ISIS

Micro-CDS/ISIS is an IR system for information storage, processing and retrieval of structured textual databases. Unesco is distributing, micro CDS/ISIS free of charge for nonprofit making organizations. Due to this and some other good features of the software, an increasing number of Information Systems (ISs) of the developing countries are using it for the creation and manipulation of bibliographic referral type and factual databases. Moreover, Neelameghan, cited by Neelameghan and Rahel (1994:4), claimed that MicroIsis can be used in developing also object-oriented databases.

In CDS/ISIS all records of a given database are stored in the master file and are assigned a unique number automatically. The master file is associated with the inverted file through a cross-reference file. The inverted file consists of six physical files of which five contain dictionary of searchable terms organized in a B-tree.

Generally the files that are maintained by CDS/ISIS can be categorized into two, based on their purpose: system files and database files. According to their persistence and importance these files can also be classified into mandatory, auxiliary or user files.

The software performs its functions using two groups of programs: user programs and system programs. The user programs consist of different individual programs that perform data entry and record editing, information retrieval, printing output, and inverted file maintenance. The system programs consist of those programs that define new databases and modify the existing ones, perform utility functions, interchange data with other systems, and provide programming facilities.

3.5.4 IR FEATURES OF CDS/ISIS SOFTWARE

CDS/ISIS is a menu driven generalized Information Storage and Retrieval system designed specifically for the computerized management of structured non-numerical databases (Unesco 1989). It has been claimed in the manual that it does more than just text processing, which are normally found in word-processing packages, because the text that CDS/ISIS processes is structured into data elements that a user defines. Data elements in CDS/ISIS are stored in fields the collection of which make up a record. The fields of CDS/ISIS are variable length so also are, the records. This feature allows high rate of disk space utilization and gives freedom in defining the maximum size of fields and also makes a field or its part optional. Allowing fields or their parts to occur more than once, to be repeatable, is also the other good feature of CDS/ISIS. The major IR functions of CDS/ISIS according to Unesco (1989:4) facilitate:

- Define databases containing the required data elements;
- Enter new records into a given database;
- Modify, correct or delete existing records;
- Automatically build and maintain fast access files for each database in order to maximize retrieval speed;
- Retrieve records by their contents, through a sophisticated search language;
- Display the records or portions thereof according to user's requirements;
- Sort the records in any sequence desired;
- Print partial or full catalogues and/or indexes;
- Develop specialized applications using the CDS/ISIS integrated programming facilities.

These facilities are classified into: user services, operating on existing databases, and system services, designed for the database administrator to create new databases and perform various system tasks. Some of the tools and techniques provided by CDS/ISIS to make these tasks easy include: Menus, Multilingual dialogue and worksheets.

The software accomplishes these tasks using a database that consists of a number of logically related but physically distinct computer files. Some of the major ones are: database

definition files that consist of field definition table (FDT), data entry worksheet(s), display format(s) and field select table(s); Master file; Inverted file; and ANY file. However, it should be noted that the software has got different system restrictions. Some of these are indicated as appendix B along with the upgrades made.

CDS/ISIS also maintains a file that provides opportunities by which users can change system setup parameters. The file is called SYSPAR.PAR. The parameters that can be specified include about 13 system parameters and 11 database parameters.

As it is a textual IR system, CDS/ISIS does not use relational model to structure databases. The data is, therefore, defined in a single file, the master file. It uses, the REF function and L formatting function to establish some simple relationships. The search mode of CDS/ISIS is inverted file based and the search language is based on boolean logic. Its IR service provides facilities to: Browse master file, display search results, change display formats, execute previous search, change dialogue language, save search results, search formulations and recall the previous ones, if required, and display terms dictionary to select terms from.

The output component of the system facilitates users to change the default line width, number of columns, column width, lines per page, and indention using system print worksheet. One can also request sorting of search results, of course not in the order of relevance. With version 3.0 and above, new print options are introduced to allow the production of ASCII files with no carriage returns other than the ones inserted through the new line commands of the user's format. It has also become possible to produce print files in Microsoft Rich Text Format (RTF), which may later be printed using a word processor accepting RTF (such as Microsoft word or Word for Windows).

3.5.5 PERFORMANCE OF CDS/ISIS IR SYSTEM

Members of the standing committee on Harmonization and standardization of Documentation and Information Systems in Africa, in their sixth meeting (PADIS 1994e) recognized that all existing database management systems had shortcomings but resisted the creation of a whole new software to bridge such shortcomings as too costly and unnecessary. They also noted that a windows version of CDS/ISIS that was being developed would open new possibilities of software integration. They moreover, realised that it was then possible to run parallel systems such as xBase and CDS/ISIS in DOS and exploit their respective strengths.

As a strength of CDS/ISIS, Rashid(1994: 62-63) attributed the retrieval speed of CDS/ISIS to its development using ASSEMBLER (80%) and PASCAL (20%). The software has also an excellent internal editor together with a mechanism of checking data validity and online help (Simui 1995: 85-86). Besides it has the capability to create integrated as well as separate databases (Rashid 1994: 83). In addition to its ability to handle bibliographic, referral type and factual database, Neelameghan, cited by Rashid (1994: 24), claimed the capability of CDS/ISIS to handle object-oriented databases.

Some of the search facilities available on CDS/ISIS include: Right truncated search; Range search; field and proximity search operators ((H), (F), (.), and ($)); relational search; ANY term search; and Free text search. In its multilingual nature Amharic, Arabic, Chinese and some other languages scripts of East Europe and of India are available, in addition to Latin/Roman scripts (Simui 1995: 27).

Its network support introduced in version 3.0 is the feature that facilitated for fast interactive IR, retrospective searching, current awareness services and selective dissemination of

information (SDI) service (Rashid 1994:93).

The software also provides output facility of both soft copy and hard copy. The user can make modifications to the search results in soft copy. Data exchange through the ISO 2709 format is also possible. To download structured texts in ASCII format, and dBASE, FOXBASE and INMAGIC databases there are appropriate programs (Rashid 1994:27).

As a limitation, Perera's 1992 evaluation of CDS/ISIS version 2.3 user interface, cited by Simui (1995:91), concluded that it is not user friendly for novice users. This is evidenced by the several attempts made to come up with friendly search interfaces of CDS/ISIS two of which are Heurisko of Elvio of UNESCO and System Interface Search Assistance (SISA) of Molla Hunegnaw. The other limitations that may be considered as low performance of the software include: the system restrictions (appendix B), its inability to handle very large databases as that of the smaller ones, the space utilization by facilitating files, and all what has been discussed in CHAPTER TWO of this thesis as compared to xbases (relational DBMSs).

Comparatively, CDS/ISIS lacks some of the above discussed capabilities of recently developed text retrieval systems. Moreover it lacks some of the good features of relational and object-oriented database management systems. PADdev IR system is based on CDS/ISIS and is therefore limited to the performance of CDS/ISIS.

CHAPTER FOUR

EVALUATION OF IR SYSTEMS

Theoretical basis for the evaluation of IR systems is discussed in this chapter in detail. The need for an evaluation had arisen, as Cooper (1973) was trying to imagine, because some clever design modification had been proposed which might substantially improve the system's retrieval performance. Martin and Parker (1978) also noted that system evaluation is necessary if the system is to stay viable and if lessons learned from the development process are to be useful for later versions of this or other systems. That is why evaluation appear to be one of the three areas of IR research that make up considerable portion of the subject (Van Rijsbergen 1975:6). As a theoretical basis of this thesis, the following parts discuss research development and other aspects of evaluation research.

4.1. RESEARCH DEVELOPMENT

Gilchrist (1996) points out that interest in evaluation has continued at high level since 1940s or earlier, when comparative examination of the merits of a classified catalogue and an alphabetical subject catalogue was a favoured papers of librarians. He also recognised the fact that with the subjects to be evaluated, styles of evaluation have changed to suit new technical developments, operations needs and theoretical attitudes.

According to a survey by Robertson, cited by Cooper (1973) and Salton (1992), the first system analytic retrieval tests took place in the mid 1950's. This can be affirmed by Lancaster's guess that the first evaluation study in IR was conducted in 1953, as cited by Chowdhury

(1994:52). Since then, an astonishing amount of attention had been devoted to the problem of retrieval system evaluation. However, scholars including Lancaster and Ellis (1994) have given the significance, as a starting point, to the Cranfield project, as these were the tests that provided the ontological and methodological framework within which evaluation of IR first developed.

Spark Jones, cited by Chowdhury (1994:52-53), summarized the trend of evaluation research of the first two decades as follows:

first decade	- evaluation of indexing system
second decade	- comparison between different indexing languages
	- evaluation of indexing exhaustivity and specificity
	- evaluation of search techniques
	- evaluation of output ranking term weighting, relevance feedback etc.
	- evaluation of efficiency (cost-effectiveness).

Gilchrist (1996) also tries to summarize four decades of information system (IR emphasized) evaluation ascertaining two main trends, and a constantly changing balance:

- The first trend is that away from reliance on a single 'scientific ' model of evaluation towards a more complex mix of quantitative and qualitative methods, based more on social science models.
- The second is the requirement to find new means of evaluation to cope with problems presented by technological advance like initially mechanized documentation systems, then computers, then online systems , and now expert systems, optical media and hypertext.

He further pointed out that the great majority of information system evaluation conducted to date have been largely quantitative in nature which in part resulted from the fact that the methodologies used were taken either from the physical sciences and technologies, or from the type of social sciences research which involves extensive quantitative surveys, with subsequent

statistical analysis.

Although it did consider the management rather than the requirement of individual users, Lancaster's MEDLARS evaluation is one early, and particularly interesting example of the evaluation of an operational retrieval system (Gilchrist 1996). The Cirt evaluation project is the other good example of an operational evaluation (Robertson and Hancock-Beaulieu 1992).

In this regard, nowadays researchers, including Salton himself (1992), started to doubt the applicability of laboratory-based evaluation of IR systems that relies on fictitious information needs and artificial relevance judgments which are more likely to be quite different from real situations, and they are advocating the user-oriented approach (eg. Smithson 1994).

4.2. CRITERIA OF EVALUATION

An evaluation is basically a judgment of worth or an appraisal of value. The process of evaluating is highly complex and subjective (Swanson 1978). Participants, if the evaluation is from their point of view, may create difficulty in that it is very difficult to assess how people will react to a new technology before they have experienced it, on the one hand they may be unaware of the problems it can solve. On the other hand, they may be carried away without appreciating the costs and limitations involved (Martin 1978).

Other than their difficulty and complexity, what distinguishes evaluation studies from other types of inquiry, scientific or otherwise, is their purposive orientation. They basically investigate the achievement of goals or expectations (Swanson 1978). Some of the specific purposes listed by Swanson (1978) are:

1. To explore techniques for increasing program effectiveness;
2. To establish a foundation for further research on the reasons for the relative

success of alternative techniques;

3. To improve the means employed for attaining objectives or to redefine sub-goals or goals in view of research findings.,

Of all, Swanson (1978) prefers the inclusion of the planned exploration of likely alternatives suggested by study results. In this line, King and Bryant, cited by Gilchrist (1996) had given four possible decisions to be supported by evaluation:

- decisions to design or implement a new system
- decisions to modify an existing system
- decisions to discontinue a system
- decisions (presumably minor modification) arising from continuous monitoring of system effectiveness.

Noting the problems of evaluating an IR system, specifically, Van Rijsbergen (1975) posed three questions, the first of which is what the above cited scholars tried to answer: 1) why evaluation?

2) what to evaluate? and 3) how to evaluate?. His answer to the first question was 'To put a measure on the benefits (or disadvantages to be got from IR system)'. Cleverdon answered the second question as early as 1966 listing the criteria of evaluation given in the first part of this thesis.

Van Rijesbergen attributed the answer to the third question to the particular retrieval strategy adopted and the form of its outputs. Categorizing evaluation works into two Gilchrist (1996) suggested that in evaluation distinctions should be made between the following counter methodologies or objectives: qualitative, laboratory, experiment, poor-oriented, macro evaluation(indicate improvement), objective, improving understanding(academic research) on the one hand; and quantitative, operational, investigation, insight-oriented, micro evaluation(produce improvement), objective and subjective and producing improvement(action research) on the other hand. He equated the former group to research while the latter to evaluation.

The first counter aspects of Gilchrist has been already considered by Schwachaw (1990)

and are given corresponding indicators. According to him indicators of the quantity are, for instance:

- number of users
- volume of uses
- revenues
- connecting hours
- number of documents delivered
- number of references

While examples of quality indicators are:

- reliability in the provision of the services
- up-to-datedness or timeliness of the delivered information
- novelty of the delivered information
- speed or frequency of the information provision
- completeness of the information
- selectivity of the information
- relevance or specificity of the information
- integrity of the information
- security of the information provision
- user friendliness, user effort
- flexibility of the information services
- accessibility of the information services.

Swanson (1978) saw indicators like the above ones as criteria that must be transformed to criterion variables to produce measurement of the degree to which a value situation meets stated goals or objectives. Gilchrist (1996) on his own compared the performance of an information system with some theoretical maximum, which, in an operational situation, means the satisfaction of the real needs of its users.

Losee (1994) states that users of document retrieval systems often measure system performance by examining one or more of a number of different measures, including precision and recall and their variants, as well as single-number measures of performance such as Van Rijsbergen's E, and Shaw's D, also system's utility to its users which was claimed by cooper (1973). Van Rijsbergen (1975) in addition, advocated the ability of the system to satisfy the user in

terms of the relevance of documents retrieved to be measured in evaluating effectiveness. Yao (1995), on the other hand, states that representation of user judgments on the usefulness of documents is fundamental to the development of reliable methods and techniques for measuring the effectiveness of IR systems. Users are basically assumed to judge a retrieved document in view of their information needs put into a query.

According to Losee (1994) the relationship between a query and a document is usually described in terms of the concept relevance. In this regard, Ellis (1990:21) restated Cleverdon's outlook that if relevance was to be employed as a performance criterion, the form of relevance judgement employed had to be objective enough to serve as the basis of the measure of effectiveness. However, again citing a series of studies on the consistency of relevance judgements undertaken by Cuadra and associates, Ellis (1990:18) shows that a number of factors affected the consistency of relevance judgements assumed above. Ellis(1990) also discussed Maron and Kuhns' probabilistic outlook towards relevance and that of Robertson and Spark Jones's model together with an attempt made by Robertson, Maron and Cooper to unify these two competing theoretical models.

Not being satisfied with evaluating only the effectiveness of IR system, Rowley (1992:387) and Cleverdon (1978) provided a Checklist and points, respectively against which a database and text retrieval software can be evaluated (both are given in the first part of this thesis). In the same sense, Chowdhury (1994:37-38) also listed a number of system parameters for different criteria. The evaluation criteria for which he associated different system parameters include: recall and precession, response-time, user effort, form of presentation, collection coverage.

In view of the last three scholars' outlook, cited above, to the evaluation of IR system, Townley's (1978) statements are worth restated here as follows: the search is only a part of the IR

system, and it is the effectiveness of the whole system not merely search effectiveness, which is sought.

