ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

SCHOOL OF INFORMATION STUDIES FOR AFRICA

COMPUTER-ASSISTED GENERATION OF VOCABULARY CONTROL TOOLS AND THEIR APPLICATION TO DATABASES IN URBAN PLANNING

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

BY

TESHOME BULA OLCHA

JUNE 1993

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

SCHOOL OF INFORMATION STUDIES FOR AFRICA

COMPUTER-ASSISTED GENERATION OF VOCABULARY CONTROL TOOLS AND THEIR APPLICATION TO DATABASES IN URBAN PLANNING

BY

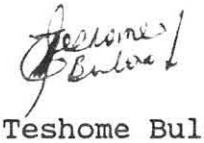
TESHOME BULA OLCHA

Approval by the Board of Examiners:

Getachew Birru	Getachew Birru
Chairman, School Graduate Committee	
A. Neelameghan	
Advisor	
G. Bhattacharyya	G Bhattacharyya
Advisor	
TAYE TADESSE	
Examiner	
S Seetharama	S Seetharama
Examiner	
Examiner	

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.



Teshome Bula Olcha

31 May, 1993

The thesis has been submitted for examination with our approval as university advisors.

A. Neelameghan

Advisor

31 May, 1993

G. Bhattacharyya

Advisor

31 May, 1993

ACKNOWLEDGEMENTS

This work would not have been completed had it not been for the unreserved assistance of many people. It is not possible to mention all but a few of them here.

I would like to express my gratitude and appreciation to Prof. A. Neelameghan for his guidance, advise, encouragement and help at all times. Many thanks are also due to my Co-advisor Dr. G. Bhattacharyya for his appreciable assistance.

I extend my heartfelt thanks to Alemayehu Waktole and Dafa Gudina who helped me in data collection, organization and encoding. I am also indebted to Ato Bekele Teklu for his help in abstracting some documents.

I am grateful to Ato Getachew Birru, Dean, SISA, Dr. Taye Tadesse, Ato Tesfaye Biru, Ato Solomon Yewondwossen of the SISA teaching staff and to the SISA office staff for their support throughout my studies in SISA.

My sincere gratitude goes to the National Urban Planning Institute (NUPI), Addis Ababa, for its financial support during my studies at SISA.

I highly appreciate the cooperation I was accorded by the staff of the National Urban Planning Institute (NUPI) in providing me information/data whenever I needed it. Especially, my deep gratitude to Ato Mekete Shiferaw and Ato Megbaru Yimer, for allowing me to use the computer facilities of NUPI, and Ato Dereje Zewdu and Ato Kassahun Tesfaye for their material assistance.

I owe much to Ato Ayalew Tessema, Ato Kumsa Burayou and Ato Fromsa Taye for their moral support.

I extend my deepest gratitude to Adde Qonjit Bogale for her moral support and encouragement that gave me strength when I felt weak.

And, all my friends and relatives deserve my appreciation for their kindness and sympathy during my study at SISA.

ABSTRACT

This thesis discusses and presents the different methods of computer assisted generation of vocabulary control tools, particularly thesaurus and it variations, as well as their application in designing and developing databases and user interfaces to support urban planning activities. The various features and attributes of on-line databases and search services are examined with a view to assessing their relevance to the information needs of specialized institutions, such as the National Urban Planning Institute (NUPI). The need for local databases to provide the type of information needed for specialized local/sectoral institutions and the role of vocabulary control tools in facilitating the design and development of such databases is discussed. The various types of information needed for preparing urban plans, their sources, and the need for an effective information support systems are discussed. Different types of vocabulary control tools, their structure and display, and procedures for designing and developing them are mentioned. Also discussed are the features of selected softwares and programs useful in the design, development and maintenance of thesauri. Particular attention is given to CDS/ISIS Pascal thesaurus programs (Thes, Mthes, Thes1 and Thes2) used in the current work. A classaurus for the field of urban planning has been developed using Mthes pascal program. The terms were collected from specialized documents in urban planning and queries collected from potential end-users of the system at NUPI. The classaurus was used to update the OECD Macrothesaurus

to enhance its utility in the urban planning field. Direct thesaurus generation, with the assistance of computer, from texts and their abstracts without using a language parser is demonstrated. This method of thesaurus generation is found to be effective and time saving. Prototype urban planning databases made up of reference and specialized object-oriented databases (OODBs) have been developed. The integration of the databases was made with the help of ABNCD+ database structure. The integrated database allows search in the different records of the database with a single query. Two user interface facilities are provided: (1) A thesaurus-like search facility to facilitate browsing and navigating through the subject areas into which the records in the OODB have been divided. They can also be applied to the reference records with modifications and editing of the search expressions provided by Mthes.pas; and (2) A structured query worksheet based on the structured divisions of urban planning information. In this context, vocabulary control tools have more potential applications than their usual applications; for example, in database design, user interface, etc. Advances in information technology and development of different thesaurus software have renewed the interest in the application of vocabulary control tools, particularly thesaurus and its variations. This study also presents examples of thesaurus generation, retrieval of records from the integrated database, the use of the user interface facility, the applications of queries to the data bases and flexible outputs from the OODB.

DEDICATION

This work is dedicated to the following persons:

* My deceased mother Warqitu Beka
* My deceased sister Jorgee Bula
* My real friend Qonjit Bogale (Bareedduu), and
* ABO, in whose existence my hope lies.

@ ********** @

TABLE OF CONTENTS

ANNEXES

Annex

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1. OBJECTIVES OF THE STUDY

1.1.1. General Objective

The overall objectives of the study are to examine:

(1) the different methods of computer-assisted generation, development and maintenance of vocabulary control tools, particularly the thesaurus and its variations; and

(2) the use of vocabulary control tools, especially the thesaurus and its variations, in the design and development of and retrieval from, databases to support urban planning activities.

1.1.2 Specific Objectives

The specific objectives of the study, formulated in conformity to the overall objectives mentioned above, include the following:

(a) to develop database(s) covering urban planning subject fields and containing a mix of bibliographic records, profiles of experts, of institutions, of research projects, as well as object-oriented tabulated data, using Micro-CDS/ISIS;

(b) to use these and other records to generate sample thesaurus and classaurus in the conventional way and by using Micro-CDS/ISIS and related Pascal programs;

(c) to apply the knowledge organization principles and the thesaurus and classaurus generated [see (b)] for the updating of the OECD Macrothesaurus which covers the urban planning field; for providing user-interface(s) in formulating search expressions for information retrieval from the database(s) mentioned at (a); and

(d) to examine the features of thesaurus generating software especially those that will be used in this study.

1.2 SCOPE OF THE STUDY

The subject coverage of the prototype databases and the associated vocabulary control tools will be the field of

"urban planning", more particularly to meet the needs of the National Urban Planning Institute (NUPI).

The databases will be developed using the software Micro-CDS/ISIS and programs written in CDS/ISIS Pascal language. The following Pascal programs will be used in this study: Thes.pas; Mthes.pas; and Thes1.pas; and Thes2.pas. Thes.pas developed by Unesco, has been made available with Micro-CDS/ISIS software; Mthes.pas, a modification of Thes.pas, was obtained from the Documentation Centre of the National Council for Research, Khartoum, Sudan; and Thes1 and Thes2 were made available along with Macrothesaurus (on diskettes) by OECD, Paris. These software packages and programs were used in developing the system components mentioned under 1.1.2.

1.3 LIMITATION OF THE STUDY

Time constraints and difficulties in accessing a large number of end-users and information generators in the various organizations in Ethiopia concerned with urban planning aspects, and responses to the surveys carried out in connection with this study, were limited even after persistent efforts. Nevertheless, the information

collected is representative of the needs of urban planning.

1.4 SIGNIFICANCE OF THE STUDY

The field of urban planning has been widening in scope, and it has been all along a multi-disciplinary subject. It requires a whole range of data and information on several physical and socio-economic aspects for preparing a plan that reflects the reality, and meant to be implemented. For such purposes, organized, comprehensive, pertinent, reliable, and timely information should be made available. Databases that contain information pertinent to the needs of urban planners can be helpful in providing such information service.

Object-oriented databases containing organized information in readily usable form can be of particular help to planners as they save search-time and effort involved in the identification and compilation of information from various sources. A user interface mechanism that can assist users in viewing their respective fields of interest in a helpfully structured way and in formulating specific search expressions facilitating retrieval of pertinent information will enhance the usefulness of information systems and users'

appreciation of such systems.

The development of local end-user-oriented databases and an appropriate user interface system, with the help of thesaurus techniques are recent developments. Hence the significance of this study for the urban planning field.

1.5 ORGANIZATION OF THE THESIS

The thesis is divided into nine chapters. The first chapter is introductory, in which the general and specific objectives, the limitations, the scope, the significance of the study are discussed. The second chapter deals with factors that form the basis, background and justification of the study. In the first part of this chapter, the various attributes and features of on-line databases and search services as well as their applications are examined with a view to assessing their relevance to information needs of local and specialized institutions like NUPI. Various search procedures and problems, and their relevance to local databases are discussed in the second part. In the last part of chapter two, the justification for the study in terms of relevance of the on-line procedures to local databases and the special needs of NUPI, is presented.

Chapter three presents the methodology adopted in the study. In chapter four, urban planning as a subject matter, the type of information required for urban planning, the pertinent sources of information, and the problems encountered in getting the information from the source are dealt with. In this connection, the need for an effective information system for urban planning is discussed.

The various vocabulary control tools, and the procedures and requirements for designing the tools are presented in chapter five. It covers the common vocabulary control tools, i.e. thesauri and schemes of classification. Chapter six is on thesaurus softwares. In this chapter, some thesaurus programs are discussed with respect to their capability and application in thesaurus development. Particular attention is given to their relevance to and use in the current work.

In chapter seven the concept, the types and attributes of databases are discussed. Specialized object-oriented database design and development as well as the underlying principles thereof are given particular attention.

The main concern and the substantive part of the current work, the design and development of the system

components, is included in chapter eight. In the first part of the chapter, the queries collected from potential end-users of the system, are summarized. The second part is on the generation of the vocabulary control tools (thesaurus system) in which the procedures for the generation of the thesaurus and the actual work done are considered. The third part deals with the design and development of integrated prototype urban planning databases. In the final part of chapter eight, the design and development of user-interface facility involving the design of a thesaurus-like search assistance facility and a structured query for the field of urban planning are dealt with.

The major conclusions drawn from the study and the recommendations are summarized in chapter nine.

CHAPTER TWO

ON-LINE DATABASES AND SEARCH SERVICES STATE-OF-THE ART AND THEIR RELEVANCE TO NUPI's NEEDS

2.1 GENERAL CONSIDERATIONS

In this section, the various attributes and features of on-line databases and search services[1], and their application are discussed with a view to assessing their relevance to information needs of local and specialized institutions in general and NUPI in particular. Selected features of on-line databases and search services and retrieval of information from them are presented as they are experienced in practice. Their relevance and use to satisfy NUPI's information needs are examined and an alternative approach is suggested.

2.1.1 On-line Databases and Search Services

In an on-line search, the searcher sits at a terminal or work station in a library, information centre, office, laboratory or even at home and connects the terminal through a local telephone, acoustic coupler or modem to a computer in another location.

---

[1] The presentation is based on the information in part 1 of the "Manual of On-line database Searching and Retrieval" prepared in 3 parts by Prof. A. Neelameghan.

CHAPTER TWO

ON-LINE DATABASES AND SEARCH SERVICES STATE-OF-THE ART AND THEIR RELEVANCE TO NUPI's NEEDS

2.1 GENERAL CONSIDERATIONS

In this section, the various attributes and features of on-line databases and search services[1], and their application are discussed with a view to assessing their relevance to information needs of local and specialized institutions in general and NUPI in particular. Selected features of on-line databases and search services and retrieval of information from them are presented as they are experienced in practice. Their relevance and use to satisfy NUPI's information needs are examined and an alternative approach is suggested.

2.1.1 On-line Databases and Search Services

In an on-line search, the searcher sits at a terminal or work station in a library, information centre, office, laboratory or even at home and connects the terminal through a local telephone, acoustic coupler or modem to a computer in another location.

---

[1] The presentation is based on the information in part 1 of the "Manual of On-line database Searching and Retrieval" prepared in 3 parts by Prof. A. Neelameghan.

On-line searches may be in a machine readable database created by an institution and accessible on a computer located in the same building or in some other location, or it may involve the use of public data networks and international telecommunications networks such as TYMNET, DATAPAK, etc. for accessing external on-line search services, that is, databases remotely located, even at global distances.

A user of an on-line database service may be an information intermediary attempting to find references or information in response to an end-user's query or it may be the end-user directly attempting on-line search.

There are a large number of commercially supported on-line search services offering a large number of publicly available databases containing millions of records, produced by database producers.

Each on-line service offers a number of databases, mounted on large mainframe computers and can be searched concurrently by a large number of searchers via remote terminals located in many parts of the world. On-line database search services are dependent on appropriate search equipment, communication systems, and availability of databases.

On-line search databases can be grouped on the basis of different attributes, such as, subject scope, geographic and chronological coverage, periodicity of release of new information by the database producer, frequency of updating, and the type of information/records they contain. A useful classification of databases is as follows (Cuadra, 1988):

2.1.1.1 Reference Databases

These databases refer or point the user to another source, such as, document, organization, individuals (experts), projects, etc. for additional information or the full text. These databases include bibliographic databases which contain citations and often with abstracts of the printed documents, such as articles in periodicals, conference papers, patents, reports, theses, newspaper items, books, etc.; and referral databases which contain reference to or addresses and descriptions of institutions, persons, audiovisual materials, and non-print media for further information.

2.1.1.2 Source Databases

These are databases that contain original source data, or full text of original information source,

or materials prepared specifically for electronic distribution. Source databases include:

- Numeric databases containing numeric information such as, original survey data, scientific/technical data (measurement), and statistically manipulated presentation of data (economic time series, production statistics of manufacturer, etc);
- Full-text databases containing records of the complete text of an item, such as, a full newspaper item, a technical specification, or a court decision, or a newsletter; and,
- Software databases containing computer programs that can be downloaded for use on a local computer.

2.1.2 Database Producers

Databases, some large ones but mostly relatively small ones, are produced by a variety of organizations: academic, commercial, national, international, etc.

In the 1960s, the requirement of the academic community had stimulated database suppliers to produce bibliographic and other files in the major subject

fields, such as, the Physical and Biological sciences, Psychology, Education, Economics, Law and other Social Sciences. Until recently, the Humanities received much less attention. The scope of the major databases was dictated by the market for the corresponding printed products, that is, the abstracting and indexing periodicals. In the 1970s specialized databases to serve the needs of business and industry began to appear, PREDICAST and PERA are examples of such databases.

Libraries and information centres of institutions may have to acquire on-line database services to suit specific needs of their respective clientele. In on-line, as users are located remotely from search services, they have no control over the scope, depth, quality, etc. of the information in the database.

Users of on-line search services are usually interested only in a small proportion of the documents cited in a given database. However, in order to ensure comprehensiveness of retrieval, users need to search not only on an entire database but also several databases at times. In many on-line services, there may be separate files for different types of material which increase the need to search multiple files so as to ensure comprehensive coverage for a given query.

2.1.3 Selection of Databases and Search Services

An important task in pre-search planning is a selection of appropriate database(s) and the search service on which the database(s) is (are) available. It is difficult to use on-line search services optimally and competently because of:

- the large number of databases available,
- the overlap and duplication among the databases offered by different search services,
- the variability among search languages,
- the variability within query languages when different version of the same language are available at the same time, and
- the variability between databases and indexing languages including different versions of the same language.

In order to meet diverse user-needs, the range of options provided by the on-line services has been expanding and it is becoming increasingly more difficult for the user to maintain a level of expertise continuously.

Characteristics of databases differ from one to the other. A variety of retrieval software and conversions may be necessary to interact effectively with and obtain satisfactory result from the systems. The systems are not sufficiently user-friendly, and vary with respect to

query language, protocols, system response, etc. These depend on the characteristics of the database and file structure.

A searcher can control certain aspects associated with the formulation of query at the computer terminal but has no control over the search language, interrogation procedure, quality of searchable data file and the range of failures attributable to the indexing language and the indexing process.

On-line retrieval is subject to the combined effects of whether a database covering the desired subject(s) is available on a particular search system, and whether the retrieval capabilities of the system meets the requirements of the end-user's query. The situation becomes more complex when, in order to ensure full coverage, etc., more than one database which use different search languages need to be searched.

During the past two decades (1970s and 1980s), sophisticated user interfaces including expert systems and browse capabilities to assist the searcher in navigating databases have been developed.

2.1.4 Criteria for Selection of Databases

There are databases which cover a very specific subject area (e.g. ENERGYLINE); there are some which are multi-disciplinary (e.g. NTIS, PASCAL); there are mission oriented (e.g. NASA) and other problem-oriented (e.g. TOXLINE) databases. It was mentioned above that there are different types of databases. There is obviously overlap among those databases in respect of coverage of subject areas. Therefore, users or searchers should be provided with the necessary information on characteristics of the databases in order to make the appropriate selection of the databases of interest to them.

The criteria for selecting databases include (Neelameghan, 1992):

Comprehensiveness: Comprehensiveness of databases may include considerations such as:

- to what extent the subject is complete with respect to the end-user's needs;
- the type of documents (periodical, reports, conference proceedings, grey literature, thesis, patents, standards, etc) are included;
- whether the principal periodicals, the most cited ones, in the field are indexed;
- whether all the documents entered in the

database indexed in-depth or some are indexed at a superficial level;

- whether the materials from particular regions of the world comprehensively covered;
- the availability of material selection by a language of interest;
- the largeness of the database. This information can be of some use in deciding which database to select if more than one database covers the same subject area; and,
- how far back in time the database begins on different on-line services. This information is useful for retrospective search purposes.

Currency: Currency is related to the frequency of updating the database. A majority are updated monthly, a few more frequently, and others quarterly or even less frequently. It is also useful to know about the time lapse between the publication, say, of articles in a periodical or a report or a proceedings of a conference and their inclusion in a database.

Data Elements: Most bibliographical databases include all the usual data elements in a unit record proposed in cataloguing rules. Some, however, include more than the others. Information on whether an abstract is provided and whether a data element, such as a descriptor used as a major concept, a company or an institution name in a business database can be directly accessed on-line, is also essential in database selection.

Indexing: Indexing features relevant to on-line database searching include:

- how comprehensive the indexing is; i.e., whether all materials entered in the database are indexed to the same level;
- the exhaustiveness (indepthness) of the indexing;
- whether the database uses a controlled vocabulary for indexing or whether it uses natural language or uncontrolled vocabulary;
- whether the database which use a vocabulary control tool also provides for search using natural language words; for example, the title, abstract, etc. If so, whether the natural language terms are placed in a separate field; and,
- whether it is possible to restrict search by type of document, or by language, or by date of publication, etc.

In a printed index, subject indexing and the degree of coordination introduced in the subject headings are kept to a minimum with maximum collocation of subjects so that the index is of a manageable size and scanning a printed page is facilitated.

Cost: The method of charging for services rendered is complex and may vary from one on-line service to another. For a given query, unless there is a substantial difference in the amount and type of information retrieved from one database rather than another, one

would select the service/database that is less expensive.

Others: In addition to the above criteria of database selection, local availability of primary sources, error rates, printing and sorting, documentation and aids are also important factors to consider in selecting databases that meet the specific needs of end-users or searchers.

2.1.5 Selection of Service Suppliers

The number of on-line services has increased during the past two decades (1970s and 1980s). The EURONET service alone has some forty-five service suppliers. When a database is unique to a service, the question of selection of service does not arise. But many databases are accessible on two or more services. In such a case, some of the criteria mentioned under selection of databases, in addition to other factors, are relevant in the selection from among the search services as well. Some of the criteria for selecting a search service are:

- coverage in terms of full or only part of a database,
- for how many years in retrospect it has been giving service,
- data elements included in a record. For the same document one service may include abstract, while another may not; similarly in respect of fields

indexed,

- searchability of data elements in a record; on one service more data elements may be searchable directly than on another,
- display/print out format available; i.e. whether the searcher can select fields for display,
- search facilities provided, such as, adjacency or proximity searching, string searching, left truncation, use of field qualifier, limiting search to a language, etc.,
- provision for on-line document ordering,
- cost/charging factors; for instance, cost of the same database varies on different services, relative communications cost to host computer, whether royalty charges on the database are included in the connect charges or billed separately; print charges per reference on-line or off-line, discount available from supplier by contracting to purchase a predetermined number of connect-hours a year on databases expected to be used extensively by the organization, etc., and
- searcher's familiarity with the system and response time are also important.

2.1.6 Local Databases

The overview of on-line databases, particularly those accessed from remote locations presented above has some implications to the present study: whether those databases and search services are relevant to meet the needs of local institutions in a developing country, such as NUPI in Ethiopia.

Despite the large number of databases of different types being available on commercial on-line services, developing countries and local institutions such as NUPI have to develop local databases of various types to meet the information and data need of their respective personnel (planners, decision makers, experts, executives, etc.).

Institutions like NUPI, which are involved in planning activity need information from different disciplines, locations (sources), for different planning purposes, etc. The sources of these data/information may be both local, national, regional, and international. In addition to the normal types of information usually available, such information as on institutional profiles, project profiles, expert profiles as well as data on specific objects (e.g. a town) is needed. Much of the data and information are locally generated, and usually not covered/indexed in international commercial databases. The data are of local interest only. There is also need to have all these different information and data in one integrated database so as to enable end-users to search and obtain the specific type of information they need from different databases and records with minimum effort. Such databases are essential and need to be designed,

developed and maintained locally to support the specific activities of the institution, in the case of NUPI, urban planning and plan preparation. Such information and data are of vital importance to planning, decision making, problem solving, research and other activities related to the socio-economic and technological development of the country.

On-line access to remote databases requires appropriate equipment, telecommunications facilities, etc. which most of the African countries do not have. And on-line commercial services are expensive for most of the institutions in Africa. This is most serious when it comes to institutions which need information/data on different subject fields necessitating the use of different databases. In brief then, NUPI should design, develop and put into use local databases which cater to the data and information needs of its clientle. As mentioned earlier, users have no control over a number of features of commercially available on-line databases.

On the other hand, the data needed by the planners, decision makers, and other professional experts in institutions such as NUPI need to be analyzed, synthesized, consolidated and presented in forms and formats convenient to them. This requires the design of

systems that provide value added information services that are user-oriented. Thus, all these factors have necessitated the development of local/institutional information systems and services in NUPI. The present work examines the application of vocabulary control tools, especially thesaurus, generated from the information sources used in NUPI, with a view to improving the quality of data and information provision for urban planning enhancing the relevance of other vocabulary control tools and, using as the basis of designing such tools in developing a user-system interface facility.

2.2 ON-LINE SEARCH PROCEDURES AND PROBLEMS

The main objective of this section is to assess the various search procedures and problems with a view to examining their relevance to local databases of specialized institutions. Some of the known procedures and problems of on-line searching are discussed and then their relavance to local databases are considered.

2.2.1 Search Language

An idea or a concept is normally expressed in a natural language for the purpose of communication in written or verbal form. A subject may be expressed by a single term or, more often, by a combination of terms representing two or more ideas. The combination of terms need to be structured according to the grammar of the natural language used for effectiveness of communication of the ideas to a native speaker of the language. In a computerized search system, the component terms representing a subject may be held in binary form. However, a user of the system, i.e. an end-user or an intermediary, seeking information about a subject, presents the query in a natural language known to him/her.

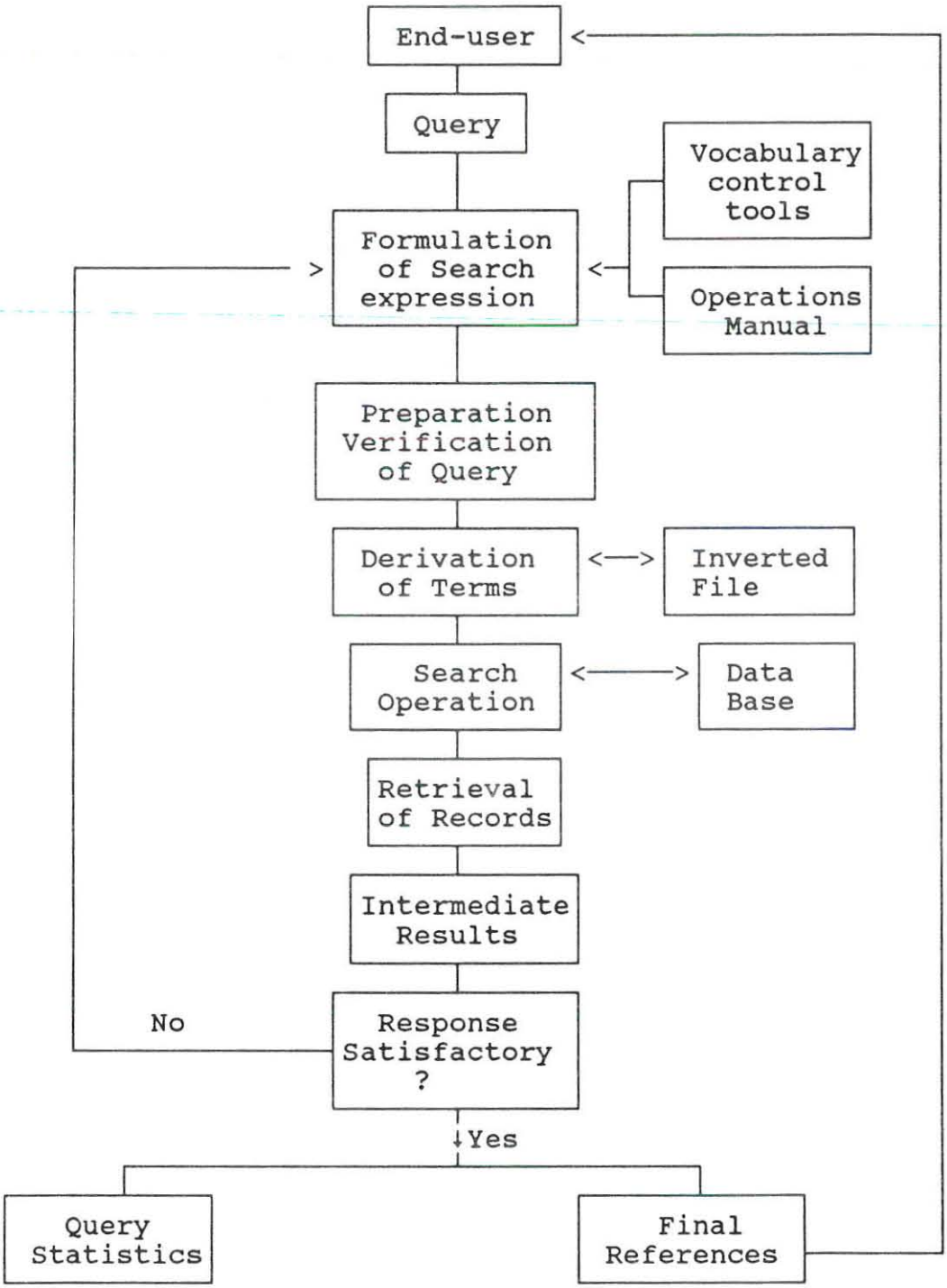
In attempting to search information from on-line databases, the user needs to understand the search languages of the database service so as to be able to express the information need in the system's query language in a way that it represents precisely his/her information need at the moment. This requires the organization of the relevant concepts or subject ideas (representing the user's information need) in the tools e.g. index, vocabulary control tools, etc., used in the retrieval of information in response to users query.

4. If possible and necessary. discussion with end-user with a view to specifying as precisely as possible the various aspects of the end-user's information need. Aids such as scheme of classification, thesaurus, known documents covering the subject area of the query, etc. can help in the interview with the end-user.
5. Selection of database(s) (and on-line service(s) if applicable) likely to yield best results.
6. Formulation of the query in the search language of each of the databases (and on-line service selected): adoption of search strategy and search expressions appropriate to the structure, organization, search language and capabilities of the information sources and services selected. Vocabulary control tools, such as thesaurus, scheme of classification, etc. associated with the database(s) to be searched can assist in this step. Participation of the end-user is important at this stage to obtain satisfactory search result.
7. The search operation: fast access files, such as index/inverted files, created by the system should be used. The end-user can assist in evaluating the intermediate search results.
8. If necessary, modification of search strategy and refinement of the search expression on the basis of the successive intermediate results obtained and their evaluation by end user.
9. Selecting the most relevant references and provision of copies of the documents as desired by end-user.
10. Recording the query(for example, in a Search Request Form), the search procedure adopted, if necessary, and the result obtained for future use and statistics of services rendered.

2.2.3 General Features of Search Languages

In a database a document is usually represented by surrogates, e.g. a catalogue entry/record. Index terms often selected from some kind of vocabulary control tool are assigned to the documents by a trained indexer or subject specialist(s). Alternatively, the term may be selected by the indexer perusing the title, abstract, summary or even the whole text of the document indexed. No vocabulary control tool may be used. These index terms and/or the related vocabulary control tools can be used for selecting terms in formulating the search expression. The terms of the search expression are matched with the index terms assigned to documents (surrogates) to effect retrieval of relevant records in response to the search made with the terms. The retrieved references are expected to represent documents pertinent to the end-user's query and those not retrieved are deemed not pertinent. Fig. 2.1 is a schematic representation of the search procedure.

Fig. 2.1: Schematic representation of search procedure and the role of Vocabulary control tools in the process



On-line services generally permit the formulation of search expressions using Boolean operators AND, OR, NOT; some services permit the use of other operators: greater than, less than, equal to, adjacency or proximity, Parenthesis, truncation, and partial string matching, as well. Also facilities such as nesting, back referencing, field qualification, and expanding a particular term are available in some systems for building logical set of documents to modify query progressively on the basis of intermediate search results obtained.

2.2.4 Devices to Improve Search Strategy

There are two devices that may be used to improve the search strategy and the result thereof; namely, recall and precision devices.

2.2.4.1 Recall Devices

Recall devices may be used to increase the number of relevant references in the retrieved set. There are basically seven types of recall devices and each can be implemented in a number of ways in information retrieval system. A summary of the devices follows:

Synonym Control

If the software used has a facility for mapping of terms, then the searcher may use a non-preferred term from the vocabulary control tools associated with the database. In some systems, such as MEDLINE, it may not be expedient to use concurrently two search aids, e.g., expanding (exploding) an access term and mapping facility. This results in a "NO POSTING" message. Terms that are not truely synonymous are treated as such (e.g. Flat and House), it may result in loss of term discrimination in the searcgh language.

In some systems synonym control is implemented by confounding synonymous and near-synonymous terms, using different terms in indexing but they may be taken as synonymous in searching. Judicious use of truncation can help; e.g., Plan? can take care of plan, plans, planner, planners, planning, etc.

Hierarchical Term Linkages

This refers to the linking of the selected descriptor to its broader and narrower terms. Establishing such relationships should be done in preparing the search expression, for example, using appropriate vocabulary control tools. Expand/explode facility of the search software automatically match generically related search

terms taken from the controlled vocabulary. This can at times lead to the inclusion of undesirable terms in search expression.

Associative Term linkage

These are similar to the related term linkages for a given descriptor in a thesaurus, and a variety of such relationships have been recognized.

Document Clusteruing

Associative and probablistic indexing apply statistical techniques to produce clusters of keywords that form the basis for automatic derivative classification for retrieval, that is, the associative groups are used in cluster-based retrieval (Van Rijsbergen, 1979). The searcher controls the search, decides on the stopping point, and produces the list of retrieved documents in a sequence ranked according to their closeness to the query.

Control of Word Forms

This is based on the prescriptive use of vocabulary in indexing. Judicious use of truncation enables matching of words with common stem. For example: Plan? will match with Plan, Plans, Planner, Planners, Planning, Planned; Urban? will match with Urban, Urbanized, Urbanization, Urbanizing, etc.

Control of word forms at search point provides the searcher different search options. However, the flexibility provided by truncation will be lost for confounding of wordforms is done automatically at the indexing stage.

Summation of Document Sets

Almost all on-line search services permit the use of Boolean operators. Different word forms can be ORed together. For example:

Urban plan OR urban policy OR Urban design

Gerry points out that:

> Under certain circumstances, the logical expression A AND B may be too constraining, in that not all wanted documents explicitly contain B. Recall can be improved by identifying alternatives for B or by simply omiting B from the search and identifying some undesirable concept C... (A AND B) OR (A NOT C), which will retrieve those documents that contain A and explicitly mentioning B while rejecting those explicitly mentioning C.
>
> ... ...
>
> Quorum logic, in which retrieval is based on "n out of x terms," is also equivalent to Boolean logic and is useful when retrieval of a polythetic group is required but the actual term combinations are unimportant. Both weighted retrieval and quorum logic have the advantage of being able to express a retrieval function in a concise way when the retrieval objective is clear.

2.2.4.2 Precision Devices

Precision devices may be used to restrict the number of records retrieved. Such devices enable the co-ordination or linking of information at various stages of document indexing, search and retrieval processes. These include logic, syntactic, weighting devices and bibliographic coupling. Each of the devices may be implemented in more than one way on on-line retrieval systems. A summary of the devices is presented in the table below.

Precision Devices (Gerry 1983: 152)

Devices	Implementation	Remark
Logic devices	a) Boolean AND and NOT	Reformulation of a search to increase precision is easier with Quorum logic
	b) Quorum logic x out of n	
	c) Contextual logic	
Syntactic devices	a) Links	
	b) Roles	
Weighting	a) At the point of indexing	Easier to vary the specificity of a search by assigning threshold and cut-off values. Weighting at the point of search provides more flexibility including the retrieval of a continuous set of documents in a sequence of their probable relevance
	b) At the point of search	
	c) At the point of retrieval	
	d) At the poin when documents enter the system	
Bibliographic coupling	Measure the associations between documents according totheir co-citation frequency	

Logic Devices

These devices include Boolean operators, e.g. AND, AND NOT as distinct from OR NOT, and appropriately applied quorum logic (a retrieval specification of "6 out of 9" will retrieve fewer records with higher probability of relevance than a specification of "4 out of 9). Logic devices are based on postcoordinate contextual analysis, an extension of Boolean logic, and takes into account positional information in the retrieval process for enhancing the associative strength between terms thus improving the precision.

Sysntactic Devices

Links and roles are used with syntactic devices. Links show which concepts are related by assigning a common letter or a number to each of them or by using fixed tags and subfield code combinations so as to avoid false coordination of terms. As there are other methods of avoiding false coordination there is relatively less interest in the use of links and roles.

Weighting Devices

With a view to presenting search output in a ranked sequence (e.g. better than with quorum logic) term weighting technique is used. Boolean logic does not provide for the presentation of retrieved records

according to the relative frequency to the query. Weighting applied at the stage of indexing a record is based on the intuitive judgement of a human indexer identifying concepts as major of minor or based on a statistical assessment of a term's weight depending on a calculated frequency of occurence of terms. Retrieval based on such weights regard weighted occurence of terms as more specific subsets of the whole set, both weighted and unweighted.

Weighting at the stage of searching permits the searcher to express an interest in a term over another for reasons unrelated to the actual use of the terms in the set of documents. A user interested in the "comparative presentation of land use pattern in Nekempte and Akaki towns" may give greater weightage to a document that presents land use patterns for the two towns than to separate documents that deal with land use pattern in Nekempte and land use pattern in Akaki.

Bibliogrphic Coupling

Bibliogrphic coupling is a measure of concepts/subjects association determined by the number of cited works that documents have in common. Bibliogrphic association may be used as a device to increase precision in retrieval in a scoring search.

The search begins with retrieval of a document known to be relevant to the query. Documents in a file/database in which the cited documents do not match in any way those in document chosen are rejected. Documents with matching cited items are presented in ranked sequence of their decreasing commonness of items with those in the bibliography of the chosen document.

On-line search systems allow the display of retrieved records on the computer console or printed off-line according to predefined format(s) or, in some cases, according to user defined format(s).

2.2.5 Limitations of Boolean Logic

As said above, most on-line information retrieval services use Boolean operators which operate on inverted file of the database.

The Boolean mechanism, however, has limitations which should be recognized (Gerrie, 1983:45)

1. A query must be fully and precisely stated as a logical expression compatible with the indexing practice adopted by the database, irrespective of whether the index is made up of derived or assigned

terms. This means an incomplete or open ended query is not permitted.

2. The Boolean retrieval mechanism tends to have an all-or-nothing quality, that is, if the logical search expression is true for a specific document, the corresponding record is retrieved; otherwise, it is not retrieved. In other words, there can be no partial matching.
3. The logic of sets forces the value-judgement regarding a document into a binary scale- either relevant or irrelevant.
4. There is no order of preference within sets of retrieved records; that is, each document retrieved is taken to be as important as any other. And although a query can be precisely expressed using Boolean logic, the set of retrieved documents is not arranged in any sequence in terms of relevance to the query. A single retrieved set can be ranked after retrieval.
5. Boolean logic makes it difficult to vary the depth of a search in order to vary the number of records or quantity of information required unless terms either are grouped in classes so that they can be automatically substituted for one another or the searcher is willing to study the properties of the index language in order to construct the right kind of query.

6. The calculus of sets (Boolean algebra) is not suited for considering doubtful classes of variable product class. String A B C may represent a different class (B A C) in a different context.

Application of faceted vocabulary control tools, knowledge classification based thesaural structures, etc. to overcome some of the limitations of Boolean logic in information retrieval are discussed by Ingwersen and Wormell (1992).

2.2.6 Mismatch Between Information Need and Search Results

The system of information retrieval from databases has two parallel analysis, i.e. the analysis of a document which is the source of the information in the databases, and the analysis of the query of an information user which is the basis for information search in the databases.

The elements of these analysis are given by Gerrie (1983: 6). The elements of document analysis include: selection of a document, conceptual analysis of a document, organization of document descriptions or citations in a file [database], and storage of documents. The elements

of query analysis include: conceptual analysis of queries, search formulation, retrieval of document descriptions through matching of the search formula and the document descriptions and, if necessary, retrieval of documents. In this activity, the document which is the source of information is described by surrogates or catalog entry that include subject terms while the query is expressed in terms of a search language (see fig. 2.1) It is here that problems of matching the two arise. Such problems relate to or are caused by a number of factors. The factors in one way or another affect the effort of information users to get precise and comprehensive information on the subject of their interest from the database(s).

The mismatch between the information need and the search result in response to such need is a serious problem in an on-line searching. Problems of on-line searching identified in Fenichel's study include (Fenichel, 1980-81):

1. Approaches to searching vary considerably, from one searcher to another even on the same system and database.

2. Major problems of most searchers are with the search strategy and not with the mechanism of the retrieval system.
3. Many searchers perform simple searches and do not browse the retrieval references to check the adequacy of a search formulation or to improve a search; that is, they do not make full use of on-line interactive facilities.
4. The on-line search process tends to be sensitive to many factors in addition to the skill of the searcher and the nature of the question. For example, the institutions's attitude to on-line searching, the changing policy and economic considerations, and the search environment, etc. all affect on-line searching.

The major causes for mismatch between the actual information need of end-user and search results occurring at different stages of the search procedure include the following (Guinchat and Menou, 1990: 314)

- Problem recognition and actual information need of end-user,
- Information need as perceived by end-user,
  - definition of problem
  - knowledge of available information
  - knowledge of information source
- Information need as expressed in end-user's query,
  - communication means used (letter, phone, in

2. Major problems of most searchers are with the search strategy and not with the mechanism of the retrieval system.
3. Many searchers perform simple searches and do not browse the retrieval references to check the adequacy of a search formulation or to improve a search; that is, they do not make full use of on-line interactive facilities.
4. The on-line search process tends to be sensitive to many factors in addition to the skill of the searcher and the nature of the question. For example, the institutions's attitude to on-line searching, the changing policy and economic considerations, and the search environment, etc. all affect on-line searching.

The major causes for mismatch between the actual information need of end-user and search results occurring at different stages of the search procedure include the following (Guinchat and Menou, 1990: 314)

- Problem recognition and actual information need of end-user,
- Information need as perceived by end-user,
  - definition of problem
  - knowledge of available information
  - knowledge of information source
- Information need as expressed in end-user's query,
  - communication means used (letter, phone, in

person, etc.)
    - conscious retention of the exact parameters of the problem
    - communication ability of end user

- Information need as interpreted by information specialist (IS),
    - subject knowledge of IS
    - IS's understanding of end-user's needs
    - IS's aptitude for effective dialogue with end-user
    - IS's capacity for expression

- Query formulation in the language of the system,
    - knowledge of search language
    - capabilities of the search language

- Formulation of search expression and strategy,
    - searchers experience in searching particular search service and database
    - adequacy of the search logic

- Searching and retrieval operations
    - capacity and flexibility of system
    - structure and organization of databases

- Intermediate results
    - adequacy of databases
    - validity of judgement of pertinence

- Final results
    - refinement of search expression
    - modification of search strategy

The unsatisfactory search results can arise from inadequate end-user and information specialist interaction, selection of inappropriate database(s), inadequacy of and/or inadequate understanding of the search language of the database service. These inadequacies necessarily result in less effective searcher-system interface, and others. A knowledge about and an understanding of the characteristics of on-line services, their databases, the search language used, the

formulation of search expression and search, and specification of user needs can, therefore, assist in adopting measures and techniques to improve search results with respect to end-user's needs. For a searcher without adequate knowledge of the design and structure of the searchable field, the capabilities of the query and indexing languages, etc., it is quite difficult to get the required level of satisfactory search results. Variations can occur across database files, across the searchable indexes, and across the query languages used. These difficulties will be complex and compounded when a number of databases are to be searched in different on-line search services.

In addition to the above, synonyms, spelling variations, abbreviations, grammatical variations, variable prefixes, preferred local terms, homonyms, and homographs are possible causes of mismatch in information retrieval from databases. These are oresented below.

Relationship	Example
1. Synonyms	Trucks, Lorries Aves, Birds File structure, File organization
2. Spelling Variations	Fiber, Fibre Center, Centre Labour, Labor On-line, online, on line

3. Abbreviations	Ltd., LTD., Limited CAD, Computer aided Design
4. Grammatical Variations	Make, Making, Makes Weed, Weeds, Weeding
5. Variable prefixes	Bisulfide, disulfide Co-author, Joint-author
6. Preferred local term	

Then there are homonyms and homographs:

Spring (season); Spring (machine element); Spring (jump) Spring (water source).

Control of these type of equivalent relationships among terms/concepts is very important and the usual means for this are through the use of vocabulary control tools.

2.2.7 The User System Interface

Another area of the problem of matching between the user's query and the document surrogates is the user interface. An end-user of a search service may not always be able to specify his/her information needs precisely. In this regard, Lancaster has this to say:

> " 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 searching strategy for machine search on the same vague basis. If a machine search is to yield useful results, we must do as much as possible to obtain request statement that explicitly delineates the actual information need. The wider the gap between stated request and information need, the less successful the search is likely to be." (Lancaster, 1968).

This points to the need to involve users in the search process. It is, therefore, highly desirable that the information specialist or search intermediary conducts an interview with the end-user and also have him/her present to assist during the searching process. The combination of the precise subject knowledge of the end-user and the systems knowledge of the searcher is likely to yield better results from on-line search. The searcher while analyzing the concepts of query can help and be helped by the end-user in clarifying and precisely expressing the information need, through a search interview. In this process, vocabulary control tools, such as, a thesaurus or scheme for classification and known papers covering the subject of the query can assist in defining the scope of the subject of the query, the relevant core concepts, related concepts, etc.

Participation of the end-user in the search process can not only facilitate zeroing on the end-user's precise information need but also exercise his/her personal point of view (Neelameghan; Divina, Pasqua-Cruz, 1983; Neelameghan, 1992). In this context, end user interaction with the system will be via the search intermediary. It is also desirable that each request for on-line search received in person, by telephone, letter, e-mail, etc., be recorded in a search request form designed for the purpose.

The search request form should be designed to elicit such information that would help the searcher in:

- analyzing the problem expressed in end-user's query,
- constructing a search profile,
- selecting the concepts appropriate to the search, and
- translating each concept into the search language of the on-line search system.

2.2.8 Search Strategy

A search strategy is an overall plan or approach to a search problem. The major approaches to searching are (Markey and Atherton: 1981; Harter, 1986, chap. 7; Hartley, 1990: 170-71):

Briefsearch

This is a search that is a fast and cheap way to get a rough idea about what a database contains on a given subject. It usually involves a single search formulation, a boolean combination of a few search terms without spending time to use vocabulary control tools to select synonyms, etc. There may be little interaction between the

searcher and the system. The number of records retrieved will be few, i.e. recall ratio will be low. It may not be an ideal search strategy for full fledged search. However, specific term(s) used in a search may result in high precision and also reasonable high recall. Brief search can also be used to retrieve a record for a document known to be relevant to a query, e.g. by words in the title, name of author, etc. Such a record can provide additional terms to build up a more in-depth or complete search strategy.

Building-block Method

This method reduces a request to a series of sub-problems. In building-block strategy each concept of the query is enlarged/enriched by synonyms and related terms (selected using appropriate vocabulary control tools) using boolean **OR** to building smaller blocks which are then **AND**ed together to produce the answer set. The principal steps in this strategy after query negotiation, formulation of objectives of the search and selecting appropriate search service and database(s) include:

- identification of major concepts and facets, and the interrelationship among them,
- formulating search strings that represent the concepts, phrases, descriptors, identifiers, qualifiers,
- determining database fields to be searched,
- for each component of the search expression, a set of postings will be created. The sets are **OR**ed,
- the sets resulting from the above step are then combined using boolean **AND** and **NOT** operators,
- intermediate results are evaluated and search terms and strategy modified, if necessary, and iterate the process to obtain desired results.

Use of vocabulary control tools is helpful and advisable in the construction of the blocks. Otherwise, it may take time in keyboarding it into the computer.

Successive Fractions Method

This search strategy involves iterative refinement of broad starting strategy. It is a method used to reduce a large set created by using **AND** or **NOT**. In this method, terms specifying language and date are useful in reducing the number of records retrieved.

terms and translate them to the artificial language that a particular search system's query language can accept.

Thus, preparing a search expression consists of translating the concepts in the search request or end-user's query into the language of the database or information retrieval service selected. If the database has an associated vocabulary control tool, such as a list of subject headings, a thesaurus, and/or a classification scheme, then it should be referred to (if not already done in preparing the search request form), in order to select appropriate near-synonymous or equivalent terms, and broader, narrower, and related terms.

When a search in a broad subject area is needed, consultation of the documentation on the database will indicate whether such features for search are provided for.

If the terms from the vocabulary control tools do not adequately cover a concept, the database may use natural language terms as index terms; i.e. free terms, text words, identifiers, etc., that may be applied in addition to or instead of, the controlled vocabulary terms.

The third approach is a compromise between a natural language search and self-contained query language. In this approach, the system displays a menu from which the searcher selects one of a number of options and in this sense the computer directs the search. Alternatively, the searcher may be directed to fill in blanks on the screen or satisfy requests for specification of parameters.

Computer directed search/querying offers minimum choice and may be suitable for beginners as it requires minimum typing and training. However, unless there are good feedback procedures, the whole process can become long and tedious.

2.3 JUSTIFICATION

2.3.1 Relevance to Local Databases

In the discussion of on-line search procedures, a number of major problems with respect to user-system interface have been indicated. These include:

- basic command language differences
- variations in the search languages and search capabilities of the system

- inadequacy of boolean based searches and retrieval
- restrictions in formulating users own output format
- lack of in-built controlled vocabulary such as thesaurus, classification scheme, subject headings, etc. in some of the databases

On the basis of various studies on information retrieval language and procedures, solutions, such as, weighting, ranking, incorporation of in-built vocabulary control tools, incorporation of user-interface aids, etc. have been suggested to mitigate or overcome the problems of user interface.

The problems identified and the solutions suggested, mostly revolve around the satisfaction of the information needs of the end-users of databases by the search services, and point to the essential link to be established between user and the system.

On the other hand, the problems identified and the solutions suggested mostly relate to bibliographic databasess their structure, format, etc. which are more or less standardized. Very little has been done in evaluating search procedures and problems for non-

bibliographic and integrated databases which may be made up of bibliographic, source and object-oriented specialized databases. Also, whether the problems that have influenced on-line bibliographic databases also affect such databases, and whether the solutions suggested may equally apply or not should be considered. It is also necessary to examine what type of search procedure, retrieval and user assistance mechanisms are required in local specilized databases.

Experience shows that users need a more helpful search procedure so as to enhance and facilitate their interaction with local databases. A more structured formulation of the search expression appears to be preferred. Such structuring of query should be the result of user-system interface and based on an analysis and knowledge of the subject concepts in the field of the user's concern. In other words, to enhance user's interaction with the system it is necessary to design a user interface system. It is thus one of the objectives of this work to examine and apply various vocabulary control tools in designing a user interface system that will hopefully enhance retrieval of information for satisfying the needs of planners, executives, etc., in the urban planning field. This is a significant issue with respect to the recent developments in object-

oriented and integrated databases. It is more so in those fields that are of a specialized type, such as NUPI, needing organized, synthesized, and consolidated information/data to prepare urban plans and make decisions of various kinds.

2.3.2 Special Needs of NUPI

The National Urban Planning Institute (NUPI) was established by proclamation No. 317/1987 of the government of Ethiopia, on 13 June 1987 as an autonomous government organization.