Most of the above criteria are used to measure or judge the performance of an IR system relatively with their optimal and/or theoretical value(qualitative or quantitative) that an ideal IR system is expected to achieve. Some of these optimal performance expected of an IR system are discussed bellow.

4.3 OPTIMAL PERFORMANCE OF IR SYSTEMS

Soergel (1994) pointed out that retrieval performance is a function of the agreement of the judgements of two parties, the indexer and the user (searcher). He also states that the foundation of all performance evaluation is a simple criterion: can one ask the questions one needs or wants to ask and get an answer with acceptable levels of successful retrieval and of distracting noise?

Gilchrist (1996) considers an IR system perfectly effective if it provides for the nearest possible coincidence between the description of a subject by a searcher and the description used to enter documents on that subject in the system. Because in such a system we could have the position in which every relevant document would be retrieved when a search was made, without the production of a single irrelevant document. This was, of course the idea of Van Rijsbergen (1975:4), Rowley (1993:129), Soergel (1994) and Meadow, cited by Chowdhury (1994). The optimum of relevance seem to be represented by Soergel's (1994) concepts of Topic relevance, Pertinence, and Utility which are central for good indexing and for assessing retrieval performance. Because they have been defined considering each and every aspects of the user's information needs.

Chowdhury (1996) adds to what have been discussed above the capability to create and maintain one or more databases containing records pertaining to the requirements of the user community as another facility of an IR system is expected to provide.

An Information Retrieval system should also provide all of the following Basic retrieval facilities listed by Rowley (1993:124)

1. Basic facilities - help, news, log-off (set up the environment)
2. Selecting search terms - by viewing index terms or thesauri
3. Entering search terms
4. combining search terms
5. specifying fields to be searched
6. Truncation
7. phrases, adjacency and proximity searches
8. Range searching and limiting
9. Displaying records
10. Search management (Reviewing the search)
11. Advanced display options (to accommodate records in full text databases)
12. Multi-file searching (including repeat searching in other databases)
13. Displaying the thesaurus
14. Hypertext (hypermedia or databases integration)

The functional requirements listed by Chowdhury (1994) , most of which are mentioned in the first part of this thesis are also the ones that can be fulfilled by an optimally acceptable user-oriented IR system.

Moreover, Listone and Schoene, cited by Chowdhury (1994) suggested that an optimally effective IR system must have:

1. Provision for prompt dissemination of information
2. Provision for filtering of information
3. Provision for right amount of information at the right time
4. Provision of active switching of information between databases.
5. Provision for receiving information in the desired form.
6. Provision for browsing
7. Provision for getting information in an economic way
8. Provision for current literature
9. Provision for access to other information systems
10. Provision for interpersonal communication
11. Provision for personalised help.

Optimally also an IR system should allow to conduct a search of any type, as discussed below.

Generally, Evaluation research is developing from being system-oriented to be user-oriented and has come out of the laboratory to the operational environment. The concentration on the recall and precision of the system is shifting to consider all the other features of the system under investigation. That is why this study is planned to adopt the user-oriented paradigm of evaluation and considers different IR features related to inputting capability; searching and retrieving; and output presentation performance of PADdev IR system.

CHAPTER FIVE

PERIPHERAL FACTORS INFLUENCING IR PERFORMANCE

Different types of searches and the data entry formats reviewed in this chapter as the major influential factors but peripheral to the algorithmic features of IR systems. These are the features where individual Information Systems get opportunities to improve the retrieval capability of the IR software developed on the basis of an algorithm. Therefore the effort of a system is reflected in these features of an IR system.

5.1. SEARCH TYPES

Townley (1978) pointed out that as far as computer systems are concerned, there are only two possible classes under the heading of response times: 'NOW' and 'NOT NOW', The online and the batch system respectively. Although almost all computer systems now-a-days were developed in the former mode and enforced the latter to disappear making the concept online synonymous with network, this thesis discusses them in the sense of their original meanings.

5.1.1 BATCH MODE SEARCH

At the early stage of computer-based information processing systems before 1960s, the searches were conducted at an information centre which had the appropriate tapes available on its own computer. Users would send their requests for information to the centre, where the searches were conducted in a batch mode. In such a search, a number of searches or user-profiles were run at one time on the computer, which produced listings of records tailor made to a client's subject

interest (Convey 1989:5). The computers, mainly the mainframes, were expensive to be used interactively, being available for searchers for a long period of time. Moreover, the storage devices used then were not direct access, but they were sequential access, and therefore could not allow a fast search that can be completed with the human's tolerance. Instead the search strategies would be entered the computers that would be left alone for a long period of time to accomplish the search, mostly overnight. We can understand that such an IR system is, obviously, poor in its response time and user-system interaction. It is also easy to see how these and the other deficiencies can negatively affect the retrieval effectiveness of the system. But the development of the hardware and software technologies in 1970 allowed conversational searches to be conducted in an online search mode. By these online IR systems, which was widespread in US and later in Europe, the batch processing of searches became virtually obsolete.

5.1.2 ONLINE SEARCH

When one is online, one is connected directly to the central processing unit (CPU) of a computer using a terminal (Convey 1989:4). In such a search, searchers interrogate the computer and receive an immediate response and allowed to have a conversation with the computer. Due to this it has been termed as conversational mode. The lack and high price of high capacity storage devices, enforced the share of large databases stored centrally and accessed through remotely located terminals. It may be for this reason that the concept online search had been used for such remote conversational search modes.

With the development of different storage devices like CD-ROM, to store a significantly large size databases, online searchings started to cease to be remote. Because the databases could

be connected directly to the system . And therefore, the conversation remained between the peripherals and the CPU of a computer, instead of terminals and central computer. As opposed to the batch mode, the online-information requestor search modes allow the information requestor to converse with a computer on his information need, even with a blurred idea of his information needs. The trend towards this is being facilitated by the advancement and simplicity of online search interfaces. It is with such development that all the advantages of user-intermediary-system or user-system direct interaction can be gained.

5.1.3 END-USER SEARCH

From their experience, Neelameghan and Rahel (1994: 38) stated that user-information system direct interaction helps in the more precise formulation of user's information need and in the searching process so as to improve the possibility of retrieving information (record) pertinent to user's needs. Saracevic et. al. (1988) also emphasized the importance of human decisions and human-system interactions as variables in processes dealing with searching for and retrieval of information. However at a given time an information requestor is usually interested in selected attributes of the object, or using the object for a particular purpose (Neelameghan and Rahel 1994:38). Similarly, Belkin et.al.(1993) noted that people exhibit a variety of behaviours when they engage in information seeking, and that effective information systems should support both these behaviours, and graceful movement among them. They have also presented the different possibilities of information seeking behaviour that need to be facilitated by an IR system as a table.

Martine and Parker (1978), moreover, argued that the non-captive searcher is not likely to become a user unless the system responds to his needs at a cost in time, energy, and money that

he can afford. They have also noted that it is not possible for the designers to negotiate with the users in order to develop detailed training and systems specifications. Instead, it is necessary to design a flexible and changeable interface that can be adapted to a wide variation in user demand.

From the users side, Rowley (1993:138) noted that the end user as a searcher should have a good perception of the information that is required, and also understand the terminology of the topic which is being sought. So that he/she can communicate with the system effectively. In general, Salton (1978) attributes the degree of system's profitable use to the interaction it can do with its users.

5.1.4 INTERMEDIARY SEARCH

An intermediary is a person conducting the online search on behalf of the end user. An intermediary search is, therefore, a search conducted by a person, in most cases librarians, information brokers etc. for requests submitted by end users.

An intermediary, to satisfy the information needs of information requestors using one or more IR systems, should be able to communicate both with the system(s) to be searched and the information requestors who have expressed or latent information needs.

Although it is in the sense of an online search, Convey (1989:79) discussed subject expertise, knowledge of databases, detailed knowledge of command languages and typing skills, logical thinking and analytical skills, curiosity and persistence, self-confidence and communication skills as requirements of a search intermediary. He also argues that a combination of the precise subject knowledge of the end user and the system knowledge of the intermediary will normally produce the best results of a search. In order to define the exact scope of a search

precisely, an intermediary should obtain the users' subject knowledge and understand their information needs. This is made difficult by the fact that those who seek information do not always specify their information needs. An intermediary undertaking the search should therefore, liaison in some depth with the information requestor , often in a face-to-face reference interview.

In this line, Lancaster, cited by Convey (1989:53), stated that while the requester himself might successfully browse through the literature on the basis of an ill-defined need, it is impossible to prepare a successful search strategy for machine search on the same vague basis. This is still true for IR systems functioning with out the advanced user interfaces. If a machine search is to yield useful results, we must do as much as possible to obtain request statements that explicitly delineate the actual information need, the wider the gap between stated request and information need, the less successful the search is likely to be.

Convey (1989:53) pointed out some of the ways through which search requests may come to the intermediaries. For example personal call, telephone, telex, letter, special online request forms, e-mail. Irrespective of the method of receiving the request an intermediary need to have as complete information on the information need of the requester as possible. He further suggested that ideally a search in specific subject be conducted by subject specialist intermediary. Actually, it would be impossible for the intermediary to be a subject specialist in every subject. But some background reading, coupled with the help from the information requestor, will avoid the situation of typing in a number of unintelliglbe terms and hoping the unintelligible records coming out of the computer to be the correct ones.

Equal to his/her effort to obtain the required information about the user's information need and understand it, an intermediary should have all the required system's knowledge so that he/she can translate user's language to that of the system's, i.e, search strategy. As users use different concepts/words to express their information needs, IR systems may also accept search strategies

in different ways like Boolean, truncation, proximity, adjacency, etc. In this line, Rowley (1993:138) emphasized that ideally the intermediary is adequately trained in the search facilities of the system.

An intermediary is also expected to modify a search strategy/statement when it is resulted in unsatisfactory search results. Which in effect means he/she should judge the search result on behalf of the information requestor, on the basis of search requests and his understanding of the information needed.

An intermediary should also reproduce the search result, which contain set or sets of relevant records, in a format suitable for the information requestor to use. Some of the output options that may be used to present search results are screen (VDU mostly) display, printing, downloading.

Here also the information requestors' information use situation, habit, or otherwise need to be known by the intermediary. That may be what made Gilchrist (1996), Citing David Bawden, consider primarily the information professionals as searchers, intermediaries, operators, providers, maintainers etc as users of IR systems.

Nowadays hardware and software technological development is coming up with very interactive user interfaces that seem to perform as an experienced intermediary through the use of expert intermediary systems like what have been discussed in the previous parts of this chapter, for example Intelligent Interface for IR (I^3R). However, professional intermediaries continue to exist as far as online IR systems exist. Of course they may shift their role to advisory, and some other similar services.

5.2. BIBLIOGRAPHIC FORMATS

A record in a database of any type has both the logical and physical format/structure. The former refers to the structure of a record by which it appears to the user either when he inputs it to the computer or when the computer displays or presents to the user, while the latter refers to the structure of a record by which it is recorded on to the physical storage media. A logical record may be either stored as it is, split into smaller parts to be stored separately or grouped with other records to be stored in the same physical location. The conversion of record structure between these formats is the task of computer programs. By record format, which implies the idea of a set of rules or conventions controlling record representation, Gredley and Hopkinson (1990: 44) mean the arrangement or structure of computer readable records of information. The purpose of most, if not all, of the formats set so far is to establish a standard of handling record structure in a database. This might be because most of the databases are designed for different purposes and to be exchanged among different information systems. Pointing out the fact that record formats may be either exchange or internal ones, Gredley and Hopkinson (1990: 46) noted that exchange formats have had great influence on internal ones. The reason for this is the fact that sharing of data is so important in the library and information field. They have also claimed that ideally, exchange formats should facilitate the exchange of data which are to be used in a wide range of different bibliographic applications, from the production of the traditional catalogue cards to records in databases which are used for online access. On top of their ideal format, we need to add the capability to handle database(s) of different types of information items as a requirement.

Most of the adopted exchange formats consist of three basic components (Chowdhury 1996):

1. A defined physical structure

2. Content designators

3. Content of the record

The most remarkably agreed upon exchange format is ISO 2709. It is the international standard format for bibliographic information interchange on magnetic tape. It has its origins in the library of congress MARC format. The ISO 2709 structure consists of the following elements: record label, directory, bibliographic data fields and record separator. It follows the structure of variable length format with content designators. Despite its wide acceptance, there are different record features that are beyond the control of ISO 2709. Some of the record structure features that are not accommodated by ISO 2709 include (Gredley and Hopkinson 1990 : 39-66).

- Linking through the embedded fields that can hold actual titles of related works or identifiers like database numbers or ISBN.
- The content of each of the data elements in a record
- ISO 2709 does not officially support the use of the today's $5^1/_4$", $3\ ^1/_2$", 8" or 3" diskettes but on agreement bases. Because it has been developed for tape storage media.
- Standardization of character sets is the practice of other standards approved by standards organizations such as ISO, ANSI, BSI, ASCII. From ASCII the international standard ISO 646: 7-bit coded character set for information processing interchange was developed.

Yet, ISO 2709 has influenced a number of computer systems in their internal structure, such as MINISIS and CDS/ISIS for microcomputers. Moreover, most of the record formats are developed based on ISO 2709. Some of these record formats are discussed below.

5.2.1 RECORD FORMATS

Probably the most used and best known exchange formats are national MARC formats like USMARC, UKMARC, AUSMARC, MALMARC, etc. UNIMARC has been developed in 1977 by IFLA to facilitate data exchange between these national MARC formats. (Hopkinson 1991: 169). The original MARC format from which all the other formats originated was developed at the Library of Congress in 1965-6. Since then more than 20 formats of common characteristics, like adherence to the ISO 2709, having official status, have appeared (Gredley and Hopkinson 1990: 70).

Hopkinson (1991: 169-170) briefly described the historical development of UNIMARC and pointed out some critique on the format. To mention some:

- It contains some redundancy
- Records using main entry like those created according to UNIMARC format, and those that do not, were to be completely incompatible.
- It is not also well promoted. Moreover, experts from developing countries could not participate in its development and therefore it is not worth being used by them.

Most MARC formats depend for the definitions of their data elements on the national cataloguing rules. UNIMARC assumes the use of a cataloguing code based on ISBD: It also supports four categories of bibliographic levels: monographic, serial, analytic and collection along with the linking mechanism between them.

Hopkinson (1991 : 171-176) briefly discussed some features of UNISIST Reference Manual and Common Communication Format (CCF) and forwarded some critique on the latter. According to him, the UNISIST Reference manual was developed to resolve the problems of secondary services and was widely circulated by Unesco. The manual was used to model the

format for DEVSIS and was adopted by the MINISIS software system. The manual contains its own cataloguing rules to define its data elements.

As it has been noted in the format (Simmons and Hopkinson 1992 : 11), the CCF was designed to provide a standard format for three major purposes:

- To permit the exchange of records between groups of information agencies
- To permit the use of a single set of computer programs to manipulate records received from various information agencies regardless of their internal record creation practices.
- To serve as the basis of a format for an agency's own bibliographic or factual database.