In the establishment proclamation, the following three broad objectives of NUPI have been clearly defined:

(1) to carry out appropriate research and studies which are necessary for the preparation of various urban plans,
(2) to prepare plans, on the basis of the researches and studies it makes, for regions, metropolitan area, and urban centres,
(3) to train the necessary manpower for the preparation and implementation of the various urban plans.

The statement of NUPI's objectives shows that the scope of operation of the Institute is very wide. The scope going up to regional level implies that the urban plans it prepares should be based on an integrative approach giving appropriate attention to the natural interaction and interdependence between a given urban centre and its surrounding regional environment.

In order to effectively discharge its duties and responsibilities, NUPI has structured itself into appropriate supporting and line organizational hierarchies. The line organizations carry out the main functions of NUPI where as the supporting structures assist the activities of the line structures. NUPI needs experts, planners, and researchers of different specialization and background.

Generally, all planning endeavours are future-oriented. This requires the availability of accurate timely, comprehensive, reliable and precise information on which the planning activity should be based in relation to its specific objective. Planning requires knowledge of the phenomena or object to be planned. Knowledge, on the other hand, depends on information and information is to be supplied by an information system of one kind or another. Only an urban planning based on the knowledge of

There are two units in the supporting staff structure of NUPI that cater to the processing and provision of information/data. These are the Computer Centre (CC) and the Documentation and Information Centre (DIC) (see figure).

The CC of NUPI is equipped with a mainframe and microcomputers which are used for a number of purposes. The mainframe computer (HP-9500 Series) is mostly used for processing data collected through field surveys and other data collection means using home developed data processing programs. It is also used for storage of the databases of processed data and for retrieving information from the databases when needed both by internal and external users. The main storage media for the databases and other data is off-line using magnetic tapes. Thus the centre holds a number of databases of processed data on different urban centres for which plans have been prepared and other special purpose projects. The microcomputers are almost exclusively used for word processing purposes. However, it is worth mentioning that these computing facilities of NUPI are under utilized or are operating below capacity.

The DIC of NUPI has such activities as the provision of library services to internal and external information

users and collecting various information materials (books, reports, documents, periodicals, maps, plans, etc.) with special attention given to the urban planning field. DIC has collections on different disciplines.

The CC and DIC of NUPI form an information system for the planning functions of NUPI. NUPI, in addition to the main objective it is entrusted with, is also responsible for the establishment of an information or data bank by collecting various documentary and non-documentary information materials from any source relating to the urban system of the country. The two units, CC and DIC, in addition to providing information to NUPI's regular planning activities, they are also expected to work towards the realization of the establishment of databank (urban information system) stipulated to be created for the urban system of the country as whole. The databank is aimed to serve as a central source of information for functions relating to urban system.

As can be seen from the enumerated objectives, NUPI is a specialized Institute having a sectoral scope of operation (i.e. urban planning). Urban planning being an integral part of the overall planning system, must cover all fields of socio-economic and physical factors. By implication, therefore, it requires planners, experts,

researchers, etc., as mentioned earlier, from various fields of specializations. The specialists also generate a considerable amount of information and data for use in the planning. These specialists need information from different areas of their respective fields of interest which are needed for their information production and planning activities. The type of information most needed by the specialists is value-added information, which is analyzed, synthesized and organized in a manner convenient for their use. Such value-added information is not obtained from the general information systems and on-line database services.

Provision of such information and information service calls for the development of local integrated database system in the specilaized areas of interest to NUPI. Most of the databases will contain data (numerical, statistical, object-oriented) extracted from different areas (surveys, computer processed, documentary), analyzed, synthesized and organized by specialists in the area of urban planning.

In interfacing with such database systems, the users find it more confortable, if the terminology used in the database records are those the users themselves use in their work, the information is presented in forms and

formats conducive to ease of use, and facilities are provided for presenting the data in different ways.

In recent times, there has been renewed interest in the application of vocabulary cotrol tools and the theoretical principles on which they are based, such as, faceted classification, thesaurus, etc. in the design and development of databases containing analyzed synthesized information, object-oriented databases, knowledge-based sytems, etc. and in developing mechanisms for user-interfaces, including expert systems (Neelameghan, 1991; Neelameghan, 1992).

This study, therefore, attempts, with specific reference to the infoamtion needs of NUPI, to:

a) computer-generate vocabulary control tools, e.g. thesaurus, from the text of materials generated/used by NUPI,
b) use the tools generated at (a) to update/enrich other thesaurus (e.g. Macrothesaurus) vis-a-vis urban planning topics; and to enrich a classaurus type thesaurus developed by conventional method,
c) develop a user-interface to a factual object-oriented database serving NUPI's interest.

It is hoped that this work will not only contribute to the improvement of information services at NUPI but also to further studies in vocabulary control tools construction, user-interface development, etc. in relation to object-oriented databases and integrated databasses.

CHAPTER THREE

METHODOLOGY

3.1 DATA COLLECTION METHOD

The data required for the purpose of this thesis are of three types. One set of data was required to create the integrated urban planning database. These are prototype databases created to show the significance of local databases to fulfil the information requirement for urban planning activities. The second set of data required was requests or queries for information from potential users of the system developed. The third type of data required was terms or vocabularies for thesaurus construction.

Appropriate forms and formats were prepared to gather data to be input into the integrated database and to collect user queries. Four questionnaire-like worksheets were prepared each collecting respectively (1) document descriptions, (2) profiles of institutions, (3) profiles of projects, and (4) profiles of persons. The forms are presented in the appendices.

The data required for the creation of the object-oriented database was collected using free formats since a common form could not be applied for that purpose. Records from

other databases were also downloaded on diskettes (see 3.3).

For the purpose of collecting terms or vocabularies required for the development of thesauri, a form, referred to as term card was used (see 8.3.2)

3.2 SAMPLING METHOD

Stratified sampling method was used to identify the potential user-requesters from NUPI. Accordingly, the experts in NUPI were grouped on the basis of their specializations and work areas (i.e. departments). Then from each group some 80 persons were selected and given the information search request forms (8.2.1).

3.3 SOURCES OF DATA

The sources of data required for the creation of the prototype databases were the National Urban Planning Institute (NUPI), the Pan-African Development Information System (PADIS), and the Ethiopian Science and Technology Commission Information and Documentation Centre (ESTC/IDC), and other organizations engaged in activities

the relevant respective chapters or sections.

3.5 METHOD OF PRESENTATION

The thesauri, the integrated urban planning database, and the user-interface facility are available on diskettes.

CHAPTER FOUR

URBAN PLANNING AND INFORMATION SUPPORT

4.1 URBAN PLANNING

4.1.1 General

Before discussing the nature of urban planning, it is useful to make clear the meaning of planning in a more general sense. It will help in avoiding ambiguity which characterizes the term 'planning' in all subject areas including the urban planning field.

According to Roberts (1985:4) planning consists of making choice among the options that appear open for the future and then securing their implementation, which depends on the allocation of necessary resources. Planning is an activity involving decision making. Elements of decision theory can, therefore, be helpful particularly in treating both the values and the uncertainty of different possible futures. In this context, the following can be quoted:

> As planning is decision making and resource allocation activity, it is political in that the options have to be selected that will not benefit equally, or equitably all the members of society (Robert, 1985:4).

Myrdals agrees with or supports the idea of Robert. His view makes it clearer that planning involves societal choice about the future, and that choice implies political intervention (quoted in Bracken, 1981: 9). To exercise choice, however, genuinely alternative views of the future must be presented and in order that selection may be purposeful and deliberate, both a knowledge and understanding of the present and explicit anticipation of the future consequences of the present action are required. Such knowledge and understanding can only be effective if the kind of information needed on each of the available alternative actions in the uncertain future is accessible at the right time. Planning as a general activity is the making of an orderly sequence of action that will lead to the achievement of stated goal(s) (Hall, 1982: 6).

Urban planning is an integral component of the comprehensive socio-economic development planning. It covers all aspects of planning in general its domain and its influence area being the urban area (including metropolitan area, suburbs, etc.). As Meyriat points out:

> ... Even when planning is sectoral (e.g., urban planning), it must take into account many other sectors which are directly or indirectly connected to the sector under consideration, all being interrelated as elements within the same national system (Meyriat, 1991: 1)

Unless it takes into consideration that, urban planning cannot have full meaning. As a result, the size of urban planning endeavour is large and its scope wide as all economic, social, and physical agents or factors in the area are involved. Urban planning concept is a terminological problem than the term planning.

One of the characteristics of urban planning is the endemic confusion among those involved in it not only about its purpose, but also about its object of concern. A long list of apparent synonyms and near synonyms can be found in the literature on urban planning. For example, physical planning, land use planning, town planning, town and country planning, environmental planning, urban planning, urban policy and urban design (Healey, 1986: 3). If one tries to analyze what actually each term means, it is not difficult to notice that they are not all synonymous but overlapping.

One could treat this proliferation of terminology as unimportant, but it is not just an issue in semantics. Each term tends to cover a different 'set' of objects of concern (Healey, 1986: 3). Thus, for instance, urban policy may refer to resource allocation without an explicit spatial dimension, while physical planning may focus on urban physical development with very little relation to social and economic policies. Some argue that

the terminological confusion is deliberate, that each of the meanings about what 'urban planning' is and does is functional to the state's needs to legitimize its activities (Scott and Rowed, 1977: 1114, Castells, 1977: 76). In this study, for practical application, urban planning is used to represent the combined senses transferred by all terms used as synonymous.

Urban planning conventionally means something more limited and precise: it refers to planning with spatial or geographical component, in which the general objective is to provide for spatial structure of activities (or of land uses) which is in some way better than the pattern existing without planning. Such planning is also known as "physical planning"; spatial planning is perhaps a more neutral and more precise term. At this stage of its development, however, urban planning is not limited to only physical and spatial aspects of the urban area but also comprises other socio-economic factors in the urban area and related environs.

4.1.2 The Urban Planning Process Cycle

The urban planning process may be viewed as community of major activities. McLoughlin (1968) gives the following generalized but all inclusive planning cycle:

1. The decision to adopt planning. Planning involves resource commitment and choice making. Thus, the planning activity must be initiated by a clear and firm decision.
2. Goal formulation and definitions of the objective(s) for urban planning. The planning should aim at accomplishing a certain objective.
3. The study with the aid of model of the urban system of possible courses of action. Such studies show how the urban system might change through time under the impact of a variety of influences arising from private and public actions, and interventions.
4. Evaluation of the courses of action in order to select an operational course by reference to assumed social values, and the estimation of cost and benefit.
5. Action to implement the plan, including both direct and continuous control of public and private proposals for change. The essence of control is to study the impact on the urban system whether or not they would deflect it from the course charted for it in the plan using the models developed in stage 3.

4.1.3 Conceptions of Urban Planning

Urban planning is conceptualized in different ways and levels, usually following the development of the discipline itself. These different conceptions are summarized in the following sections.

4.1.3.1 Comprehensive Planning

Comprehensive planning refers to planning to cover development which uses land, in order to maximize the overall benefit, and then ensuring adherence to the scheme. The techniques are mainly those of survey and fairly simple analysis and techniques as such are not greatly emphasized (Roberts, 1985: 4).

Regarding the functions of comprehensive planning, Alshuler has the following to say:

> One, to create a master plan which can guide the deliberations of specialist planners; two, to evaluate proposals of specialist planners in the light of master plan; and, three, to coordinate the planning of specialist agencies so as to ensure that their proposals reinforce each other to further the public interest. (Alshuler, 1965: 186).

In comprehensive planning, the urban planner sees himself as a guardian of public interest (Friedman, 1965, 195). Two key characteristics of comprehensive planning are

that the plan is entirely physical in scope and that it is detached from the decision making process (Hansen, 1968: 295). Hanson adds that traditional planning [comprehensive planning] ... tended to be physical in scope, detached from decision making and technically primitive. In reaction to this weaknesses of comprehensive planning, the more recent conceptions of urban planning process have developed. They are structure planning, the systems approach and advocacy planning.

4.1.3.2 Structure Planning

Structure planning is a written statement of broad strategy with the effect of the strategy discussed in relation to alternatives. Structure plans are concerned with long range (20-30 years), and, therefore, detailed allocations of site definition cannot be made in map form although diagrams may be useful supplements to the written text (Massey and Cordy, 1971). Structure or strategy planning is more activity-oriented than land use-oriented ones and is more explicitly concerned with implementation and decision making factors (Roberts, 1985: 50). Friedman views this planning conception as being still the old language, but perspective is new: planning and action are brought together and fused. Structural planners are therefore said to require

evolving policies' and 'the design of finite scheme of localized change' for execution in the near future.

In the structure planning, three broad stages are defined (Roberts, 1985: 16).

The First Stage

The survey and preliminary analysis and the formulation of objective for the strategy. This entails research into the physical potential of open country, for existing uses and for urban development, into improvement needs and possibilities in existing built-up areas, and into anticipated growth and change in population, economic and social characteristics of the urban region over the next 30 years.

The Second Stage

Strategy Formulation and Testing. Alternative strategies based on 'indications of needs, opportunity, potential and trend derived from the survey and analysis'. The alternatives are tested in terms of their feasibility, flexibility and acceptability.

The Third Stage

Advice on implementation including a monitoring system and program for development. It is worth stressing the two components of the elaborate process that are most critical for the final result: the initial selection of lines of investigation and the methods of testing alternatives. Thus structure planners take main theoretical concepts to guide their line of investigation and from the theoretical points identify representative strategies for testing.

4.1.3.3 Systems Approach

The systems approach to planning places the greatest emphasis on technical expertise - in analyzing the urban system, in forecasting future change and in simulating alternative futures. It is characterized by its view of the subject matter of planners as systems and sub-systems of main activities with their manifestations and interrelationships.

Crips and Hall (1969), exemplifying the systems approach, expressed their purpose thus: 'We are concerned to identify the system of interest to the planners through a review of recent theories of urban growth and

development, and to agree about the process of planning the urban system'.

In the planning profession, there appears to be no general agreement on either the nature or the context of the planning process although recent theory and practice are beginning, in a few places to bring some coherence to the process of urban and regional, country and town planning (Crips and Hall, 1969: 27). The coherence is founded on the systems approach. To justify their view, Crips and Hall quote from McLoughlin: 'The system that lies at the heart of the planners concern is composed of (all) types of human activity connected by flows of people, material, energy and information. The physical framework for the system consists of buildings and their cartilages, open spaces, agricultural land and other adapted spaces, while the flows are accommodated by roads, railways, pipelines, wires, cables serving as communication channel (Crips and Hall, 1969).

Undoubtedly, the systems view of planning and plan making process is now quite widely held, both among theorists and practitioners (Roberts, 1985).

4.1.3.4 Advocacy Planning

Despite the wide acceptance of the systems approach in urban planning process and systems, there are also criticisms about it especially by those who prefer 'advocacy planning'.

Advocacy planning contends that the planner should be a pleader of particular needs and approach, and that he requires techniques to build up a weighty and convincing cause to their solution (Roberts, 1985: 27). In this concept of urban planning, different interests are encouraged to debate alternative policies and plans for an area (Davindoff, 1965), providing for participatory democracy in which the state should be some form of benevolent arbiter, allowing and encouraging open informed debate about issues as seen from different view points.

The justification for such an approach is often based directly on spot-lighting the inadequacies of other approaches to urban planning. Thus, in an early plan for advocacy and plural planning, Davindoff stated:

> City planning is a means for determining policy. Appropriate policy is determined through political debate. The right courses of action is always a matter of choice, never of fact. Planners should engage in the political

> process as advocators of the interest of government and other groups. Intelligent choice about public policy would be aided if different political, social, and economic interests produced city plans. Plural plans rather than a single agency plan should be presented to the public (Davindoff, 1965: 331).

From these different conceptional approaches to urban planning, it is evident that urban planning must be evolutionary and responsive to its own history of development, as is any field of study or profession building on previous debates and experiences. It must also be reactive against whatever views have previously excessively embraced. The overall outcome is, therefore, a premature discarding of much of potential value. Planning should also take into consideration the factors that define the planners own role, i.e. appearances of new needs and opportunities, changes in the general socio-political climate, etc. In this regard, Dyckman has this to say:

> Planning must also understand something of the socio-political-cultural system in which planning is particularly understood where it takes on its meaning and eventually effectuated (Dyckman, 1969: 300).

The conclusion is that it cannot be said exactly which one is the most appropriate (right) method (concept) of

of 'reasonableness' in the exercise of choice [of alternatives], and, 'comprehensive', by which an adequate understanding of the nature of the problem has been achieved (Bracken, 1981: 10).

It is not possible to attain complete rationality. A more commonly accepted notion is that of bounded rationality (Kaplan, 1964) which essentially focuses upon reasonableness or fairness of decision making. A rational decision is one where all the various alternative courses of action are considered, the consequences resulting from them are identified and compared, and the preferred alternatives selected in the light of the most valued ends (Ratcliffe, 1981: 117). In practice, this involves the provision of relevant information to decision makers about what exists in the planning region and what may be expected under certain conditions.

In general, urban planning is aimed at efficient management of resources and rationality in directing to solve the physical, social, and economic problems of the planning area as well as to improve the living condition of the particular urban region. It is also to be noted here that efficiency cannot be meaningful without the effectiveness in the management of resource and decision making. This is achieved through a coordinated,

integrated and systematized planning of the urban centre and its area of influence. Particularly this is true with the multi-disciplinary nature of the urban planning discipline.

4.1.5 The Multi-disciplinary Nature of Urban Planning

Until the mid-1960s, urban planning was practised principally as an art within a traditional design context (Hall, 1982). In this context, its purpose was seen principally in the promotion of new development that was aesthetically pleasing and which complied with certain standards of lay-out. Step-by-step, the process developed to include in more general terms the amenity set in rather covert assumptions about service to the public interest. The pursuit of amenity inevitably became concerned with the need to ensure that actions of one individual, groups or section of a society minimally affect adversely the interests of others, i.e. the amenity of others. Urban planning subsequently became concerned with attempting to resolve conflict over needs and actions between members of society (Bracken, 1981: 11).

The needs and actions of any society is of physical, social, and economic in nature. As a result, the scope of

urban planning should encompass all these areas. Contemporary urban planning is a multi-disciplinary one. It is not sufficient to deal only with physical design and spatial analysis (such as land use patterns and allocations) as it was the case in the traditional urban planning method. The various aspect of economic activities being undertaken and planned in the region, the population dynamics including in/out-migration, the social amenities and facilities, and the multitudes of socio-economic infrastructures of the urban centre should be studied, analyzed and modelled in the plan. The integration and interaction of the urban centre with its environment (metropolitan region and rural area) must be incorporated into the plan as the centre cannot exist in isolation from its influence area or region, a self contained entity. Thus, the process of urban plan preparation involves, a systematic study and analysis of the physical, social, economic and other dynamic aspects and factors which mutually affect one another and are interdependent. This makes the planning process more rational since all the relevant information is made available through the multi-disciplinary studies and analysis.

4.2 TYPE OF INFORMATION NEEDED FOR URBAN PLANNING

Any kind of planning is conducted and operated within a society. Therefore, the planning process should take into account the characteristics of the society. The society is characterized by a number of varying and interacting elements. The process of planning also involves formulation of policies, choosing between the policies, measuring the policies, and adjusting the policies when necessary. In order to understand all these aspects of planning in a community, the planner, regardless of the subject matter and area of planning, should have access to all the data and necessary variety of information. 'Planning must be based upon knowledge, and knowledge depends upon information ...' (Ratcliffe, 1981:125). Meyriat supports this contention writing: '... planning is a decision making process, decision cannot be made without information, so information is the basis for every planning...(Meyriat, 1991: 1) Meyriat added that planning is indeed a way of dealing with information. This applies equally to urban planning.

Planning essentially is future-oriented. Such future-dependent activity is full of uncertainties. To reduce the risk involved in such an uncertain situation, to identify and know the right type of information on which

to base the planning is of great importance. The information on which the planning relies should be as complete as possible. Information is needed on both the present and recent past relating to the locality for which the plan is prepared. Urban planning being all-embracing, multi-disciplinary in nature, and having comprehensive terms of reference, a wide and detailed survey of the kind of information needed to form a sound and implementable plan is crucial for such information assists in identifying the factors that affect and/or make it reflect the actual requirements of the society. For that purpose, the area on which the information is required should cover all sectors of economic and social activity as well as the physical characteristics of the centre to be planned.

Planning at the same time is dependent on many constraints; information on political, legal, organizational, psychological aspects of the society are needed. In addition, the world background have to be taken into account, so that foreign data are also wanted. This is particularly important to learn from the experiences of other countries.

Besides socio-economic-physical data, urban planners should also resort to information of a more fundamental

and even theoretical nature. They have to know the planning methods which are evolving at the rate of economic knowledge (Meyriat, 1991:2).

Based on the experiences of other countries and, particularly, the experience and practice of the National Urban Planning Institute (NUPI), the following ranges and types of information have been identified as essential for preparing urban plans:

- Maps and plans data,
- Population and housing data,
- Social service and infrastructure data,
- Transport and communication data,
- Traffic data,
- Meteorological, geological, and hydrological data,
- Agricultural, industrial, handicrafts and commercial data,
- Employment and unemployment data,
- Historical data,
- Utilities data,
- Financial and managerial data.

The above areas of information need can be categorized into four broad groups: physical, Social, demographic, and economic data. These groups of information need areas indicate that urban planning is multi-disciplinary requiring information on several areas of human activity.

4.2.1 Physical Characteristics Information

The nature, form and scale of the urban physical environment are the canvas upon which the urban plan is painted and a knowledge and record of the physical conditions is essential for urban planning. The type of information required include:

- topology
- geology
- hydrology
- meteorology
    - rainfall
    - temperature
    - humidity
    - evaporation
    - etc
- minerals
- morphology
- areas of special interest
- location
- altitude
- rich agricultural land
- sources of population, etc.
- air pollution, noise, congestion, etc.

There are also other elements on which planning information is sought. The information on physical conditions is used to construct land use and other maps which indicate the extent of existing development and the potential trends and directions of future development. In addition, such information is also used to identify the available natural resources of the urban area. The information must be kept updated in relation to land-use, services and utilities maps. This, in turn, helps in understanding the nature of urban growth and change and the patterns of settlement over time. Maps and plans also

include topographic maps, maps and plans of networks of existing road system and utilities (water supply, electricity, telephone, sewerage, etc.), natural resources, coverage maps and others.

4.2.2 Demographic Information

Population data is nearly always the starting point and a basic requirement that influences the course that the urban planning should take. This is mainly so because, the prime objective of planning is to solve the various problems of the community residing in the urban environment. Information is required on:

- size
- structure
- density
- characteristics (dynamics)
  - birth rate
  - death rate
  - migration
    - in-migration
    - out-migration
- distribution
  - existing population distribution
  - projected future projection

The knowledge of the existing and likely future needs of the community in terms of family size, age, and structure (sex structure and sex ratio) are the yard sticks in urban planning. The growth rates, migration system, place of birth, length of residence, etc. are important features on which information is required by urban planning.

4.2.3 Housing Information

In relation to housing and its physical characteristics information is required on:

- housing size (refers to the number of housing stock or dwellings)
- housing condition
  - the quality and standard of housing units
- age of housing stock
- tenure
  - public housing
  - private housing
  - rented housing
  - owned, etc
- occupancy rates
- type of construction material used
  - roofing material
  - walling material
  - ceiling material
  - floor material
  - type of housing
  - story building
  - traditional
  - single story detached
  - single story raw
  - housing units
- basic housing infrastructures
- household size
- housing demand
- housing supply
- rent distribution
- prices of housing
- construction plan, so on.

This information forms the basis for planning future housing needs and determine policy with respect to rehabilitation, redevelopment, overspill schemes.

4.2.4 Economic Activities Information

Information on a wide range of economic activities is required for urban planning. Such information cover such topics as:

- type of economic activities
- products of economic activities
- economic sectors
    - Agriculture
    - Industry
        - small scale and handicrafts
        - modern manufacturing
    - Services
    - Commerce and trade
    - and others
- occupation
- dependency ratio
- employment status
    - by sex and age category
- consumption patterns
- labour force participation
    - by age, sex, and sector, etc.
- labour market

This information is used to identify the major economic sectors in the planning area and establish priorities for sectoral development schemes or programmes. The structure of employment status is reflected by employer and occupational categories and is used to identify occupations that are important in providing employment opportunities, information on basic and service employment as well as state of unemployment is needed. Further, information on such economic factors as the type and number of commercial and business establishments, shops, bars, restaurants, garages, workshops, flour mills and many others need to be provided.

4.2.5 Transportation and Communication

One of the principal factors contributing to the size and nature of urban development is accessibility to different location points which in turn depends upon the degree, capacity and the propensity for movement of people and materials. Thus transport and communication are concerned with such functions as origin and destination, pedestrian traffic, public and private transport and other studies. The information on movements of goods and people is applied in transportation and communication studies and planning in the assignment of traffic to networks, in testing of routes capacities and standards. Moreover, information on transportation and communication factors serves to establish systematic and efficient road networks within the urban area and between the urban area and its surroundings to facilitate movements of goods and people and communication between people in the area.

4.2.6 Information on Social Infrastructure

A major area of information requirement for urban planning is about social-infrastructure and services. Education, health, recreation, leisure and other social facilities are among the planning components.

Information on education includes information on:

- existing educational facilities
- type of existing schools and their levels
- enrolment
- number of sections
- student-class ratio
- eligible population for education
- the demand for educational facilities
- school age population
- shortfall or surplus, and so on.

In most cases, supply falls short of demand, in planning the number of schools to be constructed to cater for the future educational requirements in line with the growth of the population in the urban areas relevant data need to be made available to the planners.

Another social facility of prime requirement to the community and on which information is required, is health: existing and future requirements of health facilities such as hospitals, clinics, health stations/centres, and pharmacies.

Other social infrastructures on which information is required include cultural and recreational and other social facilities, such as, religious facilities, libraries, sports centres, theatre/cinema halls, etc, and information should cover existing situation and the future requirements of the community. Some of these

relate to the cultural practices of the particular community be affected by the planning.

4.2.7 Historical Information

For the purpose of establishing a background for the planning of a given urban centre, historical data is needed on the foundation of the town and the types of plans previously prepared for it, and their strengths and weaknesses, as well as the historical trends of development of the area. Such information will also provide an insight and understanding of the previously prepared plans and to capitalize on earlier experience.

4.2.8 Financial and Managerial Information

Financial information is useful to know the relation between the income and expenditure of a particular town, including household income and expenditure, which could show affordability and the capability of the urban centre to implement various projects to be proposed in the plan, and information on taxes, subsidies, etc. Information on management and administration is also needed to understand the structure and power hierarchy of the town for implementing the plan. Administrative information relating to the division of the urban centre in Kebeles,

for accurate, reliable, comprehensive, up-to-date and timely information. Information has been and is always available in principle. The problem is whether the right sources are used at the right time, and whether the acquired information is accurate and reliable with respect to the purpose it is intended to serve. Ideally the data required for planning, including urban planning, should be obtained first hand through survey. However, owing to the constraints of time, money, etc., it is practically impossible to do in most instances. As a result both primary and secondary sources of information are also used. The information needed may be collected by the planning institution it self or may be obtained from the an institution which is involved in collecting data on the required variables and activities of the society in a specific area. But the latter depends on the degree of organizational development of the country or area and upon the interests of the organizations in the field to be covered. In other words, if the organization which collects, organizes and maintains information does not cover the area of interest to the institution that needs information for specific purpose, then that source institution is of no uses. This necessitates collection of information from the sources by the user institution itself.

As is true with any other field, urban planning is also dependent on three major types of information sources. These are documents, institutions, and human.

4.3.1 Documentary Sources of Information

The documentary sources of information include books, periodicals, research documents and reports, indexes, abstracts, reviews , encyclopedias, directories, those available internally and externally. These documentary sources are used either to provide a background to the planning work or supplement the data collected by field survey using questionnaire and other information gathering techniques. For documentary sources access to library collections and other information and documentation centres is essential.

4.3.2 Institutional Sources of Information

Institutional sources of information are the most important in the area of urban planning. The reason is that the information required for urban planning is multi-purpose and all embracing that it is difficult to get the type of information details required from the normal documentary sources. As a result, they have to be searched for in various institutional sources.

The institutional sources of information, with particular reference to the experience of NUPI include:

- Central Statistical Authority (CSA),
- Ministry of Agriculture,
- Institute of Agricultural Research (IAR),
- Ministry of Health,
- Ministry of Education,
- Ethiopian Mapping Agency,
- National Meteorological Agency,
- Ministry of Mining and Energy,
- Ministry of Transport and Communication,
- Addis Ababa University,
- Ethiopian Electric Light and Power Authority,
- Water Supply and Sewerage Authority,
- Ministry of Labour and Social Affairs,
- Ministry of Information,
- Ministry of Trade,
- Regional Planning Offices,
- Ministry of Culture and Sports
- Many international organizations such as ILCA, ECA, and institutional databases such as PADIS,

The subsidiary units of the ministries; for example, corporations, enterprises, factories, are also major sources of information.

Even though the above list of institutions relate to Ethiopia, the records of urban planning of others countries show that the trend and emphasis on institutional sources of information in most cases are similar.

4.3.3 Human Sources of Information

Questionnaire, interviews and other survey methods are employed to collect information from the persons and authorities concerned with the planning activity. Sometimes informal sources such as peer groups and close friends are useful sources, especially in case of problem of access to the data required through formal channels for one reason or another.

4.4 PROBLEMS OF ACCESS TO INFORMATION

In the course of searching for information as it is true of other areas, there are certain factors that constrain data collection. Some of these are enumerated below:

- data often exist but difficult to obtain,
- Access problem - existing information may be of confidential in nature or assumed to be so,
- Information is often related to inconvenient or in appropriate areas in which case collection is

difficult together with the possibility of being unreliable,

- Gap in time - usually the information is out-dated or is not obtained in time, or is slow in being published, for example, statistical survey results,
- Information is organized inconveniently,
- Information about the right sources of information or the availability information itself is not available,
- Distance, i.e. information is available at remote locations, difficult to access when required,
- Unwillingness of respondents to give information or the right information,
- Poor information services of source institutions,
- Unfamiliar presentation of information.

4.5 NEED FOR AN INFORMATION SUPPORT SYSTEM

In the sections presented above, an attempt has been made to indicate the type of information needed for urban planning, the major sources for this information, and some of the obstacles in accessing the required information at the time required and in the form needed to be obtained. In this section, the need for an information support system for urban planning will be presented briefly.

In order to ensure easy access to information overcoming the access problems enumerated above, urban planning needs to be supported by an effective information support system. An information system is an integrated set of devices, technical, means and abilities intended to capture information pertinent to a given objective and to share it so that it is easily provided to decision makers at any desired moment (Meyriat, 1991:5). This should be the way the information needed by NUPI for urban planning is provided to the planners. The information and data needs of urban planning, quantitative and qualitative, and pertinent to local, national, (and international) social, political, economic, technological, cultural and other environmental concerns, should be provided for in the information bases of the information systems and services for urban planning. An information system may be one component part of a large organizational or a separate organization unit. In the case of NUPI there are two units that are engaged in the processing and provision of information: the Computer Centre and the Documentation and Information Centre. Both can be effectively directed towards the creation of an integrated information support system using the available information technology in NUPI. For that purpose, it requires the integration and unification of these two separate units. Needless to say that the effectiveness of

information system may be greatly enhanced by the application of new information technologies that do exist in NUPI. In the design and development of information system in NUPI, with a view to the type of information needed and the sources of information discussed earlier, the application of the vocabulary control tools and techniques will be very helpful. It is with this understanding that these tools and techniques to be used in the system need to be generated and developed in this framework. It is the major concern of this work also.

CHAPTER FIVE

VOCABULARY CONTROL TOOLS

Vocabulary control tools are systems which provide a standardized vocabulary to be used in the information storage and retrieval functions. They play an important role in overcoming the problem of matching the terms in the records in the databases with those in the users queries. This chapter discusses the structure and display of vocabularies, terms collection methods and some typical vocabulary control tools.

5.1 VOCABULARY STRUCTURE AND DISPLAY

Lancaster (1986: 13-17) gives three possible ways of structuring and displaying vocabularies in a vocabulary control tool. One is the systematic tree structure. This structure shows genus/species relationship and is acceptable with the exception of large hierarchies and extensive vocabularies. The second and more commonly used type is that lists terms using indentation to show the hierarchical levels. This scheme has problems in that it is not easy to grasp the term hierarchies. The third way is to arrange them in alphabetical sequence and interperse with terms in other hierarchies. In this structure and display the generic term is referred to as

For the term to be collected there are two requirements: literary warrant and user warrant. The collected term is justified only if it is known to occur in the literature of the subject field. Lancaster has extended this principle as follows:

> A term is justified only if it occurs often enough in the literature to be considedred significant and useful for retrieval purposes.

To derive terms is to consult the sources in which the terminology of the field appears in highly compact form. Such sources are dictionaries, glossaries, encyclopedias, handbooks, and comprehensive text books provided they are up-to-date. An abstracting publication in the subject field is also a good source of terms. The criteria for inclusion in the vocabulary control tool is the frequency of occurence of the term. This principle also applies equally to synonymous and near synonymous terms.

Another important source of terms is user's queries. A term is justified if it is of interest to the user of the information service (Lancaster, 1986: 26). This is useful in establishing appropriate level of specificity in the vocabulary.

User warrant implies that one must collect terms from the potential users of an information service terms that

represent their particular subject interest. This can be done by reffering to the historical requests of users or to collect them by the use of interview or questionnaire. Terms collected in this way are processed in the same way as terms collected by other means.

The terms collected from the two sources are then organized into a coherent and cohesive structure hierarchically and associatively. The process can be effectively done by the application of computer.
The use of vocabulary control devices as a mediation gives information storage and retrieval a unique logic and structure. Vickery (1971: 37) lists the standard steps to be followed in achieving this unique logic and structure as follows:

1) A decision is taken as to what kind of key will be used to identify documents;
2) Appropriate keys are derived from or assigned to each document;
3) The key so selected may be standardized and/or coded;
4) Physical record is prepared, associating the profile with an identifier for the document;
5) All the records so prepared are filed;

The significance of vocabulary control tools in database searches is undeniable. The mismatch between users query and the search result may be controlled to some extent by the application of in-built controlled vocabulary in the database. Users of databases state or express their need for information in natural language in different ways using different concepts and terms. In this context of expanding vocabulary, it becomes difficult and problematic to entertain and satisfy the needs of users for information without some kind of standardized terminology. Hence, the need for controlled vocabulary devices.

The use of vocabulary control devices is benefitial in two ways (Lancaster 1986: 7-8)):

1) To promote the consistent representation of subject matter by indexers and searchers thereby avoiding the dispersion of related materials. This is achieved through the control (merging) of synonymous and near synonymous expressions and by distinguishing among homographs;
2) To facilitate the conduct of a comprehensive search on some topic by linking together terms whose meanings are related paradigmatically and syntagmatically.

* Related term - similar but not equivalent to it; words which are often used in the same context;
* Equivalent terms - synonymous, or used synonymously in this context;
* 'Use' terms - which instructs to use an (often) equivalent word;
* 'Used for' terms - which is the converse of 'use' terms.

In addition, the context in which the term exists in the thesaurus is explained by the help of Scope Note.

Salton (cited in Townley and Gee, 1980: 21) enumerate five principles of thesaurus construction:

1) No very rare concept should be included in the thesaurus since they cannot be expected to produce too many matches between documents and search requests;
2) Very common high frequency terms should also be excluded from the thesaurus since they produce too many matches for effective retrieval. Individual high frequency terms may be replaced by more specific compound or hyphenated terms;

3) Non-significant words should be studied carefully before they are included in the text of words to be eliminated;
4) Ambiguous terms should be included only for the sense that are likely to be present in the document collections to be treated;
5) Each concept class should include only terms of roughly equivalent frequency so that the matching characteristics are approximately the same for each term within a category.

As is true in all fields of studies, change is an inherent characteristic of thesaurus. That is the structure of thesaurus has been changing from time to time which gave rise to many types of thesauri used around the world. Some of these are enumerated by Guinchat and Menou (1983: 103-106), Foskett (1982: 488-502), and Aichison (1970: 187-193) as important vocabulary control tools:

5.3.1.1 Thesaurofacet

As the name implies, Thesaurofacet is an integration of of faceted classification scheme and thesaurus. It is the result of the need to use computer techniques and post coordinate indexing which necessitated the development of

3) Non-significant words should be studied carefully before they are included in the text of words to be eliminated;
4) Ambiguous terms should be included only for the sense that are likely to be present in the document collections to be treated;
5) Each concept class should include only terms of roughly equivalent frequency so that the matching characteristics are approximately the same for each term within a category.

As is true in all fields of studies, change is an inherent characteristic of thesaurus. That is the structure of thesaurus has been changing from time to time which gave rise to many types of thesauri used around the world. Some of these are enumerated by Guinchat and Menou (1983: 103-106), Foskett (1982: 488-502), and Aichison (1970: 187-193) as important vocabulary control tools:

5.3.1.1 Thesaurofacet

As the name implies, Thesaurofacet is an integration of of faceted classification scheme and thesaurus. It is the result of the need to use computer techniques and post coordinate indexing which necessitated the development of

a classification scheme with a thesaurus. The two tools in the thesaurofacet are complementatry rather than parallel.

Foskett points out that there are several points of interest in a thesaurofacet, that a classification scheme is no longer synthetic for reasons which are relevant to analytico-synthetic classification generally (Foskett, 1982: 488). This thesaurus exploits the advantages of classical thesaurus and the hierarchical classification schemes. A classification scheme can only display one set of genus/species division. In a thesaurofacet other relations are shown in the thesaurus using the technique of BT, NT and RT cross refernces.

Thesaurofacet allows the provision of fundamental and specific facets of a subject field. Terms within these facets are arranged in hierarchies and arrays. Related terms in the structure are given for those relations that are not hierarchical.

Thesaurofacet is also a useful tool at the search stage in natural language and free text systems as an aid in program compilation (Aitchison, 1970: 192). It scores over the conventional faceted systems for this purpose. Because of the systematic approach, relationships are

displayed more clearly and precisely than the subject category lists given by conventional facet systems.

5.3.1.2 Root Thesaurus

A brief account of Root thesaurus is given by Foskett (1986: 497-502). Root thesaurus is a systematic list of terms accompanied by alphatical display. There is no fundamental difference between thesaurofacet and Root thesaurus as far as structure is concerned. There is, however, one important new feature in Root thesaurus, it is intended to be computer-based multi-lingual system, so that a new set of indicators are to be used in place of BT, NT, RT, which are language dependent.

The scheme is intended for thesaurus and the notation is not intended to be used in retrieval systems, where the alphabetic terms would be used; it does, however, make possible the main systematic sequence and facilitates translation into other languages by simply substituting the appropriate terms for their English equivalents.

The convention is to use abbreviation of English terms to indicate term relationships in thesaurus. However, since Root is a multilingual thesaurus a new set of conventions to indicate hierarchies and relationships has had to be developed. This is to use mathematical symbols mainly

which have international significance and are genuine international metalanguage.

Root has an up-dating facility. It also permits the interlingual concordance that can be produced once the necesary terms are in the database, and for the preparation of a specialized thesaurus. The user who wants more detail than is available can expand the appropriate part, discard those irrelevant, and obtain a custom-built product from the computer held databases.

5.3.1.3 Macrothesaurus

By giving a wider coverage of a certain discipline, this thesaurus merges many fields of study in an integrated whole. Macrothesaurus can be applied for several purposes. It is best used as the language for information processing. In other circumstances, it is used as a basis on which organizations build up their own vocabulary. It can also be used in integrating databases set up else where. It can further be applied as a multilingual terminological bank in documentation centres and translation services.

The application of macrothesaurus is more appropriate to organizations which deal with many disciplines. It is

also useful in institutions, whether international, regional, or national, which process or handle information from a varieties of sources and fields.

5.3.1.4 Classaurus

It is a vocabulary control tool which capitalizes on the advantage of a thesaurus and faceted classification. G. Battacharyya (1992: 24) defines this particular type of thesaurus as:

> " ... a faceted systematic scheme of hierarchical classification having all the necessary features of the conventional thesaurus excluding the enumeration of the so-called other related terms. It is supported by an alphabetical index."

This faceted thesaurus has some advantages over the conventional enumerative retrieval thesaurus.

- It brings all information concerning a particular subject or discipline in one place;
- it lists ready made subject headings,
- It can be used to construct other conventional retrieval thesauri. The most single advantage and distinct feature of classaurus is that it is readily and easily convertible to other conventional retrieval thesaurus by the

technique of picking and fixing subject headings.

Classaurus, as a typical example of faceted thesaurus, is based on the principles of analytico-synthetic approach. The classaurus classifies the substantive terms in the expressive subject propositions into their respective facet categories, that is, Discipline (=D), Entity (=E), Property (=P), and Action (=A). In these, D and E are regarded as primary categories; and P and A are regarded as secondary categories. Besides, it recognizes "modifiers" (m) to each category term if they exist; and they always go with their respective category term. Any one of the categories may be regarded as a base (=B) to bring everything on that together. The term denotative of environments, places, period (time), and physical forms are regarded as "common modifiers" (=cm). In addition, it provides a schedule of "other related subject-terms" which do not directly belong to the recognized base.

5.3.1.5 Multi-Lingual Thesaurus

A type of thesaurus that incorporates equivalent descriptors from different languages to facilitate data exchange.

5.3.1.6 Syntactic Documentary Languages

This type of thesaurus creates relationships between descriptors using symbols and sentences.

5.3.1.7 Uniterm

It is a monoterm thesaurus geared towards the reduction of descriptors drastically.

5.3.2 Scheme of Classifications

Classification schemes are categorized into two major groups; namely, enumerative scheme of classification and faceted scheme of classification.

5.3.2.1 Enumerative scheme of Classification

This scheme of subject classification groups subjects together on the basis of their relatedness or their being wanted together. Subjects, in general, are arranged in a helpful manner hierarchically with numeric or alpha-numeric descriptors assigned to code numbers. Good examples of this schemes are Dewey Decimal Classification (DDC) and the Library of Congress Classification (LCC) schemes. These schemes are generally rigid and difficult to expand as needed.

5.3.2.2 Faceted Scheme of Classification

Faceted classification is basically an attempt to act upon the realization that most subjects are compounds made up of two or more basic elements (Maltby,1968: 33). This scheme divides the subject field or each subject field into facets or categories where each facet consists of the element produced by a single characteristic of division.

Faceted classification schemes are based on the analytico-synthetic method of knowledge classification guided by principles and postulates which are helpful in a co-extensive representation of a subject. In the words of Guinchat and Menou, faceted classifications 'arrange concepts and objects into classes but they are multi-dimensional, i.e. the items are ordered within each class from different points of view' (Guinchat and Menou, 1983: 100).

A faceted classification requires the piecing together of appropriate elements from different but appropriate schedules; no attempt is made to provide ready made places for complete subjects. The scheme provides elementary terms from which the complex subjects are assembled (Foskett, 1964: 47). Faceted classification is

exemplified by the Colon Classification (CC) and, to some extent, by the Universal Decimal Classification (UDC).

5.4 APPLICATTIONS TO THE STUDY

Vocabulary control tools can be used to design an information retrieval system and to enhance retrieval of information from databases. As may be seen from the presentations in the foregoing sections of this chapter, the use of these tools has been largely in relation to bibliographic databases. There are other types of databases, such as object-oriented specialized databases, and integrated databases consisting of bibliographic and other types of records to which the tools may be applied. Particularly, this is facilitated by the use of information technologies in the development and application these tools, especially thesaurus. Therefore, in the following chapter, the various thesaurus software will be examined with respect to their application in generating vocabulary control tools and designing specialized local databases and user interface systems and also in enhancing information retrieval capabilities from the different types of databases.

CHAPTER SIX

THESAURUS SOFTWARE AND APPLICATIONS

The revolution in microelectronics, referred to as new information technology (NIT) continues to have its impact in the field of information storage and retrieval, its tools and techniques. One of the areas thus influenced by the rapid developments in NIT is vocabulary control tools particularly thesauri design, development and use.

Since the 1980's, there has been a renewal of interest in thesaurus as a tool in database design and retrieval enhancement as reflected both in the literature on the subject and the growth in the number of thesauri.

Schondorf gives several reasons for the renewal of interest in the application of thesauri (Ganzmann, 1990: 148). Among these reasons are the impact of the work on artificial intelligence (AI), the concept of knowledge representation, dissatisfaction with the result of free text retrieval and advances in computer technology which has brought about rapid decrease in cost, easier handling, and better performance both in respect of hardware and software. This advance, especially the advent of microcomputer, has greatly facilitated the tedious task of development, management and application of thesaurus both for database design and retrieval improvement.

The various thesaurus software are developed to carry out some common tasks.

6.1 TASKS OF THESAURUS SOFTWARE

Thesaurus programs have the following specific functions to fulfil (Ganzmann, 1990: 148-149):

6.1.1 Thesaurus Construction and Maintenance

Construction and maintenance/upgrading of thesaurus include a variety of tasks and activities. Word material is selected and recorded, some times part of existing thesaurus must be integrated, information as to the source of terms, language systematic grouping has to be recorded. Terms are controlled in respect of classification of meaning (homonyms, definitions) and relations must be defined between terms. Finally, thesaurus must be updated regularly with regard to relations and terminology usually arising from changes in the terminology of a given field, changing indexing and retrieval practices and failing adequacy of the vocabulary's scope or specificity.

6.1.2 Output of Thesaurus

The output of a thesaurus (or part of it) is necessary in all phases of thesaurus work, either on the screen, by the printer or also into a file in case the data are to be transferred to a word processor or for integration purposes to another system (thesaurus maintenance or retrieval system).

Most thesauri have at least one alphabetical and one systematic section, often KWIC/KWOC or hierarchical displays; sometimes graphical displays are added. The representation of three kinds of relations (equivalent, hierarchical and associative relationships) in accordance with the respective standards on thesauri and their construction has also become a common trait of most thesauri.

6.1.3 Indexing and Retrieval with Thesaurus

All thesaurus programs support the construction, maintenance and printing of thesaurus. It is possible to index document and search for information in a retrieval program with the use of printed thesaurus. The integrated thesaurus can support specific tasks associated with indexing and

retrieval, by acting as an interactive instrument for orientation about potential index and search terms, a tool for consistency control regarding the data used in indexing, searching and a tool supporting update routines.

6.1.4 Exchange, Integration and Compatibility of Vocabularies

These features are also being increasingly important in relation to thesaurus software. It entails a variety of specific tasks and functions about:

- batch input and output of computer readable data in a suitable format,
- special type of relations for multilingual thesauri (either with a dominant or equal languages) or the connection of various thesauri; and,
- use of compatible vocabularies in indexing and searching.

The tasks enumerated above are the more general and comprehensive functions performed by the various thesaurus software. In the following sections, some of

the thesaurus software packages are discussed. The limitation of the discussion to the ones presented below is due to inavailability of information on other packages.

6.2 Thesaurus Packages: An Overview

The previous section (6.1) mentioned some of the thesaurus software PROTERM, INDEX, TMS, THES, MTHES, THES1, THES2, CICADE, LIDO, BASIS, DOMESTIC. Of these TMS, PROTERM, INDEX, THES, THES1, THES2, and MTHES will be discussed here. THES, MTHES, THES1, THES2 are given more emphasis since they are used in this work. Presentation on the first three packages is based on the comparative study made by Claus Ritzler (Ritzler, 1990: 138-147). The text and the various tables and figures given in that paper are used generously.

6.2.1 TMS, PROTERM and INDEX

6.2.1.1 The Software Packages and Their Producers

In this section, three thesaurus softwares are presented: one British (TMS) and the other two PROTERM and INDEX), German programs.

6.2.1.2 Software and Hardware Requirements

PROTERM-T and TMS run under MS-DOS version 2.1 and higher. INDEX requires MS-DOS version 3.3 or another operating system such as PC-DOS, UNIX/XENIX, OS/2 and others. All the three programs can be installed on a PC-XT/AT, with a hard disk and 512 KB memory. For INDEX, the producer recommends an Enhanced Graphics Adaptor with suitable display. TMS requires one disk drive to be free for the program disk, because it is not possible to copy the program on the hard disk. Thesaurus data, however, are saved on the hard disk. TMS also needs a printer, becase it is not possible to display all kinds of views on the monitor.

6.2.1.3 User-System Interfaces

All the programs are presented with menus. The first menus the user gets are similar in all the three packages. The difference is that one presents more detail, than the other.

In Figure 6.1, with the main menu of TMS there are two parts: Construction and Review. On the construction side, we see the input components like the option to enter terms, relations and scope notes, the option to

change or delete terms, the option to add facets or even the option to re-index what is normally management of system functions. On the other side, the user will see more or less all out put functions, such as, viewing thesaurus, listing terms at different levels, creating language index and changing print parameters.

Table 6.1: Software Producer Information

1. SW Producer Information		INDEX Version 4.1	PROTERM Version 2.5	TMS Version 1.0
AwA	SW Producer Address	ERNST LUKUS Leipziger Ring 16 6054 Rodgau 3	PROGRIS Auguste-Viktoria-Str 1000 Berlin 33	PYRAMID SW Product ltd 9 CHURCH St Reading RGI UK
AwA	Status of the producer	Subcontractor	Establ. Co.	Estbl. Co.
AwA	Company Foundation	1984	1977	1982
AwA	Customer Service	Consultation Course of Instruction	Consultation	Consultation Course of Instruction
AwA	Number of References	60 Single-user Installations	24 Instal-lations	40 Installation 5 multi-user

AwA Signifies: Generally important product information. These represent the necessary requirement of a given product.