For the achievement of these purposes the format:

- Specifies the minimum set of mandatory data elements which are required to provide a useful and unambiguous record.
- Provides mandatory and optional data elements that are sufficiently flexible to accommodate varying descriptive practices.
- Permits the originating agency to create and include non- standard elements which are considered useful within its system even though they are not used by other agencies.
- Provides mechanisms for linking records and segments of records without imposing on the originating agency any uniform practice regarding the treatment of related groups of records or data elements.

As a result of interests from organizations wishing to process and exchange factual information, the CCF has been extended on take into account the different factual data elements needed (Hopkinson and Simmons 1992 :5). However, it has been indicated that the format has not been made to include data elements for recording dynamic data or statistical tables. It rather caters for the types of data that are included in directories, providing data elements for the

recording of projects, persons and institutions. Of course the other categories of factual information were promised to be included. Opposing the critique on the over-complexity of the format, Hopkinson (1991: 174) noted that its complexity is as a data entry format, which it was not intended to be. However, he pointed out three points that one should remember in evaluating the CCF. These are:

1. The CCF was not designed from first principles but was based on major existing international exchange formats and was intended to be used for the transfer of records between systems which were already capable of providing output into these major exchange formats.

2. The CCF is intended as an exchange format and as such has to contain bibliographic data for exchanging between systems. Due to this, it does not assume any particular cataloguing code.

3. The CCF is intended for exchange of bibliographic data.

The other well known record format is the MIBIS (Manual for preparing records in Microcomputer based Bibliographic Information Systems). It includes (Di Lauro and Brandon 1990):

- a database design, in the form of a list of fields for the description and management of the items in the database. The database design that aim to be as complete as possible, presenting all of the data elements that any information system might need. It is of course, possible to use all or some of the elements.
- a set of rules for the content of the fields
- guidelines on how the database may be used for information management, retrieval and dissemination
- guidelines on the operation of the system within the context of an information network.

The rules for bibliographic description in the manual adhere as closely as possible to the

Anglo-American Cataloguing Rules (AACR2). More than any other formats MIBIS gives emphasis on the provision of network capability to the database. MIBIS was designed to be compatible with the CCF, but has differences. The format has fields to contain data related to:

- networking
- core bibliographic data
- describing parent items
- subject analysis
- local processing
- acquisition

The other upcoming record format that has been designed to resolve the long living problem of database integration is the ABNCD. The fields in the ABNCD hold data related to networking and local information. Some of them are related to (Abebe et.al. 1992):

- participating centre acronym
- participating centre record number
- record status
- date record entered
- documentalist

The structure also makes provision for acquisitions information. It allows handling of records related to bibliographic data, profiles of institutions, information systems, projects and experts. Chowdhury (1996) mentioned a number of products and services that can be facilitated by the ABNCD format. Some of these are:

- Concurrent search and retrieval of different types of records.
- purchase orders and claim notices

From his comparative analysis of the major record formats, MARCs, MIBIS, CCF and ABNCD, Chowdhury (1996) concluded that only CCF and ABNCD can provide for both bibliographic and factual records, i.e the others handle only bibliographic records. Moreover, he preferred ABNCD for factual databases, because it is more detailed than the CCF.

PADIS developed its own format known as PADIS Manual for document analysis. The purpose of the manual is to give instructions on how to prepare input for the PADdev database. It is meant for documentalists who are using CDS/ISIS. If a different application is used for data entry, the manual can only be a source of inspiration for the person who compiles a manual. The PADIS manual has got four data entry worksheets (on-screen forms used to enter and/or modify data in the database or to collect the parameters required to perform some functions). Each worksheet is used to enter or edit records for different types of documents.

The PADIS manual accommodates about 51 fields that contain information of four different groups. These are:

1. Identification of documents: Field 110-170, i.e identification of the input centre, status, relationship with other records.
2. Bibliographic description: Field 210-820 eg. authors, titles, publisher.
3. Content analysis: Field 830-940 i.e the subject contents of the document
4. Identification of indexers and data input personnel: Field 950-970.

It provides only seven optional fields: form of text (510), availability (310), ancillary notes (820), Participation in other Information systems (870), proposed descriptors (920), Abstracts/French (930), and abstracts/English (940).

The manual creates relationships through a field (140) that contains the identifier of the related record (from field 110). It also provides predefined display formats in which records will be displayed or printed or saved to a disk file. At the end of the manual, alphabetical list of codes

for names of countries, language codes, annotated PADIS purpose codes are given.

5.2.2 EFFECTS OF RECORD FORMATS ON IR PERFORMANCE

From our discussion on the different record formats, it is possible to analyze that there are several IR features that are under the influence of, if not determined by, the record formats either adopted by or developed for an IR system. Some of these features include: The type of information elements to be input to the system, the size of each information elements, generally the content of a record; the relationship that a record can have with the others; the capability of the system to handle information elements of different nature, i.e database integration; number of searchable information elements; the system's responsiveness to change, i.e flexibility; the influence on most of the above record features can be reflected on the system's search effectiveness, in terms of retrieving only the relevant documents; the system's output form; output exhaustivity; and output form flexibility.

Some record formats allow a database that adopts them to contain only the bibliographic information of the entities of the database while others enable it to include the information content of the entity. Even from the former group, some are very restrictive in the number of information elements that should be recorded about the entity while others accommodate a very large number of information elements to be recorded about an entity of a database. This difference, inevitably results in a difference in the IR system's search capability in that the Recall of search made on only the title is less than that made on the database that includes the abstract which itself comes up with less recall than a search made on a full text database. Moreover, an IR system is a system that handles what is fed into it and presents only, in most cases, what it has accepted. This makes

it dependent on the record format that it adopts because it searches only what is contained in the records of the database and outputs the content of the retrieved records according to the output rules developed by the adopted format.

One of the search type that is getting greater acceptance is the hypertext/associative search. But it is clear that such a search can be implemented on a database that has the appropriate association/relationship created between different records. On the other hand the type of relationship that can exist between records or information elements (fields) is determined by the adopted record format. Most of the formats create only static relationships. Even this relationship may not be created by some record formats which limits the system to retrieve only one record at a time.

The number of information elements (fields) on the bases of which the system can perform the search, of course is related to the content of a record, also determine the search alternatives of the IR system. That is, the system's capability to specify a search to one field, two fields etc. but it intern, is limited by the number of searchable fields available in the record format. The problem of system's compatibility created by the difference in the record formats adopted by different IR systems is evidenced by the effort exerted to develop intermediary exchange formats like UNIMARC and CCF. Moreover, network ability of a database is also determined by the record format it is developed with For example, networking is facilitated by the MIBIS record format much more than the PADIS record format in that it provides a set of information elements (001-003) which are related to networking.

As it is the record format that limits what can be fed into an IR system, it is the format that limits what and how to present the retrieved record to the user like what the display formats provided by the PADIS record format are serving. The record format adopted by an IR system also limit the range of the system's output. For example ABNCD facilitates the provision of

different information products and services that are not facilitated by the PADIS format. These include purchase orders and claim notices, interlibrary loan requests.

PADdev is handled by a centralized online IR system and mostly to provide an intermediary search service. It is also developed on the basis of PADIS database format which lacks some of the good features that we can notice from the above described formats like integration, popularity and application independence.

CHAPTER SIX

THE PAN-AFRICAN DEVELOPMENT INFORMATION SYSTEM (PADIS)

The IR system that this thesis work is set to evaluate is a subsystem of PADIS. It was planned to store, maintain and retrieve bibliographic information related to African development. PADIS, as a super system, is the environment of the IR system under evaluation. With this understanding this chapter of the thesis describes the establishment, structure, objectives and databases of PADIS; the services that PADIS is offering on the basis of its databases and the activities and any other changes it is planning to make in the future. The potential sources of the data to the database and the users of the services are also briefly described in the chapter together with reviews of some/selected evaluation studies that looked into the achievement of PADIS' objectives.

6.1. ESTABLISHMENT

PADIS was established to be the Pan-African Documentation and Information System (Kangulu 1994) in 1979. Kangulu (1994) also points out that its establishment was accelerated by the creation of the DEVSIS-type systems and by the lack of an information system to deal with African development information at the continental level.

PADIS, as a Pan-African Development Information System, became operational in 1980 with assistance from the UN Development Programme (UNDP), The International Development Research Centre (IDRC), the Unesco, the African Development Bank (ADB) and ECA African member States (Kangulu 1994). It was envisaged to be an international cooperative information system (Kangulu 1994), but is functioning as ECA's programme on harnessing information for

development. Recognizing the different factors, the 14th ECA Conference of Ministers in 1979 adopted Resolution 360 (XIV) on the establishment of a Pan-African Information System.

6.2. FUNCTIONAL STRUCTURE

The existing structure of PADIS, in relation to ECA was reviewed in the first part of this thesis. To look into it a bit closely, on the basis of the 1979 ECA/OAU/IDRC Advisory Mission, PADIS was designed to operate on three levels and establishment of its Central Coordinating Office (PADIS-CCO), Sub-Regional Centres (PADIS-SRCs) and National Coordinating Centres (PADIS-NCCs) (Kangulu 1994). It also constitutes Institutional Participating Centres (PADIS-IPCs). The highest forum to which PADIS reports was the Conference of Ministers of Planning and Economic Development that meets every year, reviews ECA activities and approves ECA programme budgets (PADIS 1994d). The Regional Technical Committee for PADIS, which is an organ of the ECA, is responsible for establishing overall policy for PADIS and overseeing all aspects of its implementation (PADIS 1993b). The Sub Regional Technical Committees oversee PADIS activities in their respective subregions. These were (PADIS 1994d):

- West African Development Information System (WADIS) located in Niamey, Niger;
- Central African Development Information System (CADIS), located in Kinshasa, Zaire;
- East and Southern African Development Information System (ESADIS), located in Lusaka, Zambia;
- North African Development Information System (NADIS), located in Cairo, Egypt.

Each of these was designed to constitute the National Participating Centres (NPCs) of the respected region. NPCs are government designated institutions which perform the function of assisting national development planning by providing ready access to information required by

policy makers, planners and others engaged in national socio-economic development efforts. They are also responsible to coordinate information activities of sectoral documentation and information centres at national level (PADIS 1994d). Such responsibilities at international level are undertaken by Institutional Participating Centres (IPCs). The IPCs comprise regional and sub regional institutions responsible for coordination of information collection, classification and dissemination and development of tools and methodology in their special subjects of focus. They are African regional or sub regional institutions to which two or more states are parties, including ECA sponsored sub regional and regional institutions as well as other institutions in the region. Both governmental and non-governmental institutions which are active in development information are among the members. (PADIS 1994d, 1993C).

At the level of Central Coordinating Office (CCO) level PADIS set up a standing committee on the Harmonization and Standardization of Information Systems in Africa in 1989.

The committee deals with standardization matters on a continuous basis (PADIS 1996). The number of subcommittees that function on different aspects of the matter grew from one in 1989 (the subcommittee that existed since 1988) to seven in 1995. The following are the existing subcommittees (PADIS 1995).

Subcommittee 1:	PADIS Manual for document analysis
Subcommittee 2:	Selection and evaluation of criteria of textual database programs
Subcommittee 3:	Evaluation of textual database forms and structure
Subcommittee 4:	Selection and acquisition of microcomputer hardware configurations
Subcommittee 5:	Computer networking e-mail and online access
Subcommittee 6:	Authority files
Subcommittee 7:	Cd-Rom

The committee on the Harmonization of Documentation and Information Systems is used by the IPCs as a platform of communication with CCO (PADIS 1993b). The other important aspect of the standing committee's work has been in the area of development of conversion programmes to handle different data structures and facilitate transportability of data (PADIS

1996).

Although PADIS failed to implement most (3 out of 4) sub regional components its NPCs have covered a significant number of African states, while the number of its IPCs has exceeded 50. From Now onwards, PADIS moved to be structured under the Development Information Services Division (DISD) together with the other information related sub divisions of ECA (mentioned in the first part of this thesis). Team 1 of the division contains Databases/Archives development (including library databases) and data integration and is named as Development Information Management.

6.3. OBJECTIVES

The overall objectives and immediate objectives of PADIS have been stated in the first part of this thesis. But what has been stated by Kangulu (1994:1), as an objective of PADIS, is worth restating here which says "The main objective of PADIS was to provide ready access to information required by policy makers, technicians, planners and others engaged in the economic and social development of African countries", because it is a user oriented objective.

In the same understanding, the immediate objectives mentioned by Jusa-Sheriff (1991), but not in the first part of this thesis, are mentioned below:

- To establish a system which will improve access to both published and unpublished documents produced in Africa on questions relating to economic, social, scientific and technological aspects of development;
- To promote the improvement of the information infrastructure in African member states in order to strengthen bibliographic control of national development information output.

To achieve its objectives, PADIS was, therefore, to design an IR system based on materials produced by African countries, or in other countries about Africa and link

multidisciplinary information resources in the ECA member states (Kangulu 1994).

PADIS has, also, an objective related to standardization (third immediate objective mentioned in chapter one of this thesis). In this regard, PADIS is promoting the acceptance of its bibliographic format so that the different problems of information exchange that arise from the use of different methodologies can be minimized or avoided. This objective continued to be an agenda until now. For example, it was agenda item 24 on the Ninth conference of African planners, statisticians, and population and information specialists (ECA 1996).

6.4. DATABASES

PADIS comprises both an information system and a network. The information system is made up of its databases (PADIS 1996b). The databases that PADIS is using as information resources fall into four categories:

1) bibliographical,

2) referral,

3) numerical, and

4) full-text.

In the first category, PADIS listed about 21 databases. Out of these two are created and maintained by PADIS and ECA as in-house databases. The second category consists of about 11 databases, 7 of which are in house for PADIS or ECA. There are also one in-house and one external databases of the third type, while there is only one in-house full text database (PADIS N.D). The two bibliographic databases maintained by PADIS are PADIS development database (PADdev) and PADIS complementary databases (PADcom).

PADdev contains records from ECA documents and documents input from national participating

centres, on economic, technological and social aspects of development in Africa.

PADcom is a group of about 21 sectoral bibliographic reference databases which cover the fields of agriculture, industry, labour, social affairs and natural resources. They are databases on magnetic and optical media together with those accessed online. This group of databases is intended to complement the PADdev.

Databases of the referral type include the following:

PADexp : It is a database of personal data on African experts in all aspects of development.

PADinst : It is a database of research institutions in Africa and consultancy services on development that are potential partners in the promotion of technical cooperation between developing countries.

PADpro : This is a database of development research projects in process in African countries.

PADdab : is a database of databases at regional and sub regional institutions in Africa.

The numerical or statistical database of PADIS is **PADstat**. It is a socio-economic numerical database of time series, statistical data, on all aspects of African development focusing on trade, population and employment, agriculture, forestry, fishing, national account, finance, prices, transportation and communications, and social statistics.

The only database of the last category - full text - is referred to as **UNPRESS**. It contains the entire UN press release. The content of the database is supplied by the department of public information of the UN. It covers economic development, emergency relief, environment, international trade, law, natural resources, politics, population, science and technology, and social conditions.

The database of authority file is a special kind of database developed and maintained by

PADIS for a different purpose - facilitating retrieval performance. It is a database of controlled names of institutions included in the PADIS/ECA databases to be used by individual organization and network participants wishing to standardize the names of institutions that are involved in development activities. Only the parent body of institutions is enclosed into the database with, a purpose of avoiding maintenance anomalies (PADIS 1994g).

Regarding the relative utilization of PADIS database, Kangulu (1994) concluded that some databases like PADdab, PADexp, and PADinst are of little value, all of which are referral databases. Similarly, Seyoum (1993) had, a year before Kangulu, found out that the highest search pertaining to the bibliographic databases followed by the statistical ones, whereas the lowest number of search was manifested in the referral databases.

6.5. DATA SOURCES

The overall objective of PADIS, stated in the first chapter of this thesis indicates that all member of PADIS network would voluntarily contribute information to the information system/databases. PADIS (N.D)Information Resources also pointed out the suppliers of the listed databases. For example, UNECA and PADIS itself are the major suppliers to PADdev PADIS uses its own input to PADexp, PADinst referral databases; statistical division of ECA supplies input to the PADstat database, and PADIS uses UN press release information system and UN Department of Public Information as sources of input to the UNPRESS full text database.

6.6. USERS

In the same way as the data sources, the users of PADIS information system are indicated by its objectives stated in the first chapter of this thesis. Accordingly, the users of the system's information are the contributors of the input the member states, Institutions involved in development activities. It has also been indicated by PADIS (1993b) that PADIS users include individuals, institutions and participating centres of the PADIS network. PADIS has also announced the availability of the services, related to its immediate objectives, to non-governmental organizations as well as institutions outside the African region on a fee for service basis. Ultimately, although it is through participating centres, PADIS is designed to serve policy makers, planners and other information users working in the area of socio-economic development activities (PADIS 1993C).

6.7. SERVICES

Citing the 1982 PADIS project document, Kangulu (1994: 1-2) mentioned a number of services that PADIS would provide to African countries. The following are also listed in the 1993 version of PADIS document analysis manual as its services.

- Publication services
- Information Exchange
- Advisory Services
- Training and fellowships

One of the different categories of the activities that PADIS claims to undertake is

provision of user services. It includes printed outputs from databases, newsletters, selective dissemination of information, current awareness services, retrospective searches, Question/Answer services, hard copy and microfiche document delivery, downloading of databases on different media, consolidating and repackaging of information. It has been reported to the sixth meeting of its standing committee, that many of PADIS' efforts were directed at promoting the systematic collection, treatment and dissemination of scientific and technological and other development area information in Africa to make it accessible to interested users worldwide. To this end, PADIS in cooperation with member states and African sub regional and regional institutions have developed and adopted a series of standards including guidelines, manuals and computerized programs for database development (ECA 1996b). For example, PADOR is a program developed by PADIS to facilitate:

- Automatic creation of a database structure
- Data entry using customized worksheets
- Automated selection of authority lists by adopting the ODIN program for pick list management
- Index generation of purpose codes, sectoral codes, geographic codes, author index, shelf indexes; and
- Data export using comma delimiters by Fangorn facilities.

In general terms, creation and utilization of databases, microcomputer hardware configurations and electronic information exchange are service areas proposed by the Carnegie project being concentrated upon by PADIS(1994d).

Online access to PADIS databases is no more limited at ECA headquarters in Addis Ababa and by telecommunications link from organizations in participating centres where the databases have been installed but is possible through a pilot project of PADISnet, initiated under

a grant from IDRC. The project enabled the users at a node of the network to use modern communication to send messages, requests to database searches, exchange information over bulletin boards and participate in electronic conferences. The project also enabled PADIS to serve as a gateway for the network to access external databases and networks; and to be a testing centre for the applicability of recent IT including the use of optical storage media, scanners and facsimile transmission.

PADIS participating centres may also request the down loading of databases, and installation where necessary, on magnetic media including tape and diskette. Files may also be sent electronically to centres with the necessary equipment (PADIS 1993b). It has also been indicated in its manual (PADIS 1993) that members of PADIS network may request advisory, training and fellowship services from PADIS in fields relating to its immediate objectives.

Another service that PADIS is providing, as a compensation for the absence of full text databases is the document delivery service. The document listed in the bibliographic databases are available in the form of microfiche and hard copy on a request basis. Requests can be received through any means available to users including letters, telephone, telex, cable, facsimile or e-mail. PADIS also provides custom printed replies on the basis of searches of any of the databases. This is what it calls the question/answer services. The dominating information service of all the others provided by PADIS is the publication services. In addition to the newsletter, PADIS was publishing a hard format of more than five databases. To mention some

Databases	Publication
PADdev	DEVINDEX-Africa
PADexp	Directory of African experts
PADinst	Directory of African Development Institutions and

	its subset
	Directory of ECA-sponsored Institutions
PADdab	The Directory of African databases
PADstat	ECA statistical yearbook

Presently PADIS started to publish a CD-ROM database. This inclination to CD-ROM publication and Web-Site creation seem to take over the service dominance of distributing publications. For example, DEVINDEX-Africa will no more be published.

6.8. PADIS EVALUATION STUDIES

The different studies that have attempted to evaluate PADIS used its objectives set to be achieved at its different phases. An evaluation that has tried to assess the performance of PADIS at all of its phases is the one that was done by Kangulu, which is the last evaluation study known to the researcher of this study. Almost all of the evaluation studies conducted before the Kangulu's work have been reviewed by him. To show their difference from the present study the purpose of some of these evaluation studies are mentioned below including that of Kangulu.

The first evaluation conducted by Aiyepeku in 1983 was management oriented. It was intended to be used for policy decisions by the management of IDRC's Information Science Division. The study, therefore, reflected what PADIS was envisaged to do.

The second evaluation had an objective of assessing whether activities intended to be undertaken by PADIS during its first and second phases had been accomplished.

It was done in 1985 by PADIS. The study showed the progress in the establishment of PADIS-CCO. On the contrary, it noted the existence of the following three problems.

1. The difficulty in staff recruitment

2. The reluctance of NPCs to submit input to the PADdev database

3. The establishment of only 10 out of the intended 50 NPCs and the failure to establish almost all the intended (four) sub-regional centres.

Kangulu (1994) also cited/referred to Aiyepeku's 1988 general study of development information utilization in Anglophone Africa and pointed out that about 80% of the decision makers and development planners had been unaware of PADIS and had not used it.

In 1990 UNDP, the largest donor agency of PADIS, had requested a PADIS evaluation study that was conducted with a purpose of providing an assessment of the performance of PADIS Phase III. It specifically concentrated on the design of PADIS, its implementation, achievement and future direction.

This study reflected a significant number of achievements including: the increase in the number of NPCs from ten to 34, the increase in the number of databases designed and maintained by PADIS and the increase in the number of database search requests. The study suggested that given the financial and human resources, the objectives set out for PADIS phase III could have been achieved.

Justifying the significance of his work as the first evaluation study that covers the whole period of PADIS existence up to the time (1994) of evaluation, Kangulu conducted the fourth evaluation study with an overall objective of establishing the extent to which the objectives of PADIS in relation to the whole of Africa have been achieved.

Some of the points that have been uncovered by Kangulu are mentioned below:

- The shortage of staff at PADIS-CCO had led to the abandonment of some of the services of PADIS like the selective dissemination of information.
- Most PADIS NPCs had not sent any inputs to PADIS-CCO since they were established.

As a result most of the records in PADdev were for documents produced by the ECA and input by PADIS itself.

- No Sub-Regional database had been created even though it had been considered to be one of the functions to be undertaken by the centre. This was because of the fact that none of the Sub-Regional Centres (SRCs) were operational.

Some of the conclusions that Kangulu has drawn and are related to the present study are given below.

- Some of the PADIS methodologies are outdated. i.e, the PADIS methodologies used in recording information are difficult and unnecessarily long (this has been noted in the first part of this study).
- Recording of development information had not until then been standardised.
- People were not aware of PADIS as a source of development information and it is not used by the intended people.
- There is little information exchange among African countries.
- Definition of development information is inappropriate, i.e, Development information is largely defined in terms of the type of document, for example report, conference, paper and proceedings. Such a definition ignores the subject content of the document as the determining factor for the inclusion or exclusion of a document into the system. Furthermore some documents are excluded from PADdev purely on the basis of their physical format.
- The lack of effective participation in the PADIS network by many national statistical offices has severely affected the content of PADstat. On the other hand most of the national statistical offices do not have the necessary resources to collect, compile and distribute national statistics effectively on their own.

- Some of the databases, produced by PADIS, like PADdab, PADexp and PADinst, are of little value. They could easily be compiled by the National Centres for use within their own country.

The following are also some of the related recommendations forwarded by Kangulu.

- Core databases like PADdev, PADpro and PADstat, should be given priority over others. In particular the contents of some of the files in PADcom ought to have been downloaded and included in PADdev, because most of the PADIS users do not have direct access to files in PADcom.
- Efforts should be directed at making PADdev as comprehensive as possible.
- The available range of newer technologies on the market for the recording and dissemination of information should be explored and chosen from.

One of the priority areas that Kangulu put for PADIS is that PADIS should develop other methods for the recording and delivery of information. The standing committee (ECA 1996a) recommended that an evaluation of products and services would be useful before elaborating future activities. Partly, initiated by and based on some of the findings of the above mentioned and reviewed evaluation studies, the present study focuses on the technical assessment of the IR system developed (on the basis of CDS/ISIS) at PADIS and implemented on PADdev database.

6.9. FUTURE TRENDS OF PADIS

The Information committee of the conference of African planners, statistician, and population and information specialists (ECA 1996a) stressed, on the ninth session, that with the increasing availability of information on development issues, ECA/PADIS should make efforts to decentralize information access to reach users. Furthermore, information products needed to be developed that were easily available to the end users, both in accessibility as well as in formats they could understand. In recognition of the paucity of information infrastructure in Africa the committee decided to bring the issue of 'Information for all in Africa by the year 2010' to the level of the Conference of Ministers in 1998. Amoako (1996) pointed out that building Africa's Information Society is central to the strategic vision for the ECA and is shared by the ECA Conference of Ministers. Amoako (1996) also noted that the Ministers adopted an action framework, called African Information Society Initiative (AISI) to build Africa's information and communications infrastructure. The framework will be implemented at country level. This shows the environmental opportunity under which PADIS will operate. In this regard the new structure of ECA, puts PADIS under the Development Information Service Division (DISD). The division will serve as the centre for assembling, harmonizing and disseminating of comprehensive multi-sectoral development information to facilitate easy sharing and integration of data from different sources (ECA 1997). One of the sub-programs of the commission is to be implemented by the Development Information Services Division. The sub program (sub program 4) is named as, 'Harnessing information for development.' The major activities and outputs of the division can be classified into four broad categories (Discussion with PADIS staff and consultation of draft programs):

1. External relations and services

2. Group training
3. Development and maintenance of statistical and spatial databases,
4. Publication of catalogues, compendiums, directories and accession lists.

Regarding the publication plan of the sub program there will be about 10 recurrent, 8 non-recurrent publications and there are also 12 technical materials planned to be produced including the production of CD-ROM, establishment of infrastructure and the related info-structure for the creation of ECA Web-Site(continuation) and the continuation of acquiring, establishing and distributing dynamic regional data archives (ECA 1997).

Training plan of those that are specific to the division and require external consultant and training abroad include:

- Dynamic Web database development;
- Object-oriented programming
- Object databases,
- CGI and Java
- Wireless technology
- digital imaging and development of digital libraries
- host based technologies (routers, firewalls, etc.)

The first team in the division is named as 'Development Information Management' The activities under the team's program will focus on the establishment of regional databases and data archives service centres, catering for comprehensive information on African development. The activities will also be directed at promoting the development of linkable national databases. There is also a plan to acquire high end database management tools that can solve the problems/difficulties to make the statistical, referral and bibliographic databases of the division that are currently available in non-SQL and ODBC complaint database formats available to end users.

Generally the division is highly oriented towards IT. As a leader in information systems and primer centre of excellence in information services it was planned to be equipped with the-

state-of-the art information and communication technologies such as:

- Upgrading staff equipment to high end GUI standard;
- Full connection to the Internet;
- Upgrade of existing LAN at PADIS to Division-wide Intranet;
- Transfer of existing databases to ODBC complaint database systems;
- Digitizing library for which it requires a state of the art digitizing laboratory;
- Cyber information search site; and
- user training in up-to-date technologies.

CHAPTER SEVEN

DATA PRESENTATION AND DISCUSSION

7.1. ANALYSIS

This part of the chapter presents the analysis of the data collected about the IR system that is set to handle PADdev. Although the data on what the information requestors of the system may say on performance of this system was felt helpful, it has been found out that locating the information requestors and getting data from them needs time and a method that requires a continuous contact with them as they receive search services. Due to this, present evaluation relied on the data collected from the professional staff of PADIS who are managers, request intermediaries, promoters, searchers, database managers, maintainers and contributors for the operation of the system in one way or the other. The other data source of this evaluation is the researcher's close examination of the PADIS methodology and other documents related to the system; search and retrieval functions of the system; and the application (CDS/ISIS) that the system is presently using. To this end, questionnaires were distributed to all the nine professional staff members of PADIS. Seven of the distributed questionnaires were completed and returned with some partial response to the questions. One of the questionnaire could not be returned because of the respondent's continuous assignment to different missions outside Ethiopia. Two of the addressed professionals indicated that they have too limited relation with the IR service to be useful in its evaluation. The evaluation therefore, looks into what PADIS has done in the application of CDS/ISIS to handle PADdev; what problems stand in front of the development or operation of the PADdev IR system; what shortcomings it has; and what development should be

considered in the future both to accommodate PADIS' future plans and to benefit from the state-of-the-art development.

7.1.1. STORAGE PERFORMANCE

The first thing an IR system is required to do is to allow the user to store the data/information he/she wants to store. By storage performance here is meant the capability of the IR system in intermediating different hardware with the user in the process of converting data/information to electronic/machine readable form. The IR software that PADIS is using to handle any of its databases of development information related to Africa is CDS/ISIS. The facilities that CDS/ISIS provides for this purpose have been briefly reviewed in chapter three of this thesis. It has been indicated in the reference manual of the software that the users' effort to make use of the different facilities provided by the software determines its performance. In this regard, at the storage level PADIS created a Field Definition Table (FDT) and four Data Entry Worksheets to be used in handling PADdev. The former determines the number, type and size of data elements that can be recorded about a document. The latter is a type of electronic form which can be used by the user to code the value of each and every field defined in the FDT. The effort exerted by PADIS in the creation of these files is discussed below.

7.1.1.1 Field Definition Table (FDT)

As indicated above, FDT is a table that defines the fields of a given database and their characteristics. The creation of data entry worksheets and validation of the contents of fields of a database is controlled by FDT. When one is creating FDT one can determine the following features of an IR system

- The number of fields in a record;

- Whether each field can be subfielded and the delimiter of the subfields;

- Repeatability of each field;

- length (maximum size) and type of each field and;

- Tag of the fields

From the selected fields it is possible to categorize the type of documents that PADIS is handling, in PADdev, into the following categories:

- Monographs and chapters from them

- Serials and their components

- Multi volume monographs and their parts

- Proceedings

- Patents, and

- Academic works

The bibliographic fields identified for all of these document groups are about 38. Some are common, while others are unique to a group. In addition to the bibliographic fields, PADIS has also identified/included about 18 fields related to:

Identification of the physical document,

Content analysis and

Identification of indexers and data input personnel.