Source: Retzler, 1990

Table 6.2: Hardware and Software Requirements

2. HW and SW requirements	INDEX Ver. 4.1	PROTERM Ver. 2.5	TMS Ver. 1.0
AwA HW Basic Equipment	PC-XT/AT IBM Comp. PS/2-Ser. & others 512 KB Mem. Hard Disk Enhanced Graphics Adaptor with Suitable Display recom	PC-XT/AT IBM Comp. 512 KB Mem Hard Disk	PC-XT/AT IBM Comp. 512 KB Mem Hard Disk Min. 1 Disk drive, Print
AwA Operating System	MS-DOS Version 3.3	MS-DOS 2 upwards	MS-DOS 2.1 Upwards
AwA Other Operating Systems	e.g PC-DOS UNIX/XENIX OS/2 & others	No	NO
AwA Standard SW Basic Eqpt	INDEX Dataflex RTL Dataflex DL	PROTERM-	TMS
AwA No. of Add'l Sta. Module	8 Modules	5 Modules	No
AwA Cost of the Single User Basic Version	DM 6,200	DM 1,200	DM 1840
AwA Cost of Licenses for other SW	DM 300-2000	-	-
AwA Cost of Additional Standard Module	DM 150-3,320	DM 190-38	-
* Capable for multiple places	Yes	No	No
* Capable for network	Yes	No	No
AwA Development System Use	Dataflex	Clipper	Turbo Pascal
AwA Data Structure	Semantic Network Model	Simulated Relational	Relational with pure Hir. Str. element
* Source code Available	Master of Negotiation	No	No
AwA Quantity of Delivery	3 5.25" Diskette 1.2 MB 230pp manual Basic Set of test data	2 5.25" Diskette 360 KB 100pp manual	2 5.25" Diskette 360 KB 27pp Operating Guide
AwA Documentation and Literature - Quality - Language	Manual V.Good German	User manual Good German	User Introduction Satisfactory English

AwA Signifies: Generally important product information. These represent the necessary requirement of a given product.

Fig. 6.2: The main Menu of INDEX

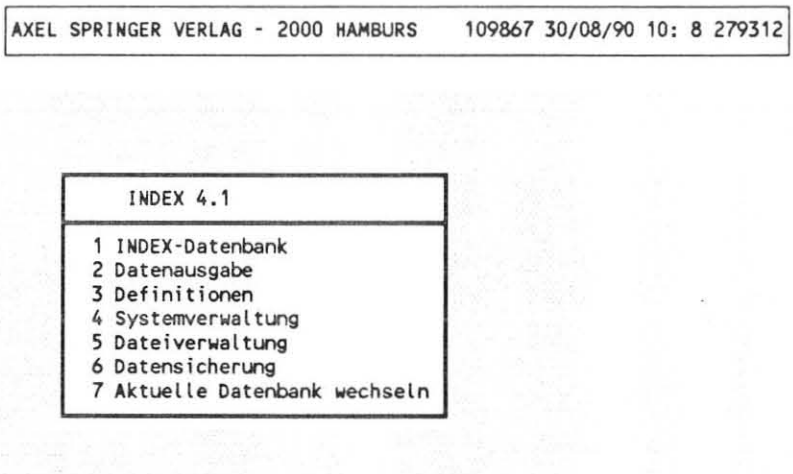


Fig. 6.3: The main Menu of PROTERM

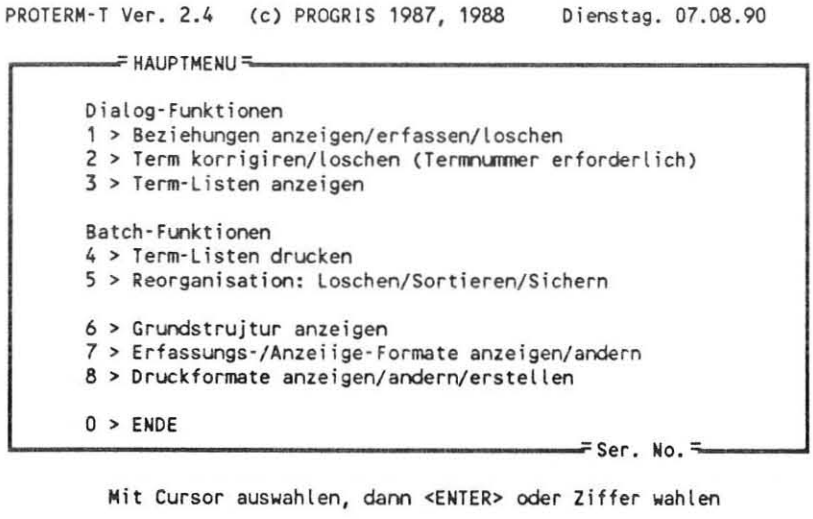
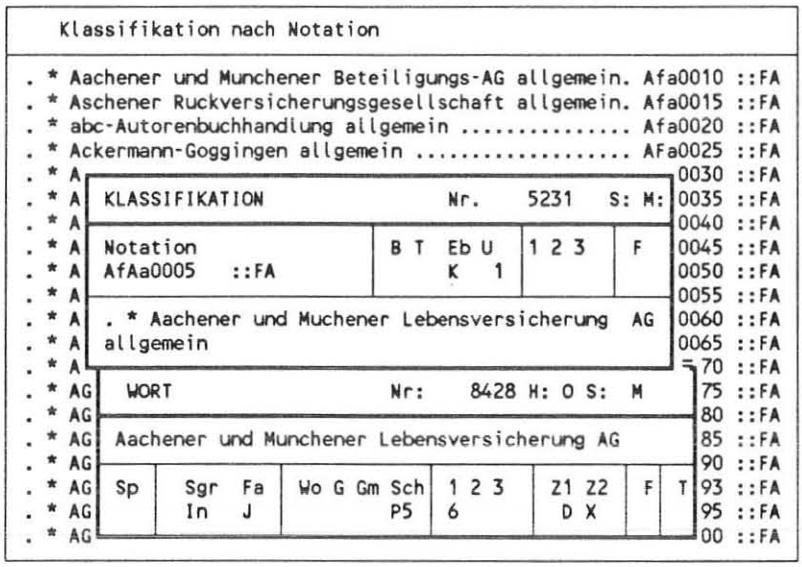


Fig. 6.4: The object in INDEX; forms for words and classifications and a list in the back.



-> Object Zoom Verbinde Liste Suche Form Mark Hilfe Weiter Ende

In figure 3, with the main menu of PROTERM, three areas are seen. First the dialog function to enter, delete, and list terms, than the batch functions for reorganization and output. The last bit contains the different management functions.

However, all the three thesaurus programs are presented in different ways to the user. INDEX works with different types of objects. First there are forms, those forms are structured into different fields and look like record cards. And there are lists (see fig.4).

INDEX provides all kinds of interactive possibilities; the screen layout at the different working levels is fully coloured. In spite of the complexity and different functions, INDEX is completely driven by using the ten function keys F1 to F10 and the cursor keys. Additionally, the user gets a menu line on the upper part of the screen by using function key F10. This main menu line with various submenus is partly equivalent to the function keys. At some stage, especially at the systems management level, the user is presented with pull down menus. All these interactive possibilities and the different usage of cursors are unified at every stage of the program. Each colour has its own meaning, aiding the user in learning and understanding the software and its working process.

PROTERM also has menus at different working levels. The options for each level are numbered upwards. The monochrome display of PROTERM and TMS operate in conjunction with different cursor keys or by entering a number or by entering different key combinations.

The user can choose different options from each menu. The interactive possibilities are unified in PROTERM. If the user wishes to jump to another menu, he must pass through all intermediate menus sequentially. It is not possible to jump between different levels as it is in INDEX

because the function keys in PROTERM are not used at all and only for statistical operations in TMS; the operation is cumbersome.

6.2.1.4 Input of Terms

The data entry mode in TMS is accessed by typing the number 10. In PROTERM and INDEX, it is selected by choosing option 1 from the main menu. The user is presented with a new display layout which is different for each program.

In TMS, the term may be entered by typing the specific term or by typing the number code for it if it is known. Terms are limited to a maximum of 40 alphanumeric characters.

The word can be described with a scope note or a facet. To register a scope note, the word has to be typed in again after pressing the function key F3. A field with maximum 1110 character space appears. To locate a facet, the user has to go back to the main menu to activate option 12, making operations inconvenient. The facet feature is primarily intended to relate to a classification scheme. However, as yet no facilities exist for sorting and printing the thesaurus by the facet.

In PROTERM, data entry, data display, data correction and data deletion are all available at one level, but it does not include term correction or term deletion. This is included in option 2 of the main menu. Each term can be 60 characters long. There is possibility to enter a scope note with a maximum of 150 characters and a notation with 6 characters. The program checks each term simultaneously at the line of data entry. A small display shows a changing term with the same sequence of characters as inserted. The program also displays a list of terms by pressing function key F2. The user can choose words from the list.

Fig. 6.5: Data entry mode in TMS

```
 Pyramid Software   0734  595633   *  *  *  *  *     10/8/1990
                                                F4  F5 F6 F7 F8 F9
R1  No.   Term                                  US  UF BT NT RT SA
     53  Volkswagen                              2   0  0  0  1  0
SN
US   54 VW                                       0   1  0  0  0  0
US   55 VW Kafer                                 0   1  0  0  0  0
RT   56 Audi                                     0   0  0  0  1  0
```

hatchemark (#) and some carriage returns. This is again a very time-consuming procedure for users who want to enter lists of terms before they start with the relations. Especially, at the beginning stage of the thesaurus project, where the word material will be collected first. It is not possible to enter short terms that are identical with the first characters of already existing terms. In this case, it needs to enter the shorter terms first or to enter them with blank space at the end of the shorter term.

Homonyms and other word with identical character sequence cannot be entered in PROTERM and TMS. In INDEX, a homonymous term can be marked in a special field with an H. In the search mode of PROTERM, no distinction is made between upper and lower case which can be very helpful.

In INDEX, a blue formula is presented to enter words which represents a 3 by 5 card from card index (Fig. 4). The formula is used for recording searching and relating the data. There are different formulae for words, classification and sources. In the formula for words, a new term is entered and in the different other 16 smaller fields specific information about the term may be added. Special information can be a language code, a subject code, a facet and others. In this case, the user is

completely free in his/her decision; and can define the categories himself. Some of the fields are indexed that is a very helpful feature for sorting and listing the material for various purposes.

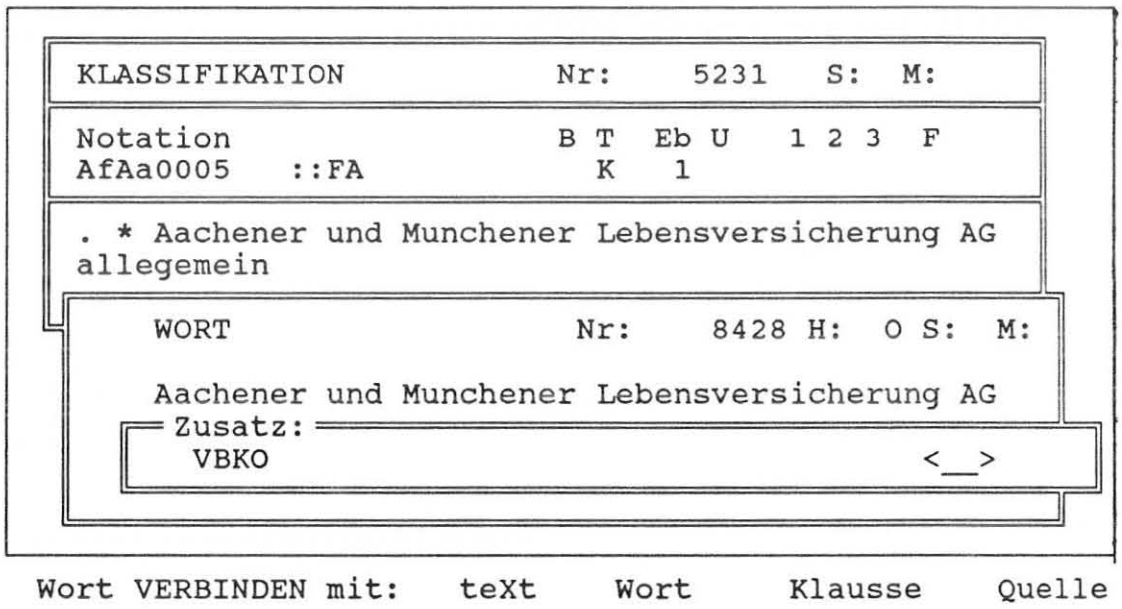
Table 6.3: the different kinds of relationships in INDEX, PROTERM and TMS

INDEX 4.1			PROTERM 2.5	TMS 1.0
1. Synonymy	Descriptor	Non-descriptor	x	x
2. Abbreviation	Full descriptor	Abbreviation		
3. Identical Use	Term	Same meaning		x
4. Quaisynonymy	Descriptor	Quasi synonym		
5. Alternative	Descriptor(Alt)	Same meaning		
6. Combination	Single descr.	Combined term	x	
7. Use	Basic Term	Example of Use		
8. Hier. Relation	Broader Term	Narrower term	x	x
9. Generic relation	Broader Term	narrower Term		
10. Partitive Relation	Whole	Part		
11. Casualty	Cause	Effect		
12. Appurtenance Rel.	Generic Term	Instance		
13. Field Relation	Field	Element		
14. Associative Rel.	Term	Related Term	x	x
15. Temporal Sequence	Predecessor	Successor		
16. Splitting	Predecessor	Successor		
17. Union	Predecessor	Successor		
18. Opposite	Term	Opposite Term		
19. Use Warning	Term	In contrast to		
20. Production	Producer	Product		
21. Rel. of Material	Material	Object		
22. Other languages	Language	Foreign Language		
23. Translation	Source Language	Target Language		
24. Systematic	Descriptor	Natation one notation/facet per term only	x?	x?

This is possible in PROTERM only by combining the term with a notation. But each notation may have only one combination with a term. Thus, if one wants to add more specific information, PROGRIS should be contacted to provide an additional module for that purpose.

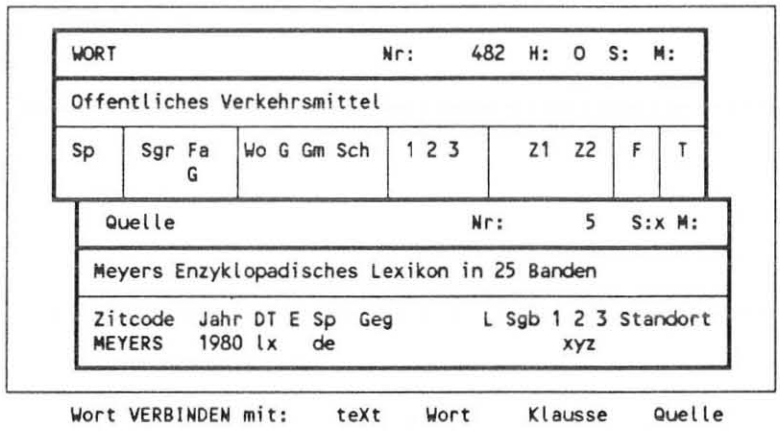
Fig. 6.8: Building a relationship in two terms in INDEX

Fig. 6.9: Building a relation between a classif. and a term i
INDEX



Wort VERBINDEN mit: teXt Wort Klausse Quelle

Fig. 6.10: Building a relation between a source and a term in INDEX



Recorded with each word saved is the entering data and the date of the last change together with the name or code of the user who did the entry or change. The maximum length of the controlled term is 50 characters even though the term itself can be as long as desired. That means, if one has two terms, which differ at character position 51 or beyond, INDEX will not accept it. The user has various possibilities to search, correct and delete terms. This can also be done in a list or in the formula. It is not a problem to switch between a list and a formula. In both types of objects, one is able to browse forward and backwards in the term material to make further selections. The user can also make words in a list for creating a separate working list. In INDEX Scope Notes have no limit to their length. To enter a Scope

note, a text field has to be opened first. This mode is supported with word processing features. They enable deletion and insertion of lines, search for a word or a phrase and more. The user has the possibility to define different kinds of notes (definitions, history note, etc.).

6.2.1.5 Relations and Consistency Control

INDEX provides the user with the choice between 24 different relationships. In this package, it is possible for the user to create and define more relationships if required. TMS and PROTERM provide 5 different relationships. In all the three programs, the minimum standard relations: synonym, hierarchical and associative relations are included.

All relations have been designed to conform to British, German and international standards. The "see also" relation in TMS is identical with the relationship "identical use" in INDEX. It is possible to define more relations in INDEX, but not in PROTERM or TMS. PROGRIS provides additional modules for additional relationships.

Moreover, all or part of the control routines may be switched off in INDEX as if you wanted to ignore special

routines. INDEX has completely user controlled consistency checks. In this part, the user is free in his/her decisions, but care must be taken when to change the default values. INDEX provides more control routines than the other programs because different descriptions of a term can be controlled as well.

In all the three programs, the reverse relation is assigned automatically after saving a relation between two terms, but the control functions are not as strict as in INDEX, especially, in TMS. For instance, once assigned, terms as non-descriptors can be related to all other terms, just like a descriptor.

In PROTERM, the consistency checks are done perfectly. If more than one relation is needed to be entered to a non-descriptor, the term has to be assigned as polysem first; otherwise, it is not possible.

Automatic reorganization of very complex relations after changing a descriptor to a non-descriptor will only be done in INDEX which is very convinient. In the other two programs, one has to delete the relations first before assigning the non-descriptor to a descriptor, then after, the relations have to be built up again. This can be very time-consuming, especially if the descriptor had a number

of relations on various levels. Unlimited number of relations to a descriptor are provided by all the three programs.

6.2.1.6 Entering the Relationships

In PROTERM, as already mentioned, each kind of relationship is obtained after having entered the term. The user can affect the order by typing the menus sign (to get the previous type of relationship) or by typing an additional carriage return (CR), then getting the next kind of relationship. The sequence and the abbreviation of the relationship can be changed by the user under the management functions. In the same mode, relationships per term can be displayed and deleted. If all the relations of a term have to be deleted at once, the whole descriptor has to be deleted under menu point three and two.

In TMS, similar to PROTERM, to enter a new relationship the user has to type the abbreviation for the relationship followed by CR, followed by the term or its number. Just as many relationships may be entered as are required. While manipulating relationships of a certain type, it is convenient for the user to exclude other relationships from the screen during editing. This can be

there are no standards for data exchange.

6.2.1.9 Output Functions

There are highly developed output functions in all the three programs. Apart from TMS, all standardized lists are possible, for displaying on the screen, for saving as a file or for the output on the printer. In addition, PROTERM and INDEX provide various sorting parameters.

The different output lists are:
(I=INDEX; P=PROTERM; T= TMS)

- Alphabetical term list	I	P	T
- Systematic term list	I	P	
- Minithesauri/Parthesauri	I	P	
- Term lists with/without relationships	I	P	
- Terms selected by different criteria	I		
- Synonym lists	I	P	
- Systematic thesaurus	I	P	T
- Alphabetic thesaurus	I	P	T
- Thesaurus with different BT/NT levels			T
- Hierarchical BT/NT lists		P	
- Different classification lists	I		
- different source register lists	I		

6.2.2 THES, MTHES THES1 and THES2

Four CDS/ISIS pascal programs for the creation and maintenance of thesaurus are of particular interest to the present study. They are THES.PAS, MTHES.PAS, THES1.PAS and THES2.PAS programs. THES and MTHES programs are for handling monolingual (English language) thesaurus. THES1 and THES2 handle trilingual thesaurus supporting English, French and Spanish languages. They can also be used in monolingual mode.

6.2.2.1 The Nature of the Programs and Their Requirements

THES, MTHES, THES1 and THES2 programs are designed and developed using CDS/ISIS Pascal language. The programs are intended to interface with Micro-CDS/ISIS databases.

Micro-CDS/ISIS requires IBM PC compatible micro-computers with at least 512 RAM and an MS-DOS version 3.0 and above.

THES and MTHES are basically similar and are presented together. The other two (THES1 and THES2) differ from THES and MTHES in some aspects and thus presented separately.

6.2.2.2 Programs THES.PAS and MTHES.PAS

THES program is the first version and MTHES is a modified version of THES. Their main objective is to provide a facility for the management of monolingual (English) thesaurus. They are designed as a menu exit in Micro-CDS/ISIS. The operation of these programs assumes that a thesaurus database called THES exists. Both these programs are menu driven with submenus and prompts displayed in the message area at the bottom of the screen.

The database which the programs assume to exist is a set of ISIS records which may contain terms with their relations (broader term, BT; narrower term, NT; related term, RT; use for, UF and scope note, SN).

The FDT of the THES database is presented below.

Field Definition Table (FDT) Dataqbase: THES

?	Tag	Name	Len	Type	Rep	Delimiters/Pattern
	1	Term	30	X		
	2	Scope note	30	X	R	
	3	USE	30	X	R	
	4	UF	30	X	R	
	5	BT	30	X	R	
	6	NT	30	X	R	
	7	RT	30	X	R	

to exit from the program.

F - This option is used to move to the first term in the term list.

S - It selects a given term and displays the term with its relations.

A - This option is used for adding relations (BT, NT, RT, UF, SN) to a term.

C - Creates a new thesaurus term with its relations

D - Depending on the cursor position, it deletes a term or a relation; however, in order to delete a hierarchical term all its relations must be deleted first.

Q - Selects a term for searching. If more than one term is selected, they are ORed together.

? - This option is used to display the current set of search terms selected, if any. These are displayed in a box at the bottom of the screen.

X - Exits from the current operation; if search terms were selected, expression is displayed in edit mode, then executed provided a database was selected when THES was called.

The thesaurus maintenance function of the two programs allows addition of new terms and relations, changing terms, updating and deleting relations.

There are two menus for each of these programs: list menu and detail menu (see figures 6.11-6.12).

Fig. 11: THES Menus

a) List Menu

```
◂┘Next   B[previous]   P[age]   S[elect]
C[reate term]   T[erm select]   X[exit]
```

b) Detail Menu

```
◂┘Next B[ack] F[irst] P[age] S[elect] T[erm select] Q[uery]
?[display query] A[dd relation] D[elete] C[reate term] X[exit]
```

Fig. 6.12: MTHES Menus

a) List Menu

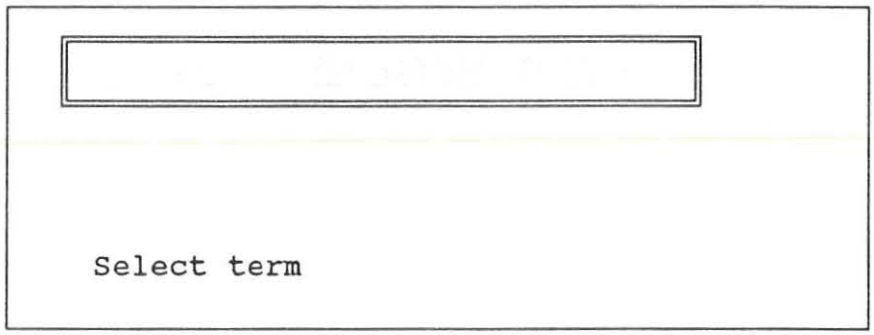
```
[C]reate  [P]age  [S]elect  [T]erm select  [M]ake file
 add te[R]m[file]                              e[X]it
```

b) Datail Menu

```
◂┘Next  [B]ack  [F]irst  [S]elect  [T]erm select  [M]ake file
  add te[R]M[file]  [C]reate term  [A])dd relation  [D]elete
 [Q]uery   sa[V]e   Query/saVe     search[/]add     e[X]it
```

The first screen layout in both THES and MTHES is similar (see figure 6.13).

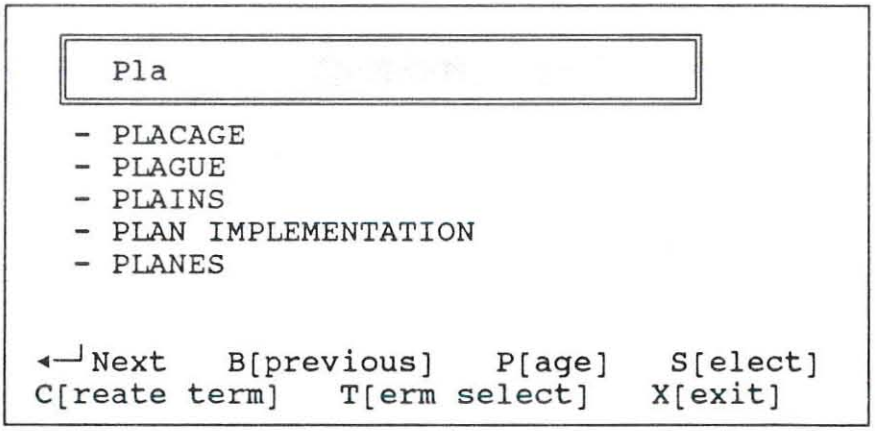
Fig. 6.13: The first Screen Layout of THES and MTHES



For example, if we enter a term stem (a string) "pla" the list will be as follows.