Some of these fields on the basis of different characteristics of the data/information, are repeatable and/or subfielded. Table 1 shows the number of repeatable and subfielded fields.

Table 1: repeatable and subfielded fields of PADdev

Field Category	No. of repeatable fields	No. of subfielded fields
1(110-170)	3	4
2(210-820	12	9
3(830-940)	2	1
4(950-970)		3

Keys: 1(110-170) - Bibliographic
2(210-820 - Content analysis
3(830-940) - Physical document identifier
4(950-970) - Indexers and data input personnel identifier

The fields are tagged with a minimum gap of 10 between fields of same category and 100 between fields of different category. The size of the fields of PADdev is limited to less than a quarter of the maximum field size that can be handled by CDS/ISIS (8000 bytes). The fields that the maximum number of subfields are the fields of personal author/affiliation and that of conference, meeting, symposium, which have four subfields. The response to the question that requests the respondents to judge the different features of the system shows that most of them (5) said that the availability of required fields to capture data about a document is good. This, in effect, means the respondents could input any data(bibliographic) about a document that they need to include into PADdev. Therefore, they feel that the storage capacity of the system is adequate for capturing data for almost all types of documents. One of the respondents said the feature is very good while the other one responded fair.

7.1.1.2 Data Entry Worksheets (DEW)

PADdev IR system uses four Data Entry Worksheets (DEW) as screen formats from which a data input operator can select according to the data she/he has at hand to input. To make a user free from remembering the name of different predefined DEWs, the system has a submenu entitled "PADdev worksheets" in the Data entry services menu of CDS/ISIS. By pressing 'T' after being in the menu of data entry services a user gets the list of the following worksheet names:

M - Monographs

A - Articles in Monograph

G - General (contains all fields)

J - Journal article

The user can select any one of these worksheets by pressing the appropriate letter. If required, the submenu provides a help option and an option to exit to the menu of Data Entry Services.

Option G represents a worksheet that contains all the fields in the database. The other worksheets contain fields related to the type of documents indicated by their name. The "Monographs" is a worksheet for all documents like books, reports, standards, audio-visual material that are in one or a known number of parts and part of a series or multi volume materials. The "A" and "J" options represent two different worksheets each containing groups of fields related to a chapter or other part of a monograph (which may be multi volume or else, series and their components; proceedings; patent; or academic works;) and an article of a journal, respectively. With most of the fields of the worksheets appropriate help messages have been associated. The questionnaire incorporates about 8 questions relating to this feature of the system.

The first one requests the respondents to judge about four data recording features of the system. The response is summarised in table 2 and table 3.

Table 2. Data recording capability of PADdev IR system

Features	V.Good	Good	Fair	Poor	V.Poor
Screen format	2	3	2		
Editor	2	3	2		
Updating facility	2	4	1		

It is understandable from the above table that most of the respondents were satisfied with the screen format of the system. This holds true to the other features shown in the table, in the sense that the screen format of the system could enabled them to identify the different fields to input the appropriate data and it provided them the fields they required; the editor and update records as they want.

Table 3 presents how the respondents rate the user-friendliness of the retrieval system of the PADdev database.

Table 3: User-friendliness of PADdev IR system

Rate	No. of respondents
Very good	3
Good	2
Poor	2
Very poor	

Almost all of the respondents did not answer the question related to its help facility. Only one respondent indicated that the help facility is not easily accessible and it has inadequate

content.

A question that asks the respondent if they think the methodology is complex to an average data input operator is answered No by almost all of the respondents. Only one respondent indicated that the methodology is complex to be used by an average data input operator in that there are unnecessary fields that confuse the input operators and the help facility is not adequate enough to guide them. Moreover, s/he pointed out that there is a need to back up usage of the system with extensive training programs. One of the respondents, who answered No, also indicated that the simplicity can only comes by the provision of basic training. From a question on the role of the respondents in the information storage and retrieval service provided by PADIS, it was understood that the doubt on the simplicity of the methodology is the view of top personnel of PADIS.

The other question is related to the respondents' satisfaction with the error checking mechanism of the system. The response shows that most (5) PADIS staff members were satisfied with it. Those who were not satisfied with the error checking mechanism of the system indicated that the error message is not as explanatory as it should be. In addition, a respondent pointed out that there are errors which the mechanism cannot check and it is not adequately documented. In this regard it is worth pointing out that PADdev IR system fully relies on the error checking facility provided by CDS/ISIS that refuses to accept any data different from the type of the field. For example, if the field is alphabetic it is not possible to input numeric data.

Self sufficiency of PADIS methodology to solve any information recording problem was affirmed by most (5) of the respondents. While it is negated by one. One of the respondents did not answer the question related to this. The respondents were also asked to judge the documentation of the methodology and its online help. With the exception of the two who judged

it as fair, most of them pointed out that it is good and very good.

The question related to the up-to-datedness of the system's user-interface is returned blank by two of the respondents. Three of the respondents indicated that it is very recent; while two of them pointed out that it is very out-of-date in view of their knowledge of the state-of-the-art development in this area. Regarding their interface preference, most (5) of the respondents preferred menu selection, which is the existing interface of the system, one preferred command language, and the other one preferred question and answer.

The last question related to the system's information recording performance, provides opportunity for the respondents to note their overall judgement about PADIS methodology and its information recording capability. Only two of them provided explanatory response. One of these two respondents indicated that it is out-of-date and not flexible to accommodate new tools of IR services while the other one pointed out that its user interface has not been kept up-to-date but the information recording ability remains good. The other respondents just said good and very good with the exception of the two who left the question blank.

7.1.2. SEARCH AND RETRIEVAL PERFORMANCE

It is notable, from a close examination of the system and the methodology, that PADIS exerted considerable effort to make the retrieval performance more than what the mere use of CDS/ISIS could do in handling PADdev. PADdev IR system has different files that facilitate the retrieval functions like the file of Field Select Table (FST), the different ANY files and stop word file. FST is the table that defines criteria for extracting one or more elements from a master file record. These elements will be used to create inverted file entries. To this end, The PADdev's FST is created out of a selected 16 fields including fields of the translated titles. Most of the selected

fields are associated with the '0' indexing technique which builds an element from each line of the format (mostly a field or sub field). The title fields and the descriptors field are associated with indexing technique '4' and '3', respectively. The former builds an element from each word of the format, while the latter builds an element from each term or phrase enclosed in slashes. It is notable that none of the extended CDS/ISIS indexing techniques are used and one of the data extraction formats is made for a non-existing field tag (70^a|). The incorporation of a linking field (field 140 Relationship to other records) is also an aspect of information retrieval facility because like a hyper record facility it displays one or more records that are not retrieved but linked through this field as part(s) of the retrieved record. More importantly PADIS developed a program that enables the user to search the database online through e-mail connection.

The first question in section III of the questionnaire, related to the search and retrieval performance of the system asked if it is easy to interpret users request to a search strategy and is responded by 5 of the respondents. One of the respondents indicated that (sometimes) users' way of expression is much complex than the search formulating capability of the system and noted that it is not possible to answer Yes or No. The other respondent noted that it is up to the searcher. Three respondents answered Yes to the question. One of those two respondents who answered No indicated that users' search request is difficult to understand. It is possible to note, therefore, that the search formulating facility of the system satisfied most of the respondents. All the five respondents to the second question, related to the retrieval feature of the system, indicated that the search formulating capability of PADdev retrieval system is as flexible as required.

Moreover, four of the respondents ranked this feature very good(1) and good (3), while only one of them indicated that it is fair. But still there are three respondents who think that cognitive and/or Best-match (natural language supported) search systems are required to further

enhance its search formulation facility.

Only two of the respondents pointed out that it is rare to complete a search for a request at a shot, while three answered 'sometimes' and two answered 'Always': The response to the question on the reason of re-conducting by refinement of the search strategy, in most cases is summarized in table 4. One user left this question blank indicating that it does not apply to his role.

Table 4: Reason of re-conducting a search

Search result	No of respondents
Very large	5
Very small	1
Irrelevant or unrelated to a request	1

The table shows that the precision of PADdev IR system is poor, especially at the first attempt of the search in that the respondents get a very large search result most of which are irrelevant to the information requesters need. This forced the intermediaries re-conduct the search repeatedly.

In another question most (4) of the respondents indicated that their effort on re-conducting the search repeatedly makes a significant difference on the size of search result and they often use previous search strategies for the purpose. Moreover, the response to the first question in section II of the questionnaire shows that almost all (two ranked it as very good and four of them said good) of the respondents were satisfied with the system's feature of previous search refinement. In response to the same question two respondents judged the document ranking feature of the system as very good and two of them as good while three of the respondents judged this feature

as very poor.

The way that the respondents refined the search strategy, when they re-conducted the search, is tabulated below.

Table 5: Methods of refining search strategy

Methods	No. of respondents
Using/changing boolean operators	4
Changing the scope of different terms	2
Using proximity search strategy	2
Using range search method	1

Although, using dictionary search, selecting terms from retrieved records are options in the question, no respondent answered these options and no respondent pointed any other method. It may be because they are familiar with dictionary terms of the system and using it does not change the result and the other option is not available in the system.

From table 5 we can generalize that search refinement is being done by the use of different boolean operators. In this regard, we have to notice that CDS/ISIS is an IR software that relies on boolean search method.

Compared to their knowledge of the state-of-the-art development in the area of search languages, most(4) of the respondents think that the available search languages are enough to conduct a search on any database effectively. One of the two respondents who answered No to the related question pointed out that at least search languages of SQL family are lacking from PADdev IR system.

Regarding the time they spend on a single searching, the respondents' answer range from 5 minutes to 30 minutes. The variation of the time requirement of a search may have resulted

from the fact that it can be affected more by the individuals typing skill, analytic thinking, concentration and other factors than the retrieval system's response time. But still, we have to notice that retrieval speed can be increased by automating different routines that require the searcher's effort.

There is a question that is directly related to the recall level of the system. Three of the respondents to this question think that they could retrieve all the relevant documents available in PADdev for all the requests they conducted the search. One of the two respondents who doubted the system's full recall attributed the existence of fallouts to the record linking problem. Yet most (4) of them ranked flexibility of the record structure(which means the structure allows a record to be linked with different other records and to change the constructed links in the way users need) as Very Good(1) and Good (3) except the two, who ranked it as poor and very poor. Regarding its precision, only four respondents answered the related question and believe that they could (two always and two sometimes) retrieve only the relevant documents for a request without any noise.

7.1.3. OUTPUT PRESENTATION

One of the IR functions in which CDS/ISIS requires users' effort and give opportunity for the user to create his own format is the function of output presentation. Due to this PADIS exerted a significant amount of effort to enable the user of PADdev IR services to have different predefined display and print formats. In this regard three display formats are developed and made easier to select. To this end, a submenu of the list of the formats has been introduced in the standard CDS/ISIS "change display format" of "Information Retrieval Services" menu. Pressing the 'C' option of the submenu displays the following options any one of which can be selected.

L - short representation

B - Long representation

T - Diagnostical (field numbers)

S - Give your own format

X - Exit

The short display format enables the user to have brief display of a record. It displays only 8 selected fields of the record.

The long display format provides detailed display of the records in the database. It, therefore, consists of most of the fields of the records. The Diagnostical format is intended for the diagnosis of a record for data entry error. Therefore, it displays the record as it is entered together with any of the control characters. The user is also given option of using his own display format. To make a user free from specifying the sorting and printing requirements from scratch, PADdev IR system provides an additional option (I-PADdev indexes). The option enables a user to produce various print/save outputs in predefined formats. When a user presses "I" from the

options of standard CDS/ISIS sorting and printing services menu, the following list of options will be displayed and can be chosen from.

S - Sectoral code index

P - Purpose code index

A - Authors index

K - Keyword index

G - Geographical index

H - Shelf list

C - Comma delimited export

E - Explanation

X - Exit

The index options enable a user to generate different indexes corresponding to their name. The H option generates a shelf list. The comma delimited export option intermediates the PADdev IR system to exchange data with other application programs such as xBase, Lotus 1-2-3 etc. with the E option the system associates explanatory text on the other menu options. Some of the CDS/ISIS tools that PADIS used in the creation of the different display formats include:

- Indention command;
- MFN command;
- Mode command;
- Spacing command;
- Conditional literals;
- Boolean expression of Absence of a field; and
- IF command.

PADdev IR system also has a program that installs all the files that it has developed to

facilitate CDS/ISIS-based database development discussed above.

Given the effort of PADIS on the output function of PADdev IR system, its staff members were asked different questions related to the output capability of the system and their response is presented below.

The first question is related to the users' request for output forms. Most of the respondents (5) indicated that users request the search result to be presented in hard copy (paper form). Those who pointed the "soft copy" indicated that users ask for CDS/ISIS, MINISIS, word processing and ASCII output formats. But nobody answered dBASE, FoxBASE, INMAGIC or any other application formats.

All of the respondents (4) to the second question of section IV indicated that they can download or save search results without any problem as requested by the users. Almost all (6 out of 7) respondents to the first question of section II of the questionnaire ranked the presentation of search output of the system as very good(2) and good(4), with the exception of the one who ranked it as fair. They have also indicated in their response to the third question of section IV that no output format compliant have been received from the user. Most of the respondents to the first question of section II ranked compatibility of the system with other IR systems as Very Good (2) and Good(4), while only one respondent indicated that it is very poor.

7.1.4. PADDEV IR SYSTEM IN GENERAL

The response to the general commentary questions on PADdev IR system is presented as follows:

Most (5) of the respondents to the question related to the utilization of PADdev database negated its well utilization. The reason that most of the respondents indicated is lack of publicity and therefore users' less awareness of its existence. Some of the respondents also attributed the problem of less utilization of PADdev to its being outdated, limited accessibility, and users' preference to libraries. Some of the factors that the respondents thought as causes of less adoption of PADIS methodology are:

- insufficient resources to introduce and maintain it;
- problem of follow-up and transfer or loss of trained staff;
- it is not based on an open technology;
- it can have less outreach and is based on non-connectivity;
- PADIS can not and is not intending to enforce its entire adoption.

On the package adopted for the system, respondents were asked if they preferred ISIS packages to handle PADdev with no consideration of financial constraints. Most (4) of them negated the idea and pointed out that they would have gone for SQL, ODBC complaint, windows based dBASE and/or access. But most (4) of them don't expect ISIS for windows will have any problem to solve and they think that PADdev IR system is up to the expected technological development. Of course there are three respondents who negated this thinking.

The respondents have also been asked if the CD-ROM production of PADdev solves any problem. Four of them said No while three respondents answered Yes. Those who answered Yes pointed out that it can solve storage and access problems.

The last question that gives opportunity for the respondents to make additional comments, opinion and facts was answered only by a respondent who would like to see PADdev on interactive web site and be accessible to all users.