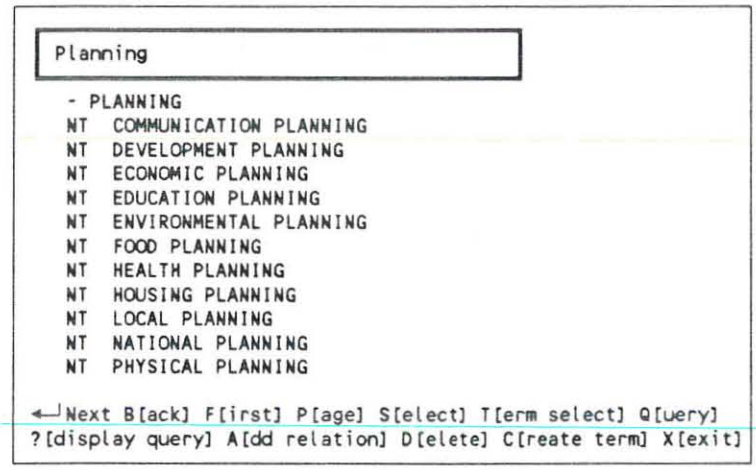
Fig. 6.14: Select mode screen layout in THES and MTHES program

a) Term select from the first screen (fig. 6.13)



If we select the term PLANNING and then press <ENTER>, the result is as given under (b) below. In this case the selected term is displayed with all its relations.

b) Term select screen layout from the alphabetical term list given in the above screen



Adding Terms and Relations

New terms can be added by typing in '**C**' on any line of both term list and detail menus. A prompt is displayed in the bottom area of the screen which requests for entry of new term. A term validity is automatically checked during this process before the term file is updated. THES and MTHES also support additions of term relations respective to BT, NT, RT, UF, USE and SN. When a relation is created, the screen is cleared and it prompts to the type of relation. The relation and the terms are filled one after the other (see fig. 6.16).

Fig. 6.15: Term create mode in THES and MTHES

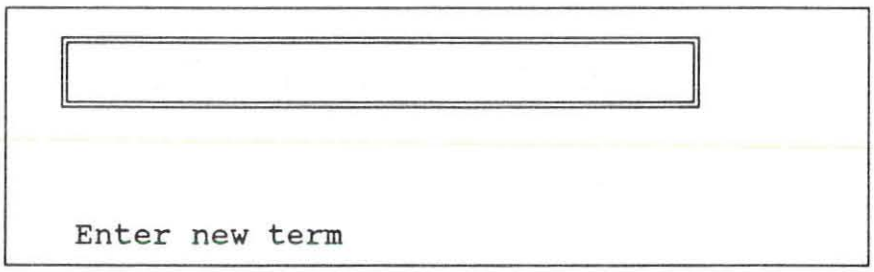
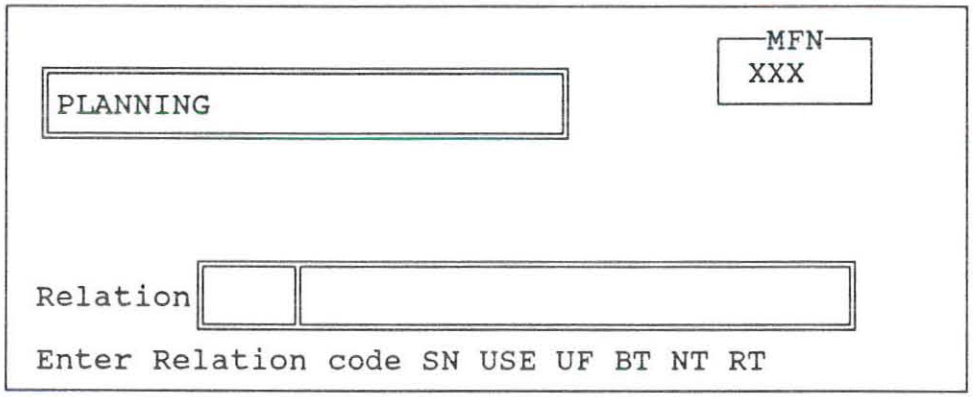
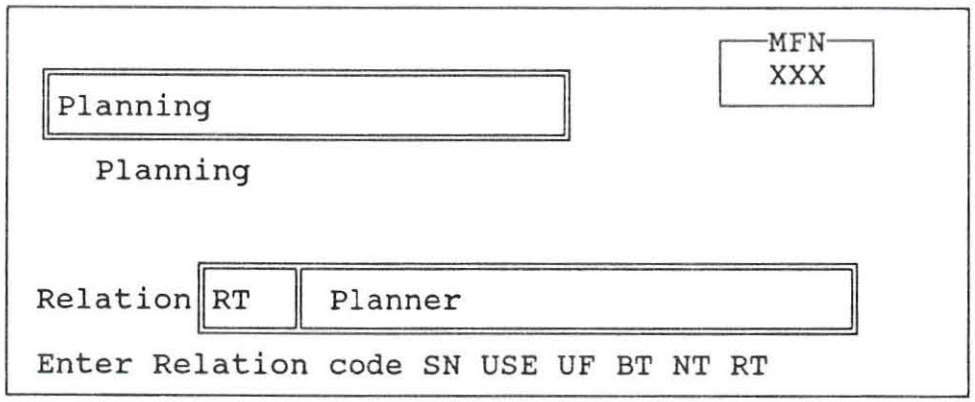
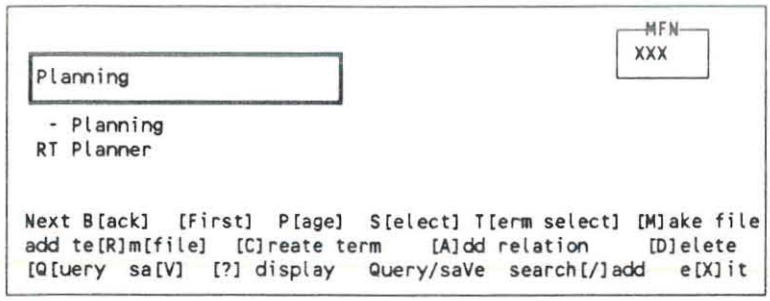


Fig. 6.16: Relation Creation mode in THES and MTHES







Deleting Terms and Relations

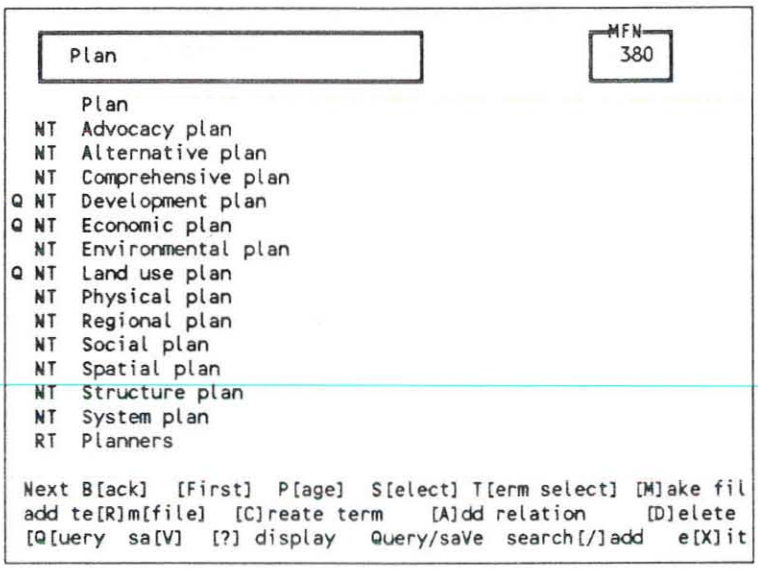
In order to delete terms, first all relations of the terms, if any, should be deleted. Any relation can be deleted by pressing option **'D'** in front of the corresponding line. Reciprocal relations are also deleted if applicable directly by the deletion of the relation.

Searching

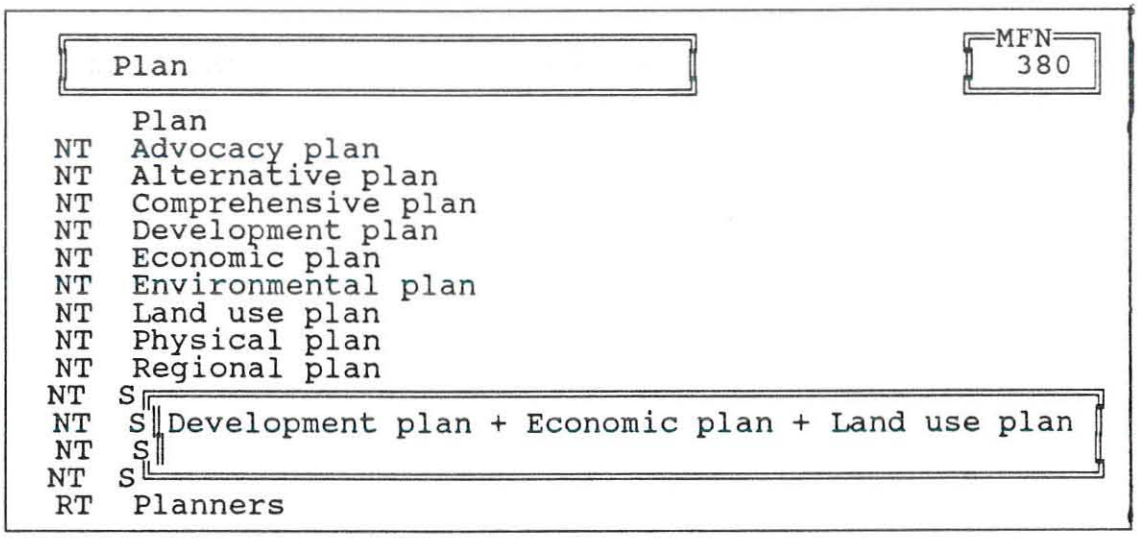
THES and MTHES allow search in a database by the use of terms in the thesaurus. The term of interest is selected by the use of option **'L'** or **'T'**. Option **S** is used to select a term with its relations. In the process of searching, pressing **'Q'** will select term(s) for searching in database(s) and **'?'** will display currently selected term(s) and **'X'** will exit from select/browse mode. Then the search expressions built-up from the selected terms are edited, if necessary, before used to search

information from database(s) (see fig. 6.17)

Fig. 6.17: Query formulation mode in THES and MTHES



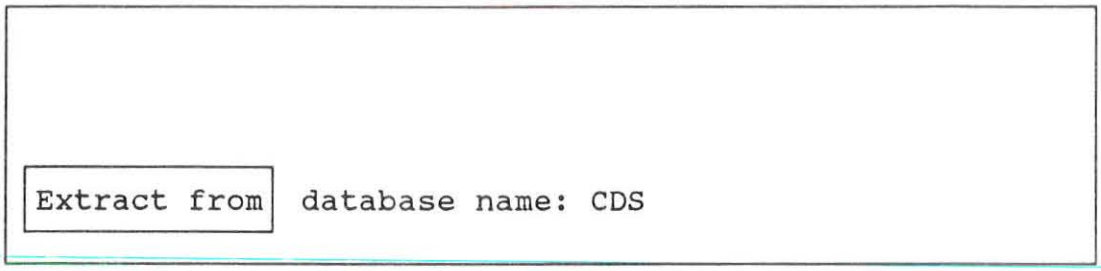
Selecting [?] will display the query as follows:



Press any key to continue

Fig. 6.18: Make file mode screen layout in MTHES

a) Asking for database from which file is to be made

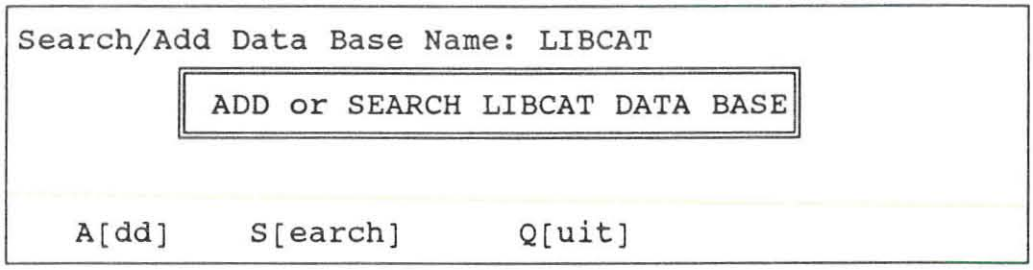


b) After entering the name of a database

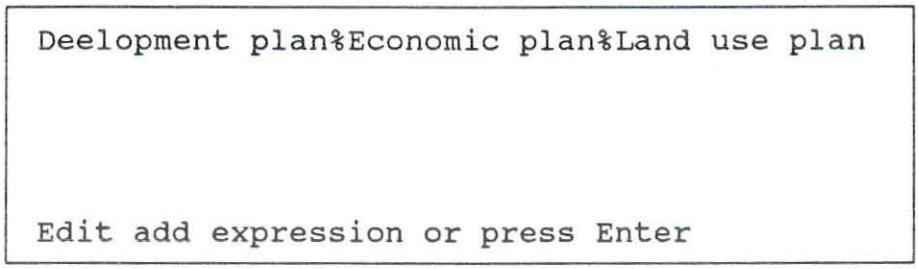
The program asks for the tag of a particular field from which we want to extract term(s).

The 'make file' option is meant to create another term file other than the existing one. The add te[**R**]m[file], as its name implies, is used to add a term or a file. Option sa[**V**]e is used to save queries and search terms, and search[**/**]add makes searches from a database using selected terms or adds selected terms to a database (see figure 6.20).

Fig. 6.19: Search/add option screen mode in MTHES



If option A[dd] is selected, for example, assuming our previous query, the result will be as follows:



With these changes and modifications, the original THES program is enhanced to perform more functions than was posssible by the original program. Terms can be selected from the thesaurus and added to a particular field (e.g. keyword field) of a database and vice versa.

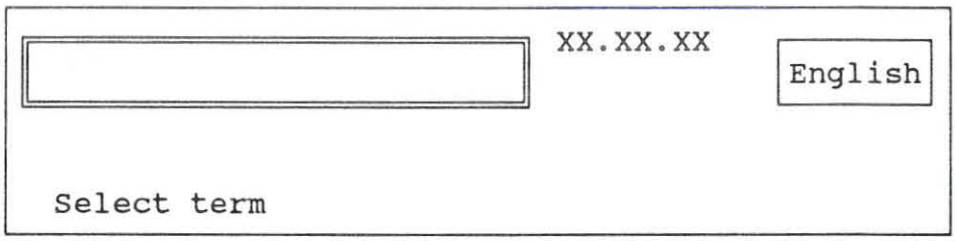
6.2.2.3 Trilingual thesaurus Package

This Package has been designed to enable thesaurus maintenance, trilingual and hierarchical search with the aid of thesaurus and keyword checking and translation. It has five components:

1) An ISIS database called THES that contains the Macrothesaurus (recently 3rd edition. 1985),
2) An ISISPAS program for thesaurus maintenance (THES1),
3) An ISISPAS program for trilingual/hierarchical searching using the thesaurus as a support (THES2),
4) Text file (EMTHES.ISO, FMTHES.ISO, SMTHES.ISO) containing a list of messages in English, French and Spanish, respectively.
5) An ISISPAS program translation and checking keyword.

All these programs are written as menu exit.

The first screen display in this sytem is the same with that of THES and MTHES:



Selecting a term "PLAN"in the above screen layout, the following screen is invoked.

```
- Plan                    08.14.06                English

NT  Advocacy plan
NT  Comprehensive plan
NT  Development plan
NT  Economic plan

P(age forward) S(elect) T(erm list) C(reate) A(dd relation) U(pdate)
 ←┘Next B(ack) F(irst0 P(age) L(ist) Z(Language) >(see SN) X(exit)
```

6.2.2.3.1 The Thesaurus Database

This database, similar to the database for THES and MTHES, is a set of ISIS records which may contain (i) a keyword in the three languages with respect to their relations (BT, NT, RT, UF), scope note (SN), facet number (optional) and date stamp, (ii) a forbidden term in one language respective to **USE** relation for that language, scope note, facet number and date stamp.

The FDT of the database THES is presented below.

Field Definition Table (FDT) Dataqbase: THES

?	Tag	Name	Len	Type	Rep	Del/Pat
	99	Facet number XX.XX.XX	10	X		
	100	Source of Term (code)	2	A		
	1	Term English	35	X		
	11	Term French	35	X		
	21	Term Spanish	35	X		
	2	Scope note English	400	X	R	
	12	Scope note French	400	X	R	
	22	Scope note Spanish	400	X	R	
	3	USE English	35	X	R	

Field Definition Table (FDT) Dataqbase: THES

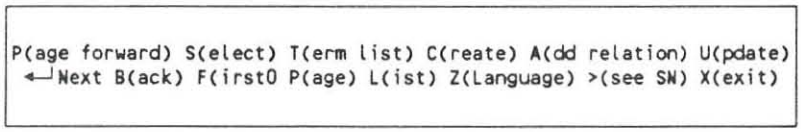
?	Tag	Name	Len	Type	Rep	Del/Pat
	13	USE French	35	X	R	
	23	USE Spanish	35	X	R	
	4	UF English	35	X	R	
	14	UF French	35	X	R	
	24	UF Spanish	35	X	R	
	5	BT English	35	X	R	
	15	BT French	35	X	R	
	25	BT Spanish	35	X	R	
	6	NT English	35	X	R	
	13	NT French	35	X	R	
	23	NT Spanish	35	X	R	
	7	RT English	35	X	R	
	17	RT French	35	X	R	
	27	RT Spanish	35	X	R	
	8	F term English	35	X	R	
	18	F Term French	35	X	R	
	28	F Term Spanish	35	X	R	
	97	Date stamp (Relations)	18	X		
	98	Date stamp (Terms)	18	X		
	80	Chapter English	100	X	R	
	81	Chapter French	100	X	R	
	82	Chapter Spanish	100	X	R	
	83	Sub Chapter English	150	X	R	
	84	Sub Chapter French	150	X	R	
	85	Sub Chapter Spanish	150	X	R	
	86	Paragrasph English	150	X		
	87	Paragrasph French	150	X		
	88	Paragrasph Spanish	150	X		
	89	Heading Record Code	2			AA

6.2.2.3.2 Thesaurus Maintenance Program

The program THES1 is used for this purpose which supports the following:

- creation of keywords in all the three languages (omitting one of them is not possible),
- creating forbidden terms in one language at a time, no translation is supported,

Fig. 6.22: THES1 Maintenance Detail Menu



On both menus, typing **Z** requests a language submenu in the message area to which the response may be **E** for English, **F** for French, and **S** for Spanish.

Fig. 6.23: THES1 and THES2 Options in ISIS Retrieval Service Menu

Service ISISRET Information Retrieval Service Menu EXGEN

L- Change Dialog Language
B- Browse Master File
T- Display Terms Dictionary
S- Search Formulation
D- Display Search Result
G- Execute Previous Search
F- Select Display format
N- Modify Display Format
R- Recall Query formulation
P- Save Search Results
K- Searching via Macrothesaurus
M- Macrothesaurus Maintenance
C- Checking/Translation Key Words

X- Exit

? [ ]

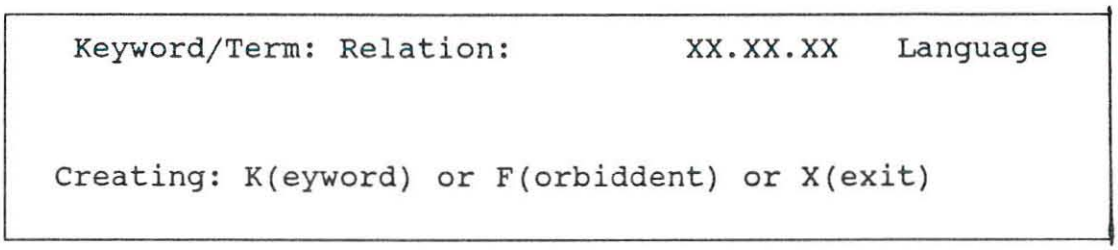
Database _____ Worksheet _____
ax MFN _____ Format _____

Adding Terms

To add new terms option **C** can be keyed in on any line of both menus. A submenu is displayed in the message area requesting the class of the term, i.e. whether it is a key word or a forbidden term. When a keyword is added, the screen is cleared and prompts are given for facet number (optional), and keyword to be entered in all the three languages. At this stage, term validity is checked automatically before the terms are updated.

Creating (adding) a forbidden term requires selection of one language in which the term will be keyed in. The screen is cleared and it prompts to facet number (optional) and the forbidden term. Validity check is performed to make sure that the term does not exist in the term file in the selected language.

Figure 6.24: Create Mode Screen Layout in THES1



Adding Terms

To add new terms option **C** can be keyed in on any line of both menus. A submenu is displayed in the message area requesting the class of the term, i.e. whether it is a key word or a forbidden term. When a keyword is added, the screen is cleared and prompts are given for facet number (optional), and keyword to be entered in all the three languages. At this stage, term validity is checked automatically before the terms are updated.

Creating (adding) a forbidden term requires selection of one language in which the term will be keyed in. The screen is cleared and it prompts to facet number (optional) and the forbidden term. Validity check is performed to make sure that the term does not exist in the term file in the selected language.

Figure 6.24: Create Mode Screen Layout in THES1

```
Keyword/Term: Relation:          XX.XX.XX   Language


Creating: K(eyword) or F(orbiddent) or X(exit)
```

When a term, e.g. Population is entered, the program will display as follows asking for the entry of the term in the other two languages.

```
Keyword/Term: Relation:          XX.XX.XX   Language
Population
            ------------
                         ----------------

Creating: K(eyword) or F(orbiddent) or X(exit)
```

Adding Relations

The A[dd relation] option of the detail menu allows to create relations, i.e. BT, NT, RT, SN. When a relation is created for one language , it is also created automatically for other languages, and reciprocal relation is created when relevant (BT/NT, RT/RT); however, the reciprocal for USE FOR (UF) is created automatically for a term in a given language by creating a USE relation for a forbidden term, and the relation is created for one language at a time and need not exist for all languages. The process for this function is the same as for THES and MTHES (see figure 6.17). The only difference is that keywords are to be entered in three languages and the facet number of the term is to be optionally provided.

Changing Terms

This function of the program allows descriptor spelling changes. Facet number and/or one or several terms may be changed. If the term has been changed, the modification is made throughout the thesaurus, and relations are maintained in alphabetical sequence.

Updating Relations

Only scope notes may be updated. Scope note lines may be increased to more than one line by pressing '>' in front of the line. For other relations, they must be deleted and created anew.

Deleting Relations

Any relation may be deleted by pressing **'D'** in front of the corresponding line: the reciprocal relations in other languages will also be deleted if applicable.

Deleting Keyterms or Forbidden Terms

In order to delete terms or forbidden terms, all relations must be deleted first. Again option **'D'** is used for this purpose.

6.2.2.3.3 Search Program

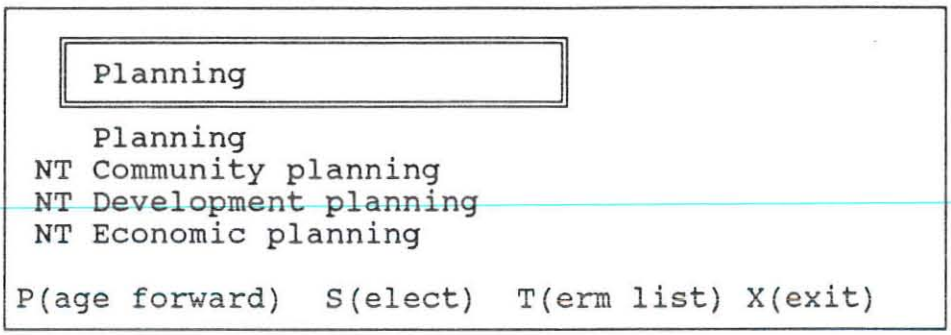
This program (THES2) is designed to assist in thesaurus scanning, term selection, and term translation. The program allows switching between current database and the thesaurus each time the thesaurus facility is invoked for term selection. For this purpose, the menu options are included in EXGEN/FXGEN/SXGEN menus (see fig. 6.22). The program is invoked in the currently selected language and there is no built-in language change while processing the thesaurus. The language of interest must be selected before starting through the standard ISIS menu option.

The same principle is applied in menus as in program THES1, i.e. **'L'**, **'T'**, **'S'**, **'>'** commands may be used to go back and forth in the thesaurus.

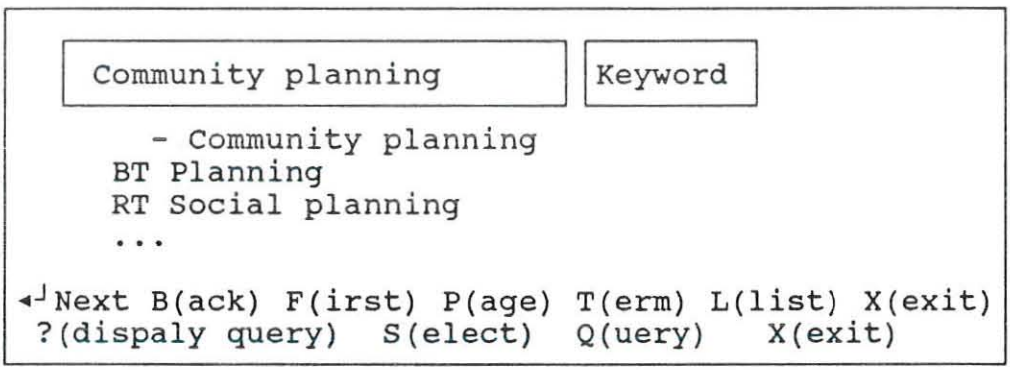
During selection and browsing of terms and relations with the appropriate commands, pressing **'Q'** will select terms for formulating searching expressions; **'?'** will display terms currently selected, **'X'** will exit from select/browse mode. Then, the search expressions built-up from the selected terms may be edited before the search is addressed to the database. The result of this process will be search in English, French and Spanish, and an **OR** on the three language set. It is possible to search in a specific language, if needed.

The screen layouts in the search process is similar to that in THES and MTHES. (see fig. 6.18)

Fig. 6.25: THES2 Select Mode Screen Layout



THES2 screen layout after selection
from a list of terms in figure fig. 25.



THES1 and THES2 programs provide for various printing outputs as shown in the following examples. In example 1, the English alphabetical thesaurus with top terms (TT)

numbered is displayed using the display format EALPHA, the number attached to TT showing the level of the term in the hierarchy; e.g. TT1 indicates the first top term, TT2 indicates the second top term and so on. Example 2 shows hierarchical display in English using the display format EHIERA with dots in place of TT, BT and NT taking two terms, i.e. household income and planning.

Example 1: Sample display of English alphabetical thesaurus using the display format EALPHA

```
HOUSEHOLD INCOME - 03.02.05
  TT1 INCOME
  BT  INCOME
  RT  FAMILY BUDGET
      STANDARD OF LIVING
```

Example 2: Sample hierarchical dispaly in English using the display format EHIERA with hierarchies shown by dots.

```
-   POPULATION
.INDIGENOUS POPULATION
.LABOUR FORCE
.OVERPOPULATION
.RURAL POPULATION$
..AGRICULTURAL POPULATION
...AGRICULTURAL WORKERS
....FARMERS
.....TENANT FARMERS
....FISHERMEN
....FORESTRY WORKERS
..RURAL WOMEN
..RURAL WORKER
...AGRICULTURAL WORKERS
....FARMERS
.....TENANT FARMERS
....FISHERMEN
....FORESTRY WORKERS
..RURAL YOUTH
.SCHOOL-AGE POPULATION
.URBAN POPULATION
```

Example 2 continued

```
-   PLANNING
.COMMUNICATION PLANNING
.DEVELOPMENT PLANNING
..DEVELOPMENT PLANS
...NATIONAL PLANS
...REGIONAL PLANS
...URBAN PLANS
..DEVELOPMENT PROJECTS
...AGRICULTURAL PROJECTS
...HOUSING PROJECTS
...INDUSTRIAL PROJECTS
...JOINT PROJECTS
...MULTIPURPOSE PROJECTS
...PILOT PROJECTS
.ECONOMIC PLANNING
..AGRICULTURAL PLANNING
...FISHERY PLANNING
...FORESTRY PLANNING
..EXPORT PLANNING
..FINANCIAL PLANNING
..INDUSTRIAL PLANNING
..MANPOWER PLANNING
..MARKET PLANNING
..PRODUCTION PLANNING
.EDUCATIONAL PLANNING
.ENVIRONMENTAL PLANNING
.FOOD PLANNING
.HEALTH PLANNING
.HOUSING PLANNING
.LOCAL PLANNING
.NATIONAL PLANNING
.PHYSICAL PLANNING
.PROGRAMME PLANNING
.REGIONAL PLANNING
.RURAL PLANNING
.SOCIAL PLANNING
..POPULATION PLANNING
.TRANSPORT PLANNING
.URBAN PLANNING
```

The various files for printing the thesaurus are given in Annex 6.1. From the annex, it can be seen that these programs have more capabilities than present here.

6.2.2.4 THESHI.PAS Program

CDS/ISIS Pascal program THESHI provides for display of terms in a hierarchy with indication of the level of the hierarchy numbered. As may be seen from the example given below, the terms are picked up by the program from different hierarchies and brought together.

Sample hierarchical display of ternms using THESHI Pascal Program.

```
-       POPULATION
- NT    INDIGENOUS POPULATION
- NT    LABOUR FORCE
- NT    OVER POPULATION
- NT    RURAL POPULATION
- NT 2     AGRICULATURAL POPULATION
- NT 3        AGRICULTURAL WORKERS
- NT 4           FARMERS
- NT 5              TENANT FARMERS
- NT 4          FISHERMEN
- NT 4          FORESTRY WORKERS
- NT 2     RURAL POPULATION
- NT 2     RURAL WORKERS
- NT 3         AGRICULATURAL WORKERS
- NT 4            FARMERS
- NT 5               TENANT FARMERS
- NT 4            FISHERMEN
- NT 4            FORESTRY WORKERS


-       FARMERS
- BT    AGRICULATURAL WORKERS
- BT 2     AGRICUALTURAL POPULATION
- BT 3        RURAL POPULATION
- BT 4           POPULATION
- BT 2     OCCUPATION
- BT 2     RURAL POPULATION
- BT 3        RURAL POPULATION
- BT 4           POPULATION
- BT 3        WORKERS
- BT 4           MANPOWER
- BT 5              HUMAN RESOURCES
```

programs is wide. One of these application areas is assisting the design of specialized local databases and user interface for information searching and retrieval. The purpose of reviewing the capabilities and applications of the above presented four thesaurus pascal programs is to assess how they can be applied for such specialized objectives.

Table 6.5: The Global Program Options Used in the four Programs

Global Option Used	THES	MTHES	THES1 and 2	Remark
Language	Mono	Mono	Trilingual	
T(erm select)	yes	yes	yes	D
L(ist term)	yes	yes	yes	D
S(elect)	yes	yes	yes	
A(dd relation)	yes	yes	yes	
C(reate term)	yes	yes	yes	D
D(elete)	yes	yes	yes	D
Q(uery)	yes	yes	yes	
?(display query)	yes	yes	yes	
?(display query/save)	no	yes	yes	
U(pdate)	no	no	yes	
Z(change language)	no	no	yes	
>(change scope note)	no	no	yes	
X(exit)	yes	yes	yes	
V(save)	no	yes	yes	
R(add term file)	no	yes	yes	
M(ake file)	no	yes	yes	
/(search/add)	no	yes	yes	
P(age)	yes	yes	yes	
F(irst)	yes	yes	yes	
B(ack)	yes	yes	yes	
↵next	yes	yes	yes	

D = Signifies that the option is slightly different in THES1 and THES2 as compared to the THES and MTHES.

CHAPTER SEVEN

THE DATABASE CONCEPT

7.1 DATA

Tscichritzis and Lochovsky define data as:

> A perception of a world can be regarded as a series of distinct, although sometimes related phenomena. From the dawn of time, human beings have shown natural inclination to describe these phenomena in some fashion whether they understand them completely or not. These descriptions of phenomena will be called data. Data correspond to discrete records, facts about phenomena from which we gain information (Tscichritzis and Lochovsky, 1982: 3).

These authors support their definition by quoting from Langefors (1977) that information is an increment of knowledge that can be inferred from data.

Data is defined in Webster's Dictionary as:

> "things known or assumed; fact or figures from which conclusion can be inferred."

The American National Standard Institute (ANSI) offers a dual definition for data as follows:

1. [Data is] a representation of facts, concepts, or instructions in a formalized manner suitable for communication, interpretation or processing by human or by automatic means.

2. [Data is] any representation such as character or analog quantities to which meaning is or might be assessed. Generally, we perform operations on data or data items to supply some information about an entity.

McDonough (1963):

> The term data is used here to represent message that can be available to the individual but which have not as yet been evaluated for their worth to him in a specific situation ...

This definition implies that when data are evaluated by different users, the information inferred will be different depending on the situations for which the resultant information is to be used or applied.

More recently, Debons and Horne (1988: 7) defined data as a product of symbols that are organized according to established rules and conventions. Weinberg and Giller (1988) describe data as a name given to a fact regardless of who might use it or how it might be used. Whether the

fact is a law of nature, a measurement, a word written on an application form, or just something we assume to be true, an available fact is data.

All these definitions in one way or another, refer to data as being raw facts and figures from which we get useful information through analysis and evaluation.

7.2 DATABASE

A database is a generalized integrated collection of data together with its description, which is managed in such a way that it can fulfil the different needs of its users (Bowley, 1990: 59).

According to Davis (1980: 10) the most widely quoted definition of database is that it is:

> A common pool of shared data in which the data is interrelated, where each item of data is only stored once and which provides a service to a wide range of applications.

A database can also be viewed in another way in terms of the structure of a database itself. A database is made-up of data elements, fields, records, and files. One or more data elements make up a field, one or more fields together make up a record, one or more records make up a

file, one or more files make up a database, and one or more databases may be needed to meet users' interests (Neelameghan, 1992: 203-204).

In order to build, handle and manage a given database, we need a database management system (DBMS) which is a generalized software. A DBMS is a system that generates, runs, and maintains a database and as such the system must include all of the software needed for this purpose (Rowley, 1990: 59). That is, it is a system that provides facilities to define database schemes and provides operations that can be used to transform one database to another, usually of the same schema (Tscichritzis and Lochovsky, 1982: 13). A DBMS requires the database structure to be described in terms of record types, data items and relationships. These are defined using a data definition language (DDL) and the resulting definition is often referred to as a schema or database schema (Ralson, 1983: 443).

The concept of database emerged in the late 1960s as users' demand for information coincided with advances in computer technology and increased expertise in software engineering (Rowley, 1990: 59). Traditional file based systems were found inadequate, particularly for generating high level planning and control information. Some of the disadvantages of conventional file structure

- additional programming effort required to establish and maintain relationships between files, provide access control and recovery facilities;
- it is not easy to interface with data communication systems;
- it is difficult to enforce standards.

The shift towards database systems has been brought about by the increasing variety of information needs of users. These needs imposed on the data processing requirements in two ways (Davis, 1980: 9):

- the need to provide management [users] with consolidated and consistent information. It is claimed that a growing number of organizations are coming to realize that if information, which is normally dispersed, is brought together, interrelated and made available, then considerable benefits can result;
- the need to reduce maintenance and to speed up the implementation of new applications. Most data processing departments, it is claimed, are faced with growing problems of program maintenance and increasing lead times for new applications.

7.3 DATA ENTITY, DATA ATTRIBUTE AND DATA VALUE

In defining a data structure in a database, the terms entity, attributes, and value are used. An entity is something about which it is desirable to store data. An entity must be uniquely defined, but may vary from physical object (such as town) to a more abstract concept (such as social services). A data entity can also be defined as a set of objects that share one or more attributes in common among themselves and about which an organization may be interested in collecting descriptive data for use. (Neelameghan, 1992: 204).

A data attribute is a property or characteristic of an entity. The descriptive data collected about an entity is data about attributes of the entity or attribute data (Rowley, 1990: 65, Neelameghan, 1992: 204).

Data values are the specific data of an attribute describing an entity. Values may be measurement or descriptive information (Brackett, 1987: 15).

7.4 DATA MODELS

In order to describe and clarify the structure of a database, it is useful to define the concept of data

7.3 DATA ENTITY, DATA ATTRIBUTE AND DATA VALUE

In defining a data structure in a database, the terms entity, attributes, and value are used. An entity is something about which it is desirable to store data. An entity must be uniquely defined, but may vary from physical object (such as town) to a more abstract concept (such as social services). A data entity can also be defined as a set of objects that share one or more attributes in common among themselves and about which an organization may be interested in collecting descriptive data for use. (Neelameghan, 1992: 204).

A data attribute is a property or characteristic of an entity. The descriptive data collected about an entity is data about attributes of the entity or attribute data (Rowley, 1990: 65, Neelameghan, 1992: 204).

Data values are the specific data of an attribute describing an entity. Values may be measurement or descriptive information (Brackett, 1987: 15).

7.4 DATA MODELS

In order to describe and clarify the structure of a database, it is useful to define the concept of data

models. A data model is a schema to represent the real world using information concepts and structures (Neelameghan, 1992: 205). Korth and Silberschatz (1986: 6) define a data model as ... a collection of conceptual tools for describing data, data relationships, data semantics, and data constraints.

Macleod and Reuber (1987: 162) describe a data model as follows:

> A data model is comprised of object types, operations and integrity rules. The object type defines the local structure, the operator defines what operations or transformations can be performed on the object type, and the integrity rules define any additional constraint on the data (such as, the constraints of uniqueness of the data).

Harington referred to a data model in a simplified way 'a way of depicting data relationships' (Harington, 1989: 63).

Data models are generally classified into three major categories: object based logical models, record based logical models and physical data models. Each of these major categories is further classified into sub-data models. It is not within the scope of this paper to describe all of the three categories of the data models and their sub-divisions. Only the record based data model

is discussed here.

The record based data model is divided into three well known and accepted standard data models. These are: hierarchical model, network model, and relational model.

7.4.1 Hierarchical Data Model

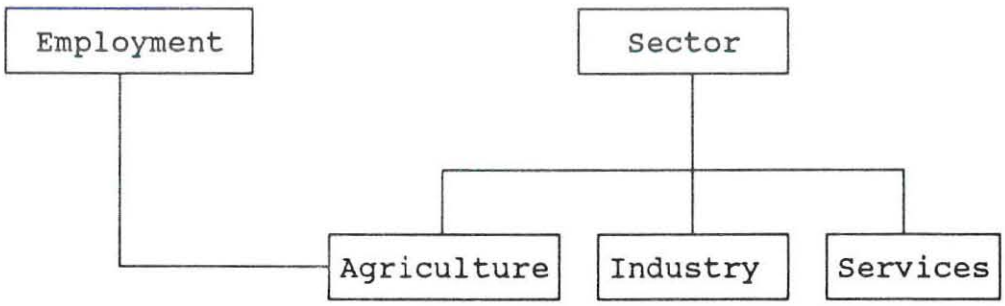
A hierarchical model consists of a collections of records which are connected with each other through links. A record is similar to the record in the network model. Each record is a collection of fields (attributes), each of which contains only one data value. A link is an association between exactly two records. Thus, a link is similar to the link concept in the network model (Korth and Silberschatz, 1986: 145).

Hierarchical model is based on the assumption that all data can be represented within a single hierarchy. The pure hierarchy model allows only one-to-many relationships, although variations permit more than one hierarchy and allow interconnections between different hierarchies. The model has some relevance to document organizations (Macleod and Reuber, 1987: 163).

Thus a link can be viewed as a restricted (binary) form of relationships.

Network model permits in principle many-to-many relationships which, as well as permitting hierarchies, also allows data to be shared among hierarchies (Brodie et al, 1984). The network model is also relevant to document collections. One useful application would be in permitting different logical organizations to coexist as they do in real world situations (Macleod and Reuber, 1987: 163). The records in the network data model are organized as collections of arbitrary graphs.

Using the example mentioned in the hierarchical data model, the network data model can be shown as follows:



In the example, attribute agriculture created a network relationship between the employment and sector entities. This model can represent one-to-many or many-to-many relationships.

7.4.3 The Relational Data Model

From an historical perspective, a relational data model is relatively recent. This model was developed by the mathematician E. F. Codd (1970: 377-387). Codd took concepts from mathematical set theory to describe a way to model data relationships that he believed to be intuitive and natural to people. The only component of a relational database is a two dimensional table made up of rows and columns, the intersection of which is a field. The term for such a table in mathematical set theory is a relation, from which relational database systems get their name (Harington, 1989).

A relational database is nothing more than a collection of relations. The data and the relationships among data are represented by a collection of tables each of which has a number of columns and rows with unique names. In this model, tables and relations are basic data types. Each table has structure. A row in a table represents a relationships among a set of values and since a table is a collection of such relationship, there is close correspondence between the concept of table and the mathematical concept of relation.

Taking, as an example, the entity employment in a data base, the relational model can be represented with the following table.

Relational data model: Employment relation:

Town	Government Employment	Private Employment	Self Employment	Other Employment
Assela	3,365	1,262	3,038	213
Ambo	1,497	606	1,664	126
Akaki	8,896	2,245	5,236	602

Each row in the table represents a single entity as can be seen from the relation table shown above.

7.4.4 The Difference Between the Three Data Models

The difference among the three data models is discussed by Macleod and Silberschatz (1987: 163). The fundamental difference between relational and the other two models is the degree of flexibility of retrieval. In the hierarchical and network models, the relationships among objects are predetermined and are represented by explicit pointers. Thus, query languages are simple, almost procedural in nature and are mainly focused on navigation. The hierarchical model differs from network model in that the records are organized as collections of trees rather than arbitrary graphs.

2.5.1 One-to-One Relationships

This relationship states that two data objects relate exclusively to one another. The data objects may be of the same or different types. Among the common one-to-one relationship is that between town and town plan. This relationship can be visualized as follows:

As can be seen from the diagram, there is a unary relationship on each side.

The actual data elements (attributes) that describe each data object are not usually shown in the diagram. The relationships are between data objects and not between individual data elements.

Even though all data models can accommodate one-to-one relationships, these rarely appear in database schemes. Except in unusual conditions, this relationship suggests that the data objects are not really distinct at all, and that they should be managed as single data objects.

7.5.2 One-to-Many Relationships

This relationship states that one data object is related to many data objects of the same or different type. In this relation, there is a unary association on one direction and a multiple association in the other direction. An example of this relationship is given in the following diagram.

It can be seen from the diagram that a town may have many residents while the resident lives in only one town at a time. The town data object is referred to as the parent in this relationship, and the resident data object is referred to as a child. (In one-to-many relationship, a parent data object has many children, but a child data object has only one parent). This type of relationships may be found in business database environments.

7.5.3 Many-to-Many Relationships

If a many-to-many relationship exists, then many data objects of one type relate to many data objects of another type. There is multiple association in both directions as shown in the following diagram:

In many-to-many relationships, each parent object (example, Town) has many child data objects (example, Planner). This means that a town can be planned by more than one planner, thus relating to many planners. On the other hand, each child data object can be related to many parent data objects. Therefore, in our example, a planner can plan more than one town, relating to many towns.

7.6 THE NEW DATABASE APPLICATIONS

In addition to the general database frameworks discussed above, there are other data base types which do not fit into the general nature of database systems and approaches. These are often known as specialized databases due to their specific area of application. Such specialized databases are discussed by Korth and Silberschatz (1986: 465-478). Four specialized database types are given as follows:

7.6.1 Design Databases

In a computer aided design (CAD) system, a large amount of data must be stored to represent the item being designed. The data describing a design are interrelated

In many-to-many relationships, each parent object (example, Town) has many child data objects (example, Planner). This means that a town can be planned by more than one planner, thus relating to many planners. On the other hand, each child data object can be related to many parent data objects. Therefore, in our example, a planner can plan more than one town, relating to many towns.

7.6 THE NEW DATABASE APPLICATIONS

In addition to the general database frameworks discussed above, there are other data base types which do not fit into the general nature of database systems and approaches. These are often known as specialized databases due to their specific area of application. Such specialized databases are discussed by Korth and Silberschatz (1986: 465-478). Four specialized database types are given as follows:

7.6.1 Design Databases

In a computer aided design (CAD) system, a large amount of data must be stored to represent the item being designed. The data describing a design are interrelated

in a complex manner. Furthermore, there is a need to retain not only the current design, but also a record of a previous version of the design. As a result, the data processing approach to databases is not adequate.

7.6.2 Knowledge Databases

In artificial intelligence and expert systems, information is represented as facts expressed in logic. This collection of facts can be viewed as a database containing knowledge, or as a knowledge base.

7.6.3 Multimedia Databases

Data of a graphical nature may be stored in a database together with textual data. Such a database may be accessed based upon the structure of a graphical data item. Database language designed for data processing applications are not adequate for the purpose. Similar problems arise for audio data, design data, and other types of data with complex substructure.

7.6.4 Environment Modelling

Databases are being used as a component of systems design to automate software development and to simplify user

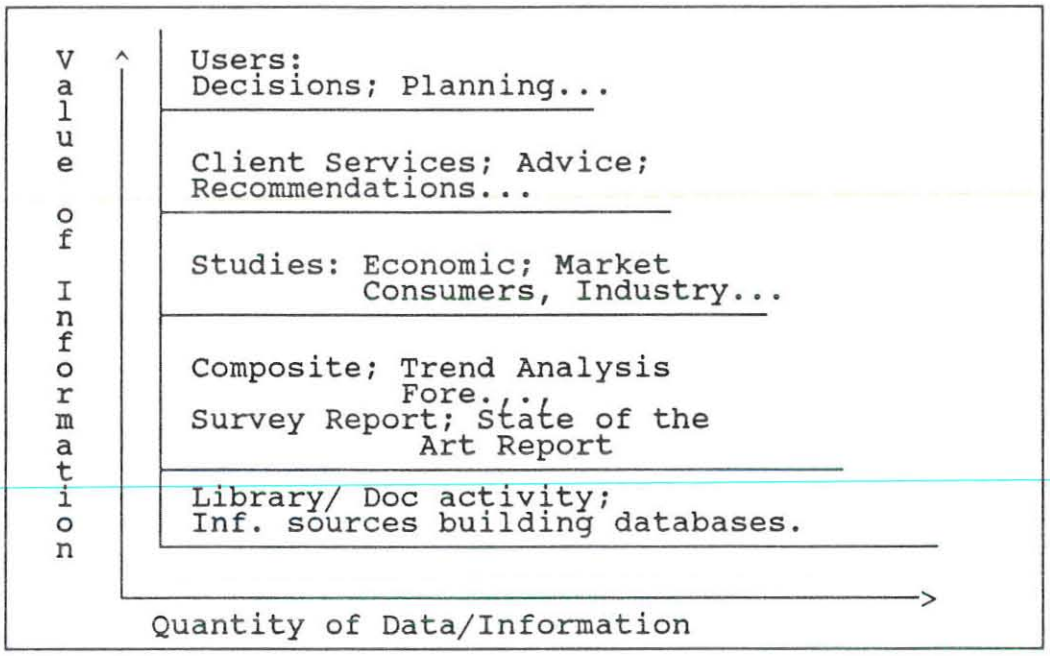
interaction with a complex information system. Representations of the components of these environments in a database require extensions of the data models.

7.7 OBJECT-ORIENTED DATABASES (OODB)

Databases containing information on an object are getting prominence in recent times. This is the result of the growing needs of planners, experts, decision makers, researchers and other specialists to get value added information; that is, synthesized, evaluated and repackaged information.

Specialized non-bibliographic object oriented databases (OODB) form the basis for the provision of value added information services. An object oriented database usually provides information on an object which may be a town, school, health centre, population of a given urban area, urban employment, market, etc. End-users are mostly interested in selected attributes of the object, with a view to manipulating, modifying, or using the object for particular purposes. Such databases can be home grown on micro computers to meet the needs of specialized user groups and effectively supplement or be integrated with bibliographic and referral type databases and services there of (Neelameghan, 1992: 260-261)

Figure 7.1: Information Hierarchy



Source: A. Neelameghan, 1992.

OODB design has a unique characteristics as compared to the more generalized bibliographic databases. One of its unique characteristics is the need and level of user involvement. Small but specialized institutions can design object-oriented database on their respective information need or to serve their user groups thanks to the availability of powerful information technology (IT) particularly microcomputers, at a reasonable cost, together with sophisticated portable software packages. The design of such databases and provisions of value added information, however , require (Neelameghan, 1992: 261):

- the involvement of end-users at various stages of the design and development of the database,
- close interaction between information professionals (IP) and end-users,
- that the subject background and work experience of the information specialists (IS) be generally compatible with those of the end-users services,
- that the training and education of IP should cover in depth information extraction, analysis, synthesis, and repackaging methods, the theories and principles of knowledge organization, natural language interface with information system, as well as IT applications in designing OODBs and preparation of value added product.