7.2. DISCUSSION

Data has been collected on the different features of PADdev IR system from different sources. The collected data have been analyzed in section 7.1 of this chapter. This part is planned to present the researcher's interpretation of the described data.

In the first part of the analysis section what PADIS did are presented, related to the data/information recording feature of its IR system specific to PADdev database; and what the staff members, as intermediaries of the system, feel and observed in the case of their handling the system.

The two files, Field Definition Table (FDT) and Data Entry Worksheets (DEWs), created by PADIS in this regard are part of its methodology that governs any information system, including libraries, which need to adopt the methodology, what to record about a document and to some extent what type of documents to include in its database. The fields included determine the purpose for which the record, and therefore, the database can be used.

In this regard, it was observed that the FDT created and being used by PADdev IR system has none of the fields into which any of a library's purpose, activity, or house-keeping functions can be recorded. Even to be used by development information systems, the FDT created for PADdev requires more fields that can be used to record (even in a descriptive way) the different attributes/features of several other document formats like films (both documentary and

commercial), cassettes (which may require such attributes as the period, in what mass media it has been presented, etc.), archival manuscripts being produced by different organizations, museum objects, cultural and traditional saying, lecture notes, and other similar documents that are difficult to categorize to any of the document groups that PADIS decided to include in PADdev. It is clear, and has been commented by Kangulu (1994)that development information should not be understood in such a way that discrimination will be made on the bases of the physical format, situation of their existence, and any other feature of a document but its content.

It is notable, therefore, that the FDT created by PADIS and is being promoted to be used widely in Africa requires more fields that will enable users to record data/information about the documents that are dominantly existing in the continent. Moreover, the different fields that are considered in the 1996 version of the manual have not been included in the FDT that is being implemented using the PADOR program.

The number of optional fields included in the PADIS methodology is very small and limits its flexibility. Of course it is understandable, from the critics on the Common Communication Format (CCF), that the number of optional fields make a format complex to be used as a data entry format. But it is possible to make it simpler by creating different optional Data Entry Worksheets (DEWs) more than what PADIS has so far done.

The satisfaction of most of the respondents with the availability of required fields to record data/information about any type of documents may be because PADIS does not provide them with documents different from the type that is predefined to input/include to the database. In this regard, getting response from different information systems which adopted PADIS methodology, is better considered and is expected to uncover several suggestions of data/information recording.

The contribution made by PADIS in upgrading the user-friendliness of PADdev IR

system through the introduction of a submenu into the menu of data entry services should be appreciated because it facilitates automatic selection of any of the PADdev's DEWs. The creation of different DEWs is also an aspect that makes the data entry work easier by providing alternatives with selected fields related to only a type of documents. This makes a data input operator free from the burden of jumping over the unnecessary fields in recording different types of documents.

Of course, it is clear that any group of fields considered by a DEW comprise of a record and is indexed and retrieved independent of any other fields that are not considered in the worksheet. Therefore, it should be linked to the other records very carefully and exhaustively so that it can be retrieved (displayed) by different users' approach to information search and/or can retrieve/display the other related records in the database.

Different writers have pointed out that CDS/ISIS is user-friendly and has satisfactory editor for the users who are familiar with it but it is the other way for novice users. The former idea is supported by the satisfaction of most of the respondents. The same effect is reflected in their thinking of the simplicity of PADIS methodology which can be used by an average data input operator. However, it is contradictory to the fact that the researcher himself faced difficulty in trying to create a record using the methodology. To illustrate some of the difficulties: the need to remember PADIS purpose, Geographic, sectoral and other codes; the need to have basic knowledge of CDS/ISIS. Of course top positioned respondents, as it can be understood from the first question of section I, recognize the fact that the methodology should be accompanied by an extensive training on its use and it is otherwise complex for an average data entry operator. Moreover, the PADIS training program coordinator attributed the less adoption of the methodology to the lose or transfer of trained personnel from the National Participating Centres(

NPCs). Besides Kangulu (1994)found out the fact that the methodology had not been adopted by a good number of NPCs because of its unnecessary fields like PADIS purpose codes. In this regard, making the methodology communicative and simpler is better than investing on training programs.

Although the respondents who think that the user interface (the interface of user-system interaction) is very recent/up-to-date, were more than the respondents who indicated the out-datedness of the interface, it is worth noting the state-of-the-art development in the area reviewed in chapter three of this thesis.

Regarding the search and retrieval performance of the system, the FST of PADdev revealed that some of the fields that contain searchable terms/concepts about a document like abstracts, development project, patent holder, etc. are not inverted by the FST.

Most of the respondents were satisfied with the search facility and flexibility of this facility of the system, and they think that search languages of CDS/ISIS are enough to search any type of databases effectively. However they have indicated the fact that cognitive and/or best-match and other interactive search formulating technologies are required to enhance search formulation performance of PADdev IR system, which indicates the need to incorporate IR technological developments to the system.

Although only two of the respondents indicated that they completed searches with the first search conduct for a request only rarely, most of them indicated that the searches were either very large or contain irrelevant documents to the request and they are required to re-conduct the search by using or changing different boolean operators. Of course, most of them need less than 20 minutes which they could tolerate.

Although the recall and precision of an IR system is evaluated quantitatively, it was pointed out by most of the respondents that PADdev IR system has good recall and precision in

that they could mostly retrieve all the available documents relevant to a request and they could retrieve only the relevant documents without any noise which can be achieved only by repeatedly conducted search. This could, however, be handled by the incorporation of expert intermediary systems like THOMAS.

As for the information recording feature of PADdev IR system, PADIS created predefined output formats to make its methodology adoptable and be used directly; and to create a standard of presenting IR system outputs. Among the predefined output formats the "comma delimited export" format intermediates the database with other databases developed by other application like dBase and Lotus for the purpose of data export. But the respondents have indicated that users request search results in paper form mostly, even those who requested soft copy did not request any of the applications which can be intermediated by this data exporting format.

PADIS has utilized different features and tools of CDS/ISIS to handle PADdev. But still there are capabilities of the software that would have been exploited to upgrade the performance of the system.

The factors, pointed by the respondents, that affected the utilization of the IR services of PADdev are more of promotional problems. This confirms some of the findings of Kangulu (1994) that indicated that very limited number of potential users know about the existence of PADIS as a source of development information.

Although some of the respondents recognized the problem of storage space and accessibility of the database that can be solved by the production of its CD-ROM version, most of them negated its problem solving capability. Of course, it is notable that although CD-ROM increases database accessibility and to some extent problem of storage space, it has problem of recency because incorporation of new records to the database is possible only with the next

version/publication, which is not the case with a networked database.

In this regard, Distributed Database Management technology is worth considering by PADIS so that the plan of Development Information Services Division that states 'Information for all in Africa by the year 2010' can be materialized by making PADdev accessible to lots of African information users and be updated easily. This is what has been commented by one of the respondents.

The IR features that PADIS would have considered to upgrade the performance of PADdev IR system are presented bellow, including programming facility of CDS/ISIS and its capability to be extended to the object-oriented level; the different programmes of different scale that are developed to assist the software; and the different technological development of IR systems.

7.2.1. PADDEV IR FEATURES THAT WOULD HAVE BEEN CONSIDERED

With due regard to what PADIS has done in the creation, maintenance and retrieval of different databases of development information, some features of IR system that would have been considered by PADIS for a better performance of PADdev IR system are worth reviewing. To this end, this part reviews what could have been done to utilize the IR capabilities of CDS/ISIS more, what could have been done to benefit from the technological development in the area of IR technology and what is felt by PADIS staff. It was pointed out in this thesis and other literature that CDS/ISIS provides different facilities and tools to create and maintain textual databases and provide retrieval services. Some of its facilities and tools can only be used directly while most of them require users' effort to make use of them to their optimal capability. In this line, PADIS would have considered the following points in the use of CDS/ISIS to handle PADdev database

with a better performance of its IR system.

In the creation of FDT and facilitating data entry function of the system, PADIS would have developed a program that facilitates repeatability of subfields within a field. Some of the subfields that require such a program include:

- name of the library within the field of 120 (location of document) because different libraries, which use the same classification scheme, give same call number for the same document.
- bibliographic level of the host document of which the document is a part within the field 130 (Bibliographic level indicator) because a document of the record may be a part in different host documents, for example an article published in a journal and reprinted in a monograph of key papers.
- all the subfields of the field 140 (Relationship to other records) for the same reason as the above one.
- all the subfields of the field 210 (personal author and affiliation) because different authors may have same affiliation, different affiliations of different authors may be in the same country, etc.
- the same for the subfields of 220 (corporate author), 310 (personal author of source document), 320 (corporate author of source document).

It would have also been possible to facilitate different proximity searches for a word or words in a part of long texts like abstract and summary by making their fields repeatable when the FDT is created. This would have had a facility for less complex natural language search.

CDS/ISIS allows a user to change/update the existing FDT of a database when required, but any major modification of the FDT requires changes in the database. Although CDS/ISIS

handles some of these change requirements automatically and through the use of its ISISXCH import/export services, most of them are to be made manually. However, it is clear that making such changes, as changing the type of the value already recorded before changing the type of the field, deleting all the extra occurrences of a field which is later made non-repeatable, for example, is cumbersome and needs time for large databases. In this regard, PADIS would have exerted some effort to automate all the routines of making the changes accompanied with different modifications of the FDT.

In consideration of some problems inherited by CDS/ISIS different performance increasing programs have been developed in the different corners of the world that PADIS would have incorporated the helpful ones to PADdev IR system. To mention some:

- Mola Hunegnaw's System Interface Search Assistance (SISA) developed at the School of Information Studies for Africa (SISA). It enables an information requestor to conduct a search on different databases easily and summarizes the history of the searches made by each user in each of the databases.
- Conversion interfaces developed by Neelameghan's group at SISA together with ABNCD format. It is, of course, an FST that intermediates the conversion of data developed by other formats to ABNCD format and vice versa.
- There are also programs for converting data from widely used record formats to the ISO 2709, and also from databases created with DBASE, INMAGIC and other international databases such as chemical Abstracts, BIOSIS, and AGRIS to ISO 2709 format (Abebe et. al. 1992).

It would have also been possible to automate control of capitalization at least for some fields like those related to personal and corporate authors.

CDS/ISIS would have also been used to handle all the PADIS databases either by an

interfacing program or by developing PADIS methodology into a database integration format, like ABNCD, and bring all the databases together.

After a number of searches are conducted, it is possible in CDS/ISIS to recall and display the result of any of these previous searches. But the search needs to be re-executed and made current to be displayed. In handling large databases, this is a time taking routine. Therefore, PADIS would have developed a program that could hold all the information required for the display of a search result along with the associated search statement so that any previous search results could be displayed directly. The other feature of CDS/ISIS that adds time requirement with the size of the database to be handled is updating the inverted file. PADIS is using CDS/ISIS to handle a significantly large database - PADdev. Therefore, Inverted File update would have been automated by a program that could enhance the updating speed or automatically update the inverted file after any single user action that calls upon the update.

Regarding space utilization, what can be referred to as a problem of CDS/ISIS to handle large databases is the space required for the inverted file. PADIS would have solved this problem by the use of compression techniques.

It has been discussed in chapter two of this thesis that it is possible to develop object orientation with CDS/ISIS and therefore would have been considered by PADIS so that the users of its IR services could have gained the benefits of object-oriented technology.

The above-discussed considerations that would have been looked into by PADIS are those associated to the use of CDS/ISIS. But a close look into the state-of-the-art development in the area of IR and the dramatically increasing requirements of handling development information uncovers the following IR technological developments that PADIS would have considered to further enhance IR services of development information provided on the basis of PADdev.

The developments and benefits of developing full text, relational and object-oriented databases have been reviewed in chapter two of this thesis along with the capabilities and benefits of Distributed Database Management Systems (DDBMSs). Chapter one of this thesis also points out Heaney's proposal of converting bibliographic formats like MARC into their object-oriented version. As PADIS is responsible to handle databases for a continental development information, it would have considered the harvest of such technological developments.

Considering the benefits of full text databases and IR system and the achieved technological developments related to full text IR systems, reviewed in chapter three, PADIS would have started to incorporate full texts into PADdev. Together with the consideration of full text IR services or even for the bibliographic ones PADIS would have introduced IR applications those which facilitate keyword frequency searching, natural language and cognitive search formulation, and application of artificial intelligence to their user interface. The development of these IR technologies and the possibilities of combining hypertext based searching with that of inverted file based and/or multimedia have been reviewed in chapter three of this thesis. The chapter also reviews the development of concept-based search, free text scan, knowledge based, Best-matching and object-oriented IR services in the first quarter of the 1990s.

The staff's response to a question related to their feeling about what would have been done, indicates that most (4) of them think that natural representation of database item's relationships using object orientation, incorporation of hypertext search facilities and full text search facilities could improve the IR capabilities of the PADdev IR system. Most of them (4) also recommended PADdev database and its retrieval system would have been developed to a full text with a hypertext and/or hypermedia retrieval facilities.

Moreover, six out of the seven respondents to the related questions indicated that there was a need to incorporate a relevance ranking facility to the PADdev IR system. The same

number of respondents recommended PADdev to be a distributed/decentralized database and integrated with the other databases of PADIS. As benefits of the integration, the respondents pointed out that searching each database for different or same request separately would have not been required, and the other databases would have contributed more in satisfying users' information needs.

Most of the respondents expect a significant change/increase in the use rate of PADdev IR services in the near future by its accessibility through the Internet. Some of the respondents recommended the incorporation of dynamic database interfaces such as Web database and JAVA Scripts to accommodate the resulting change/increase in the use rate of the database.

CHAPTER EIGHT

CONCLUSIONS AND RECOMMENDATIONS

This chapter presents the conclusions drawn from the findings/discussions of the thesis and the researcher's recommendations.

8.1. CONCLUSIONS

In line with the objectives of the thesis observations made on each of its chapter are presented to give bases for the conclusions that are to be drawn.

From the review of different database types and different approaches to database designs, it is possible to observe that:

- Different types of information is organized into different types of databases that need different information retrieval capabilities.
- Different database types are developed for different purposes. For example bibliographic databases can be used only to refer the user to other sources of information such as documents and organizations or individuals.
- Integration of a number of different types of databases is possible to get the benefits associated with the different types of databases and different ways of integration are available.
- There are DBMSs like DDBMSs that enable an effective data sharing between geographically scattered information users with a minimum delay of access and storage space cost.
- Different information retrieval capabilities are associated with different approaches to

database designs.

- Object-oriented approach to database design now joined the other two competing approaches, Viz relational and textual, with all its popularity and other emerging facilities.

In the same way the following can be noticed from the review of information retrieval systems presented in chapter three of the thesis and theoretical background of IR evaluation presented in chapter four.

- The development of information retrieval capabilities from handling only bibliographic data to handling a knowledge base at any degree of complexity is facilitated by the accompanying development of computer hardware, the price of which is affording.
- In these days IR systems are capable to communicate with their users in any way that the user is in need. This development of IR capabilities is encouraging the user community to rely on information for every of their decisions by avoiding the information anxiety that results from the difficulty to get the required information.
- The use of local and external IR systems has merits and demerits associated with each of them. Therefore, the better alternative is to facilitate access to the different local databases. This can, of course, be facilitated by applications like Distributed database Management Systems(DDBMS).
- Although they are being revised rapidly, ISIS family IR systems in general and CDS/ISIS in particular lack a lot of information retrieval capabilities.
- Evaluation of IR systems is continuously shifting from concentration on the algorithmic retrieval capabilities of the system towards looking into the overall capabilities of an IR system. As a result IR evaluation researches are coming out of the laboratory environment to the real world and started to take the different features of the system as

criteria of evaluation instead of limiting themselves to recall and precision.

- Evaluation of IR systems may be done either for academic purposes to improve understanding or application/action purposes to bring improvement. But both of the evaluation types are meant to enable information systems to take action against any failure of IR systems.

It is also possible to note the following points from the discussion on the search types and database formats.

- As they are aspects of an IR system the search type and the formats that the system adopts affect its performance. The former introduces the human factor of the system, while the latter incorporates the rules that guide the functions of the system.
- Although the degree varies with the system's interactivity the performance of an IR system depends on both the intermediary's knowledge of users information need and the system's search languages; and the users' knowledge of the subject area of his/her information and/or the system's search languages.
- The degree to which the format it adopts is open-ended can determined flexibility of an IR system. For example ABNCD gives ways for a system to be used either as an IR system or as a library housekeeping system or both more than the PADIS format/methodology. To what degree the format is open-ended in turn is mostly affected by the number of fields that it allows to be included in a record and flexibility of relationships it can create between records.

From the reviews of the different aspects of PADIS, including its future trend it is observable that:

- As an international development information system, PADIS is handling different databases of African development information on the basis of which it provides

publication and search services.

- Different evaluation researches have indicated PADIS' successes and failures at the implementation of its different phases. But that of Kangulu's emphasized some shortcomings of the system. His findings have now provided some inputs for the new Development Information Services Division under which PADIS is now structured.
- The plans set for the Development Information Services Division require the acquisition and incorporation of the state-of-the-art development in the information handling and processing area, which would have been considered by PADIS.

The following are conclusions that are drawn on the bases of the above observations and data analysis and discussion presented by chapter seven of this thesis.

- PADIS has done what is to be done to use CDS/ISIS to handle a bibliographic database at a time and to provide most of the services it has planned at the outset. It has also developed a program that enables users to search PADdev online through e-mail connectivity using the new networking feature of CDS/ISIS.
- PADIS methodology provides different facilities that enabled PADdev IR system to use CDS/ ISIS straight forward and it can now be installed automatically by a program developed by PADIS under the name PADOR.
- PADIS' staff members, whose day-to-day functions are tied to the system, are satisfied with the capability of PADdev IR system to handle a bibliographic database and its facility for the services that are based on such databases. PADIS upgraded the user-friendliness of PADdev IR system more than what it would have been by the mere use of CDS/ISIS.
- However, the methodology can handle only bibliographic data of mainly information

sources of paper format. It is also not able to incorporate dynamic relationships between different fields of different databases even, of one database.

- PADIS methodology, therefore, requires features of database integration so that it can satisfy the users need to integrate PADdev with other databases of PADIS. It also requires features of handling full text and dynamic data, and its flexibility needs improvement to facilitate the development of PADdev IR system to the level of full text and object-oriented database, and to enable the system incorporate features of hypertext, hypermedia and other IR technologies of the-state-of-the-art development.
- CDS/ISIS provided facilities for data/information recording and search output presentation to PADdev IR system that could satisfy the intermediaries. Although its boolean-based search and retrieval capability is also satisfactory, there is a need to incorporate interactive user-interfaces to its search formulating function; and hypertext, best-match, SQL, and other non-boolean retrieval technologies to its retrieval feature. Some of its features like interactivity are further upgraded by different programs developed in CDS/ISIS Pascal in the different corners of the world. PADIS would have, therefore, explored these programs and incorporated them to PADdev IR system to enhance its interactivity. The other features like SQL can be exploited by making the system use other DBMSs like dbase in parallel with CDS/ISIS. The incorporation of the recent IR user interfaces and retrieval features like best-match, full text scanner need to be explored. From the professional intermediaries point of view, PADdev IR system, however, is found to be user-friendly.
- Although it is gained by conducting searches with refinement of search strategies repeatedly, the effectiveness of PADdev IR system, in terms of " recall" and "precision" is satisfactory, from the professional intermediaries point of view. PADdev IR system has

got satisfactory predefined output formats for different purposes, such as printing indexes, saving or exporting records.

8.2 RECOMMENDATIONS

On the basis of the reviews made and the data analysis and discussions of what is being felt about PADdev IR system, the following considerations are recommended.

- PADIS should make PADdev IR system able to handle all development information in any form of their existence instead of limiting it to textual bibliographic information.
- PADdev should be made capable of handling the dynamically changing requirements of development information services. To this end, therefore, it should incorporate features of object orientation. Moreover, PADIS methodology should be revised to its object-oriented version.
- PADdev should be integrated with all the other PADIS' databases which may require the update of PADIS methodology to make it include all the fields of the other databases.
- PADIS should be a coordinator of decentralized African development information systems that develop and maintain development databases and make their database accessible to the other information systems, especially its Sub-Regional Centres(SRCs). They should, of course, have a good access to the databases at PADIS. All the databases handled by PADIS, especially PADdev, should be made accessed by different African networks and the Internet. So that their utilisation rate will be improved.
- To materialize these recommendations, PADIS should either develop or acquire the required software, and of course hardware, that can handle large decentralized databases

of full text, sound, image and other form of development information. These newly coming software and hardware technologies should, however, be evaluated for their compatibility with the existing ISIS-based IR system and their capability to enhance the different features of the system.

- PADIS should explore recent developments that upgrades user-system interaction of African IR systems and incorporate or develop the ones that enable any African development worker to communicate with different information systems so that information anxiety will be minimized, if not removed from Africa.
- To avoid the need to train all of the potential and actual users of PADIS methodology and to minimise the influence of intermediaries' analytical skills, search strategy formulation; typing skill, etc. on the performance of the system, PADIS methodology and PADdev IR system should be developed to their interactive version by incorporating appropriate IR technologies.

It is understandable that the above recommendations, require time and long range financial plan which go along the plans of Development Information Services Division. The following are, therefore, recommendations that should be considered immediately.

- The different programs that can upgrade the performance of different features of CDS/ISIS, which are developed in different corners of the world , should be well evaluated and incorporated to PADdev IR system.
- The potential capabilities including programming facilities, of CDS/ISIS should be utilized to further upgrade the performance of PADdev IR system.
- PADIS should update its methodology by:
  - adding different optional fields to make it flexible and accommodate different fields so that it can be adopted by different information systems, including

libraries, to use it for various purposes. Of course, there may be a need to add some more Data Entry Worksheets of fields of different purposes.

- adding a field of maximum size which can be a record of summaries or abstracts longer than 1500 characters and link it to the record of its bibliographic information. This can make the database more informative.

- PADIS should develop different programs that can handle the different update anomalies and data conversion between PADdev and any other databases developed by other formats and applications but which may be sources of input to PADdev.

REFERENCES

Abebe Rorissa. et.al.1992. ABNCD+: A prototype for an integrated information storage and retrieval system. Microcomputers for information management. 9(3): 137-159

Amoako, K.Y. 1996. Africa's Information Society Initiative: An Action framework to build Africa's Information and communication Infrastructure. In Global Information Infrastructure Commission (GIIC) Report on the Global Information Society and Development Forum. Midrand, South Africa. 14 May 1996. Washington CSIS.

Belkin, N.J.; P.G. Marchaetti and C.Loal. 1993. Braque: Design of an interface to support user interaction in information retrieval. Information Processing & Management 29(3): 325-344

Bourdeau, Robert H. and Betty H.C. Cheng. 1995. A formal semantics for object model diagrams. IEEE transactions on software engineering 21(10): 799-821.

Chen, Patrick Shicheng. 1994 on inference rules of logic based information retrieval systems. Information processing & management. 30(1): 43-59.

Chowdhury, G.G. 1994. Information retrieval system. Calcutta: Indian Association of Special Libraries and Information Centres (IASLIC).

_____. 1996. Record Formats for integrated databases a review and comparison. Information development. 12(4):218-223.

Cleverdon, C.W. 1978(1974). User evaluation of information retrieval system. In :154-165.

Convey, John. 1989. Online Information retrieval: An introductory manual to principles and practice. 3rd ed. London: Clive Bingley.

Cooper, Williams. 1973. On Selecting a measure of Retrieval effectiveness. JASIS. 87-100

Croft, W.Bruce, and Howard R. Turtle. 1993. Retrieval Strategies for hyper text. Information processing & management 29(3): 313-324.

Deco, Claudia; Cristina Bender; and Francisco Crespo. 1995. Why microisis?: Textual databases versus relational databases INFOISIS. 1(3): 42-55.

Di Lauro, A. and E Brandon ed. 1990. Manual for preparing records in microcomputer-based bibliographic information system. Ottawa: IDRC.

ECA. 1997. Revised programme of work for 1997. Activities by category. ECA.

__. 1995. Promotion of norms and standards for the exchange of development information in Africa. Ninth session of the conference of African planners, statisticians and population and Information specialists. Addis Ababa. 11-16 March 1996.

__. 1996b. Why standardize textual and statistical information? 7th meeting of the standing committee on standardization and harmonization of Information systems in Africa. Addis Ababa, 18-20 November 1996.

__. 1996a Report of the ninth session of the conference of African planners, statisticians, and population and information specialists. Ninth session of the conference of African planners, statisticians and population and Information specialists. Addis Ababa. 11-16 March 1996.

Ellis, David. 1990. New horizons in information retrieval. London: The Library Association

_____. 1994. Paradigms in information retrieval research. In Encyclopaedia of library and information science. Supplement 17. Vol.54 edited by Allen Kent. New York: Marcel Dekket, Inc.

Fayad, Mohammed E. et.al 1994. Adapting an object-oriented development method. IEEE software. May (1994): 68-76

Fox, Edward A. 1989. Research and development of information retrieval models and their application. Information processing and management. 25(1): 1-5.

Gauch, Susan. 1992. Intelligent Information retrieval: An introduction. JASIS. 43(2): 175-182

Gredley, Ellen; Alan Hopkinson. 1990. Exchanging bibliographic data: MARC and other international formats. Ottawa: Canadian Library Association.

Gilchrist, Alan. 1996. Blissful spiders: Forty years of information retrieval. Managing Information. 3(5): 35-36

Heaney, Michael. 1995. Object-oriented Cataloguing. Information Technology and Libraries. 14(3): 135-153.

Hillary, Retting. 1996. The Jasmine jump. VAR business 12(19): 37-

Hopkinson, Alan. 1991. Information Transfer and exchange formats. In standards for the international exchange of bibliographic information. Edited by I.C. McJlwaine. London: The Library Association. P.169-176

Ingangi, Francis. 1991.State of art report of the information sector in Africa. Paper presented at a seminar on National Policy on Information Systems and services in Ethiopia 27-29 Nov. 1991. Addis Ababa, Ethiopia.

Jusu-Sheriff, Gladys. 1991. Standardization: An African viewpoint. In standards for the international exchange of bibliographic information. Edited by I.C. McJlwaine. London: The Library Association. p.116-122

Kangulu, Christopher Mwelwa. 1994. An evaluation of the Pan-African DevelopmentInformation System (PADIS) with particular reference to Zambia. Doctoral thesis, University of Sheffield, Department of Information Studies.

Koshafian, Setrag. 1993. Object-oriented databases. New york: John Wiley & Sons, inc.

Krasowski, Michael D. 1991. Distribution processing for Distributed Database. Information management V.12. Warrent, Gorham & LaMONT. 4-06-40 : 4-06-85

Lawani, S.M. et.al. 1987. An investigation into the factors affecting the precision of automated title searches. Paper presented at the 25th Annual conference of the Nigerian Library Association (NLA). Abeokuta.

Leung, C.H.C., D.Hibler and N. Mwara. 1992. Picture retrieval by content description. Journal of Information Science. 18(20): 111-119.

Lin, Chang-Yang. 1991. Coverting to a relational environment. In Information management V.12. Warrent, Gorham & LaMONT. 4-06-40

Losee, Robert M. 1994. Upper bounds for Retrieval performance and their use measuring performance and generating optimal boolean queries: can it get any better than this ? Information processing and management 30(2): 193-203

Martin, Thomas H. and Edwin B. Parker. 1978. Designing for user acceptance of An interactive bibliographic search facility. In key papers in the design and evaluation of information systems edited by Donals W.King. New York: knowledge industry publications inc: 388-395.

Meadows, Jack 1992. Purpose and practices of text retrieval. Journal of Information Science. 18(2): 83-87.

Milet, Lynn K. and Francis A. Harvey. 1989. An exploration and analysis of therelationships among object oriented programming, Hypermedia, and Hypertalk. Paper presented at the international Meeting of the Association for the development of computer-based instructional systems. (31st, Washington, DC, November 12-16, 1989).

Muya, Wagura. 1993. information management-Retrieval. In Information Resourcesharing and Networking.2nd ed. edited by Augustes Musana and Lutz Hutemann. Bonn and Arusha: DSE. P. 61-68.

Neelameghan, A. and Rahel Bekele: 1994. Design and development of object oriented database: Application of the principles and postulates of the General Theory of Knowledge Classification. Addis Ababa, SISA, AAU

Object-oriented and object-relational DBMSs. 1996. DBMS. 9(6): 76-78.

Oliver, Martin S. and Sebastiann H. Von Salms. 1994. A taxonomy forsecure object-oriented databases. ACM Transactions on Database Systems, 19(1): 3-46

Ozkarahan, Esten. 1995. Multimedia Document Retrieval. Information & Management. 31(1): 113-131.

PADIS. 1993a. Software packages for processing and exchange of development information in Africa. Paper presented on sub-regional workshop processing tools, norms, standards and modalities for sub-regional exchange of development information for counties covered by the Niamey based Mulpoc. 4-8 October 1993. Niamey, Niger.

__. 1993b. Manual for Document Analysis. UN. Addis Ababa: Ethiopia.

__. 1993C. Report on PADIS activities since the fifth meeting of the regional technical committee for PADIS. Addis Ababa: PADIS.

__. 1994a. Evaluation of Microcomputer Database Software. Addis Ababa, Ethiopia

__. 1994b. Co-ordination and compatibility through the standing committee on the Harmonization of Documentation and information systems in Africa. Sixth meeting of the standing committee on Harmonization and standardization of Documentation and Information systems in Africa., Addis Ababa. 14-18 Nov.

__. 1994c, Report on the authority file. A paper presented on the Sixth Meeting of the Standing Committee on Harmonization and Standardization of Documentation and Information Systems in Africa. 14-18 November 1994. Addis Ababa

__. 1994d. Standardization activities in UNECA/PADIS with special reference to the carnegie project. Sixth Meeting of the Standing Committee on the Harmonization of documentation and IS in Africa sub-committee 1: PADIS Manual for document analysis. Addis Ababa: 14-18 Nov.

___. 1994e. Report on selection and acqisition of microcomputer hardware. Sixth meeting of the and standing committee on Harmonization standardization of Documentation and Information systems in Africa., Addis Ababa. 14-18 Nov.

___. 1996a. Improved access to development information in Africa. Paper presented on the Ninth session of the joint conference of African planners, Statisticians, population and Information specialists. 1-16 March 1996. Addis Ababa, Ethiopia.

___. 1996b. Special initiative on Africa Launched in Addis Ababa. PADIS Newsletter 11(1): 1.

___. 1995. CDS/ISIS aid. PADIS newsletter 10(3): 6.

___. n.d. PADIS Information Resources. Addis Ababa: PADiS-UNECA.

___. n.d. Greater Horn of Africa Electronic Communication network project. Capacity building for electronic communication in Africa UNECA.

Pao, Miranda Lee. 1989. Concepts of Information Retrieval.Englewood, Colorado, Libraries unlimited, inc.

Rashid, Mohamed 1994. Development of some utilities for Micro- CDS/ISIS software. Master's thesis. AAU, SISA

Reid, Glen J. 1996. The relational Vs. OO debate. Computing Canada. 22(22): 50-60.

Robertson, S.E. and M.M.Hancock-Beaulieu 1992. On the evaluation of IR systems. Information Processing and Management. 28(4): 457-466

Rowley, Jennifer. 1993. Computers for libraries 3rd ed. London: LA. publishing

_____. 1992. Organizing knowledge: An introduction to information retrieval. 2nd ed. England: Ashgate Publishing Limited.

Salton, Gerard. 1992. The state of retrieval system evaluation. Information processing and management. 28(4): 441-449

_____. 1978(1972). The 'Gernerality' effect and the retrieval evaluation for large collections. In key papers in the design and evaluation of information systems edited by Donals W.King. New York: knowledge industry publications inc: 168-179

Saracevic, Tefko; Paul Kantor; 46 Alice Y. Chamis and Donna Trivision. 1988. A study of information seeking and retrieving: A background and methodology. JASIS 39(3): 161-176.

Schwachaw, Werner. 1990. Problems in evaluating the quality of information services. In Information Quality: definitions and dimensions edited by Irene Wormell. London: Taylor Graham.

Sharp, John R. 1965. Some fundamental of information retrieval. London: Andre Deutsh

Shaw, W.M. Jr. 1986. On the foundation of evaluation JASIS. 37(5): 346-348.

Seyoum G/Hiwot. 1993. The contributions of PADIS to African development information. unpublished.

Sievert, Mary Ellen C. 1996. Full-text information retrieval. Introduction. JASIS. 47(4): 262-262.

Silva, Mauricio J.V. and C. Robert Calson 1995. MOODD, a method for object- oriented database design. Data & knowledge engineering 17(1995): 159-181.

Simmon, Peter and Alen Hopkinson, eds. 1992. CCF/B: The Common Communication Format for Bibliographic Information and CCF/F: The Common Communication Format for Factual Information. Paris: Unesco.

Simui, Mugageta. 1995. A micro-ISIS based information system for supporting research in the institute for African studies, university of Zambia. Maser's Thesis. AAU, SISA.

Smitten, Alan F. 1994. Natural language processing used in information retrieval tasks. An overview of Achievements to date. In Encyclopaedia of library and information science. Supplement 17. Vol.55 edited by Allen Kent. New York: Marcel Dekket, Inc.: 220-238

Smithson, Steve. 1994. Information retrieval evaluation in practice: a case study approach. Information processing and management. 30(2): 205-221

Soergel, Dagobert. 1976. Is user satisfaction a Hobgoblin?. JASIS. 27 (July- August): 256-259.

_____. 1994. Indexing and Retrieval performance: The logical evidence. JASIS. 45(8): 589-599

Spinello. Richard A. 1991. Full-text Retrieval Systems for document management. In Information management V.12. Warrent, Gorham & LaMONT. 4-06-75

Strzalkolkowski, Tomek. 1995. Natural language information retrieval. Information processing and management. 31(3) 397-417.

Swanson, Rawena Weiss. 1978: 43 Performing Evaluation Studies in Information Science. In key papers in the design and evaluation of information systems edited by Donals W.King. New York: knowledge industry publications inc: 58-74

Townley, Helen M. 1978. Systems Analysis for Information Retrieval. London: Andre Deutsch Lmt.

Unesco. 1989. Mini-micro CDS/ISIS Reference manual (version 2.3). Paris: unesco.

Van Rijsbergen, C.J. 1975. Information retrieval. London: Butterworths.

Yao, Y.Y. 1995. Measuring Retrieval Effectiveness based on user preference of documents. JASIS. 46(2): 133-145.

APPENDIX A

SYSTEM RESTRICTION IN EFFECT IN VERSION 3.07(Rashid 1994:26)

Data Items	**Limitations**
Maximum number of databases	unlimited
Maximum number of records in database	16 million
Maximum record size	8000B
Maximum number of fields (defined in FDT) (excluding repetitions of repeatable fields	200
Maximum number of FST lines	200
Maximum field size	8000B
Maximum number of fields in worksheet page	19
Maximum number of pages in a worksheet	20
Maximum size of display format	8000B*
Maximum number of stop word in stop word file	799
Maximum number of identifiers in an ISIS Pascal program	200
Maximum number of instructions in an ISIS Pascal program (including all programs called in by a uses statements)	10000
Maximum number of loaded programs	10
Maximum number or real constants	200
Maximum Pascal /run-time stack	2000
ISIS Pascal dynamic string area	16932
Worksheet work area size	3KB
Maximum Hit record size	4000*
Maximum number of characters per field	8000B
Maximum number of indexes	1
Maximum number of indexed characters per term	30
Maximum number of literals	132
Maximum number of sort keys	256

maximum number of characters in search expression	250
Maximum number for a field tag	32767
Maximum number of printed characters in a single page	300
Maximum length of pattern type field	20B
Maximum field length in a worksheet	8000B
Maximum length of the name of database, display format, or worksheet	6B
Maximum length of a field name in FDT	30B

Notes: *** Changes in the current MicroIsis software**

APPENDIX B

COMPONENTS OF A BIBLIOGRAPHIC RECORD

- document number
- title
- author
- source reference
- abstracts
- full text
- indexing words or phrases
- citation, or number of references
- organization originating the document, or author's address, or both
- language of the source document
- local information such as special classification numbers, or location

APPENDIX C

ASSESSMENT OF PADIS INFORMATION STORAGE AND RETRIEVAL SYSTEM: The case of PADdev bibliographic database

This questionnaire is prepared to obtain facts and opinion from retrieval service providers of the PADdev of Pan-African Development Information System (PADIS), who are staff members of PADIS. The objective is to attempt to evaluate the performance of the retrieval system, from the professionals point of view.

You are, therefore, kindly requested to assist by providing all facts and opinions that you think may be helpful to evaluate the PADdev retrieval system.

All the information provided by you is used only for the mentioned purpose, thus, please feel free in this regard.

Encircle the letter(s) of your choice or underline Yes or No as the case might be. For the open-ended questions you can use the back side of the page, if any space shortage occurs, but, please associate the continuation with the appropriate question number.

THANK YOU FOR YOUR COOPERATION

I. GENERAL

1. What is your role in the information retrieval service that PADIS is offering on the basis of the PADdev database?

   _____

2. How long have you been on this job? _____

3. Do you have any other experience on information retrieval services?(Yes /No)
   If Yes, what was the retrieval package(s) that you were using? _____

_____

II. INFORMATION RECORDING CAPABILITY

1. Please tick under your choice

FEATURES[1]	VERY GOOD	GOOD	FAIRE	POOR	V. POOR
Screen format	—	—	—	—	—
Editor	—	—	—	—	—
Updating facility	—	—	—	—	—
Coverage (availability of required fields)	—	—	—	—	—
User-friendliness	—	—	—	—	—
Search formulation	—	—	—	—	—
Refining previous search strategy	—	—	—	—	—
Ranking of retrieved documents	—	—	—	—	—
Flexibility of the record structure	—	—	—	—	—
Compatibility with other IR systems	—	—	—	—	—
Presentation of search output	—	—	—	—	—

2. Do you think PADIS methodology is complex to an average data input operator? (Yes / No)

If Yes, what aspect of the methodology may be difficult?

a. there are unnecessary fields that confuse the input operators

b. PADIS purpose codes have unnecessarily complicated the methodology

c. the help facility is not adequate enough to guide

d. others, please specify. _____

[1] Please consider the features in view of the PADdev IR system.

3. Is the error checking mechanism of the system satisfactory? (Yes / No)

If No, what may be the deficiency?

a. the error message is not as explanatory as it should be

b. there are errors that it can not dictate

c. it is unnecessarily detail that it interrupts the inputting process

d. others, please specify. _____

4. Could PADIS methodology solve any information recording problem you encountered as a guideline of an IR system? (Yes / No)

If No, please mention some of the problems that it couldn't solve. _____

_____

5. How do you judge the documentation of the methodology and its online help?

a. very good c. faire e. very poor

b. good d. poor

6. How do you rate the user-interface of the system as compared to the state-of-the-art development that you know?

a. very out-of-date b. very recent c. you have no idea

7. Which of the following interface options do you prefer most?

a. menu selection b. command language

c. WIMP interfaces d. form filing e. question and answer

8. Please note your overall judgement about PADIS methodology and its information recording capability. _____

III. RETRIEVAL PERFORMANCE

1. Is it easy to interpret users' request to a search strategy? (Yes / No)

If No, the reason may be:

a. users' search request is difficulty to understanding

a. users way of expression is much complex than the search capability

available

b. users' terms may, sometimes, not have equivalent terms in the systems dictionary

c. others, please specify. _____

2. How do you rate the search formulating capability of PADdev retrieval system?

a. very limited b. not as expected

c. as flexible as you want d. uncomfortable

e. others, please specify. _____

3. Do you think Cognitive and/or Best-match(natural language supported) search systems can solve the problems mentioned in question number III 1 above? (Yes / No)

4. How often do you complete a search for a request by the first search?

a. always b. sometimes c. rarely d. not at all

5. What is the reason for which you re-conduct the search by refinement of the search strategy mostly?

a. The search result is very small

b. The search output is very large

c. Almost all retrieved documents are irrelevant/dissimilar with request

d. others, please specify._____

6. If you don't complete a search for a request at a shot, how do you refine search strategies? (you can answer more than one)

a. using/changing boolean operators

b. changing the scope of different terms in the search strategy

c. using proximity search strategy

d. specifying search fields

e. using range search method

f. using dictionary search

g. selecting terms from retrieved records

h. others, please specify._____

7. Do you often use previous search strategies for refinement? (Yes / No)

If No, what may be the reason? _____

8. Does conducting the search again and again by search refinement make a significant difference on the number of search results? (Yes / NO).

9. How much time do you spend on searching the PADdev database for a request, on the average?

a. 5 minuets b. 10 minuets c. 15 minuets d. 30 minuets

e. 1 hour and more

Is it acceptable? (Yes /No)

10. Could you get **ALL** the relevant documents available in the PADdev database? (Yes / No)

If No, what do you think the reason could be?

a. the database has not been indexed properly

b. the retrieval software is not effective

c. the records are not well linked

d. others, please specify._____

11. Do you think natural representation of database items' relationships using object-orientation, hypertext search facilities, and full-text search facilities could improve the information storage and retrieval capability of the PADdev IR system? (Yes / No)

12. How often could you retrieve **ONLY** the relevant documents for a request (i.e. without the irrelevant ones)?

a. always b. sometimes c. rarely d. never

If your answer is 'c' or 'd', the reason could be:

a. improper indexing

b. the retrieval software can not discriminate the relevant from the non-relevant

c. inappropriate record linking

d. others, please specify. _____

13. Do you think search languages available at the PADdev retrieval system, like boolean, proximity, etc. are enough to search a database effectively as compared to your knowledge of the state-of-the-art development in the area of search languages? (Yes/No)

If No, what search languages are lacking? _____

14. Do you recommend PADdev database and the retrieval system handling it to be developed to a full-text with a hypertext and/or hypermedia retrieval facilities?(Yes/No)

If No, what other options of improvement do you suggest? _____

15. Is there a need to incorporate a relevant ranking facility (that sorts the search result according to the documents' relevance to the users' request) to the PADdev IR system? (Yes / No)

16. How do you rate the user-friendliness of the retrieval system of the PADdev database?

a. very good b. good c. poor d. very poor

17. How do you evaluate the help facility of PADdev IR system?

a. it is not easily accessible

b. the content is not adequate

c. it is not necessary at all

d. others, please specify. _____

18. Do you recommend PADdev be a distributed/decentralized database? (Yes/No)

19. Is integrating PADdev with the other databases helpful?

(Yes / No)

If Yes, what problems can it solve?

a. searching each databases for different or same request separately will not be required

b. the other databases contribute more in satisfying users information need

c. others, please specify. _____

20. Do you think PADdev's accessibility through the Internet changed its use rate?(Yes/No)

If Yes, do you think the current IR system can accommodate the increasing use load in the future? (Yes / No)

If No, what IR software do you recommend? _____

IV. OUTPUT CAPABILITY

1. Which output format is requested mostly?

a. hard copy (paper form)

b. soft copy (electronic form)

c. microform (cassette or film)

If your answer is 'b' which application format is requested mostly?

a. CDS/ISIS b. MINISIS c. word perfect

d. ASCII e. dBASE f. FoxBASE

g. INMAGIC database formats

h. others, please specify. _____

2. Do you download/save search results without any problem, as requested? (Yes / No)

If No, what problems have you encountered? _____

3. Has there been any output format compliant you have received from users? (Yes / No)

If Yes, what is the general feature of the complaints? _____

V. COMMENTARY QUESTIONS

1. Do you think PADdev is being well utilized? (Yes / No)

   If No, What do you think is the reason for less utilization of the database? _____

_____

2. PADIS methodology is not well adopted by National Coordinating Centres. What do you think the reason may be? _____

3. If there would have been no financial constraints, would you still prefer ISIS packages to handle PADdev? (Yes / No)

   If No, what IR packages/approaches would you prefer?_____

4. Do you think the would be coming ISIS for windows will have lots of problems to solve? (Yes / No)

   If Yes, please point out some of the basic problems you expect to be solved by ISIS for windows. _____

5. Do you think PADdev IR system is up to the expected technological development? (Yes / No)
6. Do you think the CD-ROM production of PADdev a problem solving action?(Yes/No)

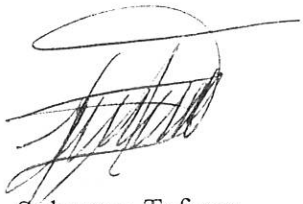
   If Yes, what problems can it solve? _____

7. Please write any other additional comments, opinion, and facts about PADdev IR system. _____

_____

Declaration

THIS THESIS IS MY ORIGINAL WORK AND HAS NOT BEEN SUBMITTED FOR A DEGREE IN ANY OTHER UNIVERSITY



Solomon Teferra
May 16, 1997

THE THESIS HAS BEEN SUBMITTED FOR EXAMINATION WITH MY APPROVAL AS UNIVERSITY ADVISOR

GBADE A. ALABI