In designing a bibliographic database as a component of an information storage and retrieval system, involvement of end-users is usually minimal, as mentioned earlier, except in occasional assistance in the selection of subject descriptors and/or preparing abstracts in specialized fields. The other bibliographic data elements are mostly normalized (e.g. cataloguing rules, authority lists, vocabulary control tools, etc.) and these are applicable to documents in all subjects. This does not apply to OODB. the data element in an OODB (usually the attributes of the object and method of manipulating them)

are not the same for all objects. For example, the data element of a record in a database for urban population will be different from those for urban housing, and the latter will differ from those of a record on urban employment.

Source of information for input to most of the fields in bibliographic record is mainly the document to be catalogued. On the other hand, for a record in a non-bibliographic object-oriented database, the input data may be obtained from different sources, including remarks of individuals, even for a single field, and the data may already be in an analyzed, evaluated, and organized form.

OODBs attempt to give information/data themselves rather than just reference to other sources unlike the bibliographic and referral databases. The output from bibliographic databases follow a standardized format (e.g. AARC2 format). In specialized OODB, end-users may be provided with different types of output formats on the basis of their specific need.

The information provided by specialized OODBs is mostly of local interest and use (e.g. a database on a plan of a given urban area; forest resource, hospital patient, etc.) and may not be sourced in abstracting services at

the international and national levels. Therefore, such OODBs are usually home grown at the institutional or small specialized geographical level (Neelameghan, 1992: 261).

These and other factors necessitates very close involvement of end-users at various stages in the design of non-bibliographic specialized OODBs.

End user involvement in the design and development of an object oriented databases is needed, among other things, in some or all of the following aspects (Neelameghan,1991).

- to define the specialized subject areas in which databases are required at the institutional (individual) small group level,
- to develop reference framework(s) to analyze and synthesize the information, including views of experts in a pattern acceptable to end-users,
- to identify and advise on relevant sources of information including views of experts, sometimes for providing the data/information itself,
- to assist in the analysis, evaluation, filtering and organization of the data for input into the system,

the international and national levels. Therefore, such OODBs are usually home grown at the institutional or small specialized geographical level (Neelameghan, 1992: 261).

These and other factors necessitates very close involvement of end-users at various stages in the design of non-bibliographic specialized OODBs.

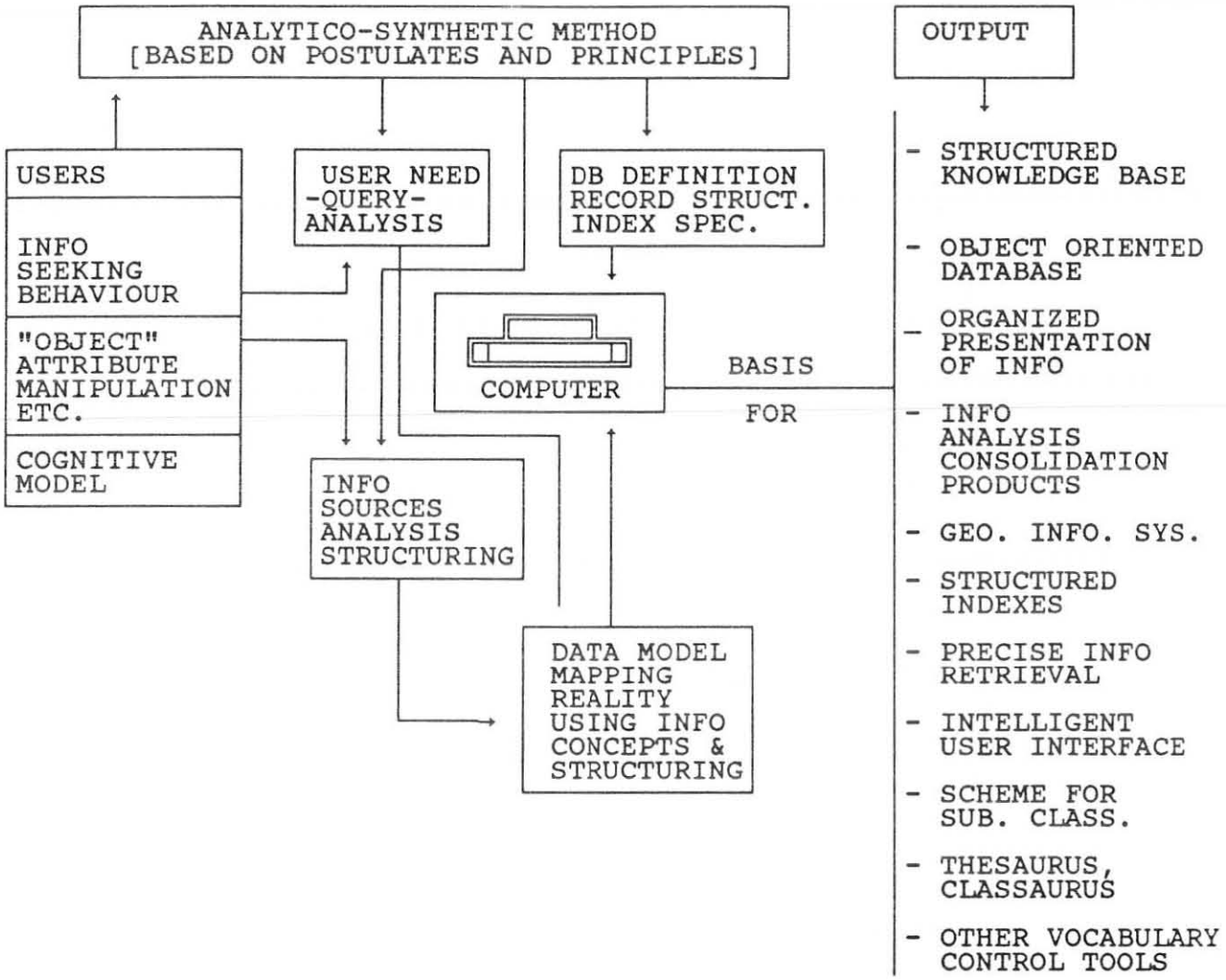
End user involvement in the design and development of an object oriented databases is needed, among other things, in some or all of the following aspects (Neelameghan,1991).

- to define the specialized subject areas in which databases are required at the institutional (individual) small group level,
- to develop reference framework(s) to analyze and synthesize the information, including views of experts in a pattern acceptable to end-users,
- to identify and advise on relevant sources of information including views of experts, sometimes for providing the data/information itself,
- to assist in the analysis, evaluation, filtering and organization of the data for input into the system,

- to advise on the field/data elements of record which is essentially the selection of attributes, other concepts relating to the object(s) of the database,
- for data rating,
- to provide the type of queries that the system should be designed to respond to,
- to advise on the standard terminology to be used in the system; to specify the type of output and their format that may be frequently required with a view to formulating a predefined output format while at the same time providing for the formulation of other formats as well.

As already mentioned, the basic construct of OODB is an object on which information is needed. This object should be manipulated, analyzed and represented with its unique attributes and attribute values. Based on the identified attributes of the object, the kind of information that is needed to be collected and stored (organized) on the object's attributes are analyzed source wise and structured in a helpful way. This structured information is then mapped to database model.

Fig. 7.2: OODB Design Process



Source: A. Neelameghan

These analyses and structuring processes can be based on the postulates and principles of analytico-synthetic method of knowledge classification and organization as shown in fig. 7.2.

Users need information on a given object. This information seeking behaviour of users is initiated in the form of query. The information need of users presented in the query can also be analyzed and structured in the same way as mentioned above. Then, the result of the analyses and structuring are also mapped to a data model. In order to store information in a database, the parameters of the database need to be defined in terms of record structures, indexing specifications, etc. Such a database can be the basis for generating various products and services (fig. 7.2).

The outputs from this system are many including OODB, intelligent interface, thesaurus, classaurus, and others.

The above process of analysis and structuring of an object on which information seeking behaviour of users can be summarized with the following:

> ... the theories and principles and techniques of concept categorization, classification and knowledge organization, and of the analytico-synthetic facet approach as well as theory of methods of vocabulary control and natural language indexing developed for organizing and retrieving subject information from bibliographic databases are helpful in designing non-bibliographic OODBs (Neelameghan, 1991).

The object of this work, as mentioned in many parts, is the generation of vocabulary control tools and examination of their application in designing object-oriented and other local databases and user interface. The system development includes the creation of a prototype OODB and as such the idea conveyed in the quotation is very important in this study.

> ... the theories and principles and techniques of concept categorization, classification and knowledge organization, and of the analytico-synthetic facet approach as well as theory of methods of vocabulary control and natural language indexing developed for organizing and retrieving subject information from bibliographic databases are helpful in designing non-bibliographic OODBs (Neelameghan, 1991).

The object of this work, as mentioned in many parts, is the generation of vocabulary control tools and examination of their application in designing object-oriented and other local databases and user interface. The system development includes the creation of a prototype OODB and as such the idea conveyed in the quotation is very important in this study.

CHAPTER EIGHT

DESIGN AND DEVELOPMENT OF SYSTEM COMPONENTS

8.1 OVERVIEW

This chapter presents the major system components designed and developed in this study. The components of the system are divided into four major parts. The first component is the examination of the queries collected from potential users of the system. The second component discusses the generation of vocabulary control tools. This component of the system is made up of the generation of classaurus, updating of macrothesaurus and a preliminary thesaurus from abstracts of documents on the urban planning. The third part is on the design and development of prototype integrated urban planning database. This part includes reference and specialized object-oriented databases. The fourth component of the system is the design and development of a search assistance facility which is of two types: a thesaurus-like search assistance and a structured user request (or query) form. Each component is presented in a separate section.

8.2 QUERY SURVEY

A search request form was designed to collect information on information needs from the target group of users of the developed system. The planners, researchers, and experts in NUPI were asked to fill in the form the type of information actually needed for their planning and research activities. The query collection was intended for three major purposes:

(1) To examine whether the terms or concepts included in the macrothesaurus and the classaurus actually represent the actual terms or concepts (vocabularies) the end-users use when they present requests for information. This was done by analyzing and categorizing the key terms or concepts the requesters used in the queries they presented. Facet analysis and categorization of the terms (or subject concepts) into facets (discipline, entity, property and action terms), was done, and their presence in the thesaurus was checked. The result was that most of them, i.e. from about 200 terms 166 (roughly 85 %), were already included in the classaurus (for more information see Annex 2). This showed that the thesaurus terms are representative of the actual subject terms used by the end-users themselves in their queries.

(2) To use the terms collected to maintain the macrothesaurus and the classaurus. This is actually related to the first purpose. Those terms which were not included in the database of terms were then added to maintain or update the macrothesaurus and the classaurus. The terms used by the requesters in their queries collected also served as indicators to some other relevant terms (subject concepts) which were added. Such terms are marked by an asterisk in Annex 2.

(3) To match the vocabulary control tools generated with the urban integrated database created. This was done to test how the system developed worked. (The details of this matching will be discussed later in section 8.4 of this chapter).

Thirty-two potential users completed and returned the query forms. In this section, a summary of the survey of queries is presented.

8.2.1 User-Requesters Background

The user-requesters background is summarized in table 8.1 and 8.2 under two categories. In table 8.1, the specialization of the requesters is presented with their respective frequency and percentage. As this table shows,

three areas of specialization account for the majority: economics, geography and, architecture and town planning in that sequence. The rest are almost equally distributed. The figures and ratio of the requesters' specialization is attributable to the total number of employees specializing in a field in NUPI in general. Therefore, it is inexpedient to make any inference from the figures and the ratios in table 8.1.

Table 8.1: Requesters Fields of Specialization

Field	Number	Percent
Geography (including GIS Physical Geography)	8	25.00
Architect and Town Planning	6	18.75
Economics (including economic policy planning, regional development planning)	9	28.13
Engineering (including Hydraulics)	2	6.25
History	1	3.12
Law	1	3.12
Demography	2	6.25
Statistics	2	6.25
Urban Design	1	3.12
Total	32	100.00

The main point to be noted and important implication is the variety of disciplines with which urban planning is concerned. This attests to the fact that urban planning is a multi-disciplinary field.

Table 8.2 summarizes the functional designations of the requesters. As the table shows, economic researchers,

physical planning researchers and architect-town planners are in the majority. The title researcher here refers to experts with different expertise and experiences who conduct studies and research activities on different aspects of urban areas and regions which are the basis for the plan preparation.

Table 8.2: Functional Designation of Requesters

Post	Number	percent
Researchers		
Physical planning Researchers	7	21.88
Population and Social Affairs Researchers	2	6.25
Economic Researchers	11	34.88
Historical Researcher	1	3.12
Architect-planners	6	6.25
Engineers	3	9.38
Technicians	1	3.12
Legal Expert	1	3.12
Total	32	100.00

Architects, planners, engineers, and technicians are those who produce the actual final urban plans. The legal expert advises the planners and decision makers on the rules and regulations that govern urban planning and other related policy areas. Even though the number of queries is few, the people working in different functional areas were represented as shown in table 8.2.

physical planning researchers and architect-town planners are in the majority. The title researcher here refers to experts with different expertise and experiences who conduct studies and research activities on different aspects of urban areas and regions which are the basis for the plan preparation.

Table 8.2: Functional Designation of Requesters

Post	Number	percent
Researchers		
Physical planning Researchers	7	21.88
Population and Social Affairs Researchers	2	6.25
Economic Researchers	11	34.88
Historical Researcher	1	3.12
Architect-planners	6	6.25
Engineers	3	9.38
Technicians	1	3.12
Legal Expert	1	3.12
Total	32	100.00

Architects, planners, engineers, and technicians are those who produce the actual final urban plans. The legal expert advises the planners and decision makers on the rules and regulations that govern urban planning and other related policy areas. Even though the number of queries is few, the people working in different functional areas were represented as shown in table 8.2.

8.2.2 Type of Information Requested

Earlier in chapter four, the major types of information required for urban planning in NUPI were discussed. The information needed as represented in queries was analyzed in line with those categories and the result is given in table 8.3.

Table 8.3: Type of Information Requested

Type of information	Number*	Percent
Physical characteristics information	18	33.96
Housing information	5	9.43
Demographic information	6	11.32
Economic activities information	2	3.77
Social infrastructure information	7	13.21
Legal/policy information	2	3.77
Public utilities and facilities information	5	9.43
Financial information	3	5.77
Historical information	1	1.89
Maps and plans information	1	1.89
Transport and communication information	2	3.77
Statistical information	1	1.89
Total	53	100.00

* In a query more than one type of information was requested and the total adds up to more than the number of queries.

From table 8.3, it is evident that all the major categories of subject areas on which information is required is covered by the queries. One area of information need, statistical information, is a new subject not included in the categories mentioned earlier. This particular information is needed for questionnaire and urban survey design. It is more related to

methodological information which may not be needed permanently or frequently.

The most frequently required type of information is physical information followed by social infrastructure, demography, housing, public utilities and facilities, and others in that sequence. But, the relative figures of the type of information do not mean any thing by themselves. Every bit of information required for urban planning is equally important in whatever magnitude or volume it may be required. Without all the types of information available in the forms and volume required, an urban plan cannot be completed. The difference in the frequency of request for a particular information is attributable to the number of requesters from that particular area, as mentioned earlier. Had the number of people who filled in the query form in the physical planning field been less, the result could have been different.

The significant inference from table 8.3 is the distribution of the type of information covered by the requests. The requests covered all the major areas of urban planning. This is another supporting evidence of the correctness of the major categories of the type of information needed for the preparation of urban plans.

In some queries, more than one type of information need areas were included. For instance, population and housing. In such cases, double counting was used, one for demography and one for housing. Thus the total frequency added up (i.e. 53) to more than the total number of requesters and queries (i.e. 32).

8.2.3 Coverage of Information

Each requester was provided with a choice between two alternatives regarding the coverage of the search result in response to his/her need stated in the query. Of the total requesters (32), 23 or 71.87% wanted the information to be specific to the search topic and 9 or 28.13% wanted the information to their request to be comprehensive in its coverage containing some information items that may not be very relevant to the search topic. This is presented in table 8.4. This has implication to the current work.

Table 8.4: Coverage of Information

Coverage	Number	Percent
Comprehensive Specific	9 23	28.13 71.87
Total	32	100.00

Persons concerned with a specialized sector, such as urban planning, need information mostly specific to the purpose. Such information is provided by an information system that provides value-added information products. Such information products are not usually possible to obtain from general bibliographic and international information services. It requires the design of special purpose and end-user specific local databases. This is what the information need pattern of NUPI's experts require.

8.2.4 Purpose for Which Information is Needed

The purpose for which requesters wanted information is categorized into four major groups: research, problem identification, planning and, developing new system/method of planning. This is presented in table 8.5. One requester did not state the purpose of the request for information.

Table 8.5: Purpose of Requesting Information

Purpose	Number	Percent
Research	5	15.63
Planning	17	53.13
Problem identification	7	21.88
Develop new system/method of planning	2	6.25
Not stated	1	3.12
Total	32	100.00

The requesters stated the purpose for which they needed the information in different ways. The four major purpose categories (Table 8.6) are summaries of the differently stated purposes.

Table 8.6: Purpose of Request in Detail

Purpose Category	Specific Purpose
1. Planning (Urban planning)	- For physical planning - For planning - For urban plan preparation - For urban plan purpose - To give planning solution to basic urban problems - Service centre planning - Urban planning - To feed information to planners - Master plan preparation
2. Research Project	- Research project - Policy research
3. Problem identification	- Problem identification - Problem identification and planning proposal - Problem identification and documentation - Identification of urban legal/ policy issues problems
4. Develop new planning systems and methods	- Develop native planning methodology based on indigenous tradition - Reorganize the existing urban planning system

As the table shows, the planning purpose in general refers to urban planning stated in different phrases. The information needed for research concerns two things: research project and policy research. Research project includes urban plan project and project for the preparation of paper for undergraduate degree program which is also directly related to urban planning.

Information need for problem identification relates to definition of urban problems which form a basis for plan proposal, identification of a particular problem of an urban area and documenting it for future consideration, and identification of problems relating to legal and policy issues in the urban sector of the country with respect to its impact on urban planning.

The other major area of information requirement is to develop new planning systems and methods. Information is needed to examine whether native (indigenous) urban planning methods exist or not and to develop a system of urban planning based on the indigenous building traditions (cultures) and techniques. The other is to propose a new system or reorganizing the existing systems of urban planning and (propose new systems) and methods. The two are closely related.

8.3 COMPUTER-GENERATION OF VOCABULARY CONTROL TOOLS

As mentioned earlier in chapter five, there are several types of vocabulary control tools. It is not the purpose of the study to cover all of them, but two types of vocabulary control tools: Classaurus and macrothesaurus. The computer-assisted generation and maintence of these tools is presented in the following sections.

8.3.1 Procedure for Generation

The design and development of any vocabulary control tool should follow a systematic procedure. The steps are more or less standardized and widely practised (Lancaster, 1972). A similar procedure was followed in the present study. The scope of subject coverage of the vocabulary control tools was defined as urban planning (see 1.2).

(1) The first step was the identification of sources to be used for selecting terms/vocabularies in the field of urban planning. The sources are documents produced by specialists in the field of urban planning, users' queries and other vocabulary control tools. Various documents, reports, papers, articles, etc. were identified that cover the major and sub-discipline of urban planning (housing planning, population planning, transport planning, etc.). As discussed in 8.2, query form was designed and distributed to potential users of the system to be developed (NUPI experts) to fill in the type of information they need for their actual work. These queries were also used as sources of vocabularies. The macrothesaurus of OECD was also consulted as a source of terms.

In the generation of this tool, the two thesaurus development requirements, literary warrant and user warrant, were ensured to justify the inclusion of a term or subject concept in the thesaurus.

The sources were books, research papers, documents, reports, etc. in the field of urban planning. To the extent possible, up-to-date sources were consulted to select new and emerging concepts in the field. The frequency of term occurrence in the literature was taken as the criteria for selection. The subject terms that are often used in a document by the author experts and which the potential end-users may use in their request for information are considered relevant for selection.

The second requirement, i.e. the user warranty requirement, was met in two ways. The first is that the literature from which the terms were collected are the products of the experts or specialists, including those at NUPI, in the urban planning field. Normally, when any vocabulary control tool is to be designed, the actual users should be fully consulted. In the present work, NUPI staff represented both the generators of information/information products (documents) as well as users of urban information and urban information systems.

Type: Action

Migration
> External migration
= Out-migration
> Internal migration
= In-migration

Example: Migration figures both internal and external illustrate the inherent difficulties of applying quantitative analysis.

Source: Town and Country Planning
by Ratcliffe, J.

Type: Entity

Land use map
< Map

Example: It is important that land use maps are kept up-to-date.

Source: Town and Country Planning
by Ratcliffe, J.

Type: Entity

Housing
> Private housing
> Public housing

Example: Housing planning and analysis is one aspect of urban planning.

Source: Housing Analysis
by Cullingworth, J. B.

decision making, etc. functions (explained in section 8.2 earlier). The terms in the queries were analyzed and checked for their frequency of use by the requesters. Then, those substantive terms used frequently were considered relevant for their inclusion in the thesaurus. To a limited extent, the actual users were also consulted on the use of the terms in their practical planning activities (see section 8.2).

The queries as sources of terms to be included in the thesaurus were considered after the classaurus had been developed. Thus, they were used as a means of testing the relevance of the classaurus terms to the end-users.

The terms collected from the above sources were organized into structures hierarchically and associatively. The development of classaurus follows a systematic faceted classification and organization of knowledge. Thus, the presentation of the classaurus developed has been made into different schedules: schedule of basic disciplines, schedule of entity terms, schedule of property terms and schedule of action terms. The meaning of the concepts discipline, property, entity and action as used in this work is given as follows (G. Bhattacharyya, 1981: 12-13):

To keep each schedule separately, unique identification codes were used: DIS, for schedule of basic discipline; ATT, for schedule of property terms; ENT, for schedule of entity terms, and ACT, for schedule of action terms. To browse terms in a given schedule, it is only necessary to use the regular CDS\ISIS search expressing process (i.e. using S or T) and enter the approprite code; for example, ENT to brouse the whole entity terms. This will invoke the display of the alphabetical listing of the terms in that particular schedule.

A sample of the terms selected and categorized into schedules is presented below. In the example, BD stands for basic discipline; P stands for property term; E stands for entity term; and A stands for action term.

Title: Urban and Regional Planning: A Systems Approach

Urban planning (BD)
Regional planning (BD)
Approach (P)
Planning (A)
Urban (E)
Region (E)
System (E)

Title: Principles and Practice of Town and Country Planning

Country Planning = Rural Planning (BD)
Town Planning = Urban Planning (BD)
Principles (P)
Town = Urban (E) Country = Rural (E)
Practice (A)

Title: Recent Development in Land use Planning Modelling

Urban planning (BD - implicit)
Land use Planning (BD - sub division of Urban planning)
Land (E)
Development (A)
Modelling (A)
Use (A)

Sample record of ENTIY Terms continued.

NT Hotel areas
NT Industrial area
NT Parking area
NT Prison area
NT Recreational area
NT Religious area
NT Residential area
NT Rural area
NT Slum area
NT Suburban area

SCHOOLS (0001)
NT Elementary schools
NT Higher secondary schools
NT School sections
NT Junior schools
NT Secondary schools
RT Enrolment
RT School age population
RT School population (eligible)
RT Student-class ratio

Sample Records of PROPERTY Terms

EQUILIBRIUM (0692)
NT Inter-community equilibrium
NT Intra-community equilibrium

INTRA-COMMUNITY EQUILIBRIUM (0693)
BT Equilibrium
RT Inter-community equilibrium

INTER-COMMUNITY EQUILIBRIUM (0694)
BT Equilibrium
RT Intra-community equilibrium

Sample Records of ACTION Terms

ANALYSIS (0766)
NT Cost benefit analysis
NT Critical path analysis
NT Impact analysis
NT Input output analysis
NT Sensitivity analysis
NT Threshold analysis

COST BENEFIT ANALYSIS (0767)
BT Analysis
RT Cost benefit

As can be seen from the exaple above, the terms in the classaurus are arranged alphabetically in each level of term arrays (e.g. **analysis** and **cost benefit analysis** are arranged alphabetically, and the terms in each primary division are also arranged alphabetically).

The classaurus developed in this work has the following applications. It can be used:

- to improve and enhance retrieval from databases covering the subject of the classaurus,
- to maintain or update a macrothesaurus,
- alongside a macrothesaurus to link queries to databases and to enhance the retrieval of information from databases,
- to design a user interface for the urban planning databases created,
- to facilitate design any database in the urban planning field.

Application of vocabulary control tools in general is at two levels: it can be applied at database design stage or at retrieval stage or at both stages. In this study, the two thesauri are applied only at retrieval stage. The classaurus and the macrothesaurus are applied to improve retrieval capability from databases. They are used as interface between users and the records in the databases.

descriptor is followed by scope note (SN) when necessary, hierarchical terms (broader term, BT; narrower term, NT), or associative terms (related terms, RT) to which it is related.

The second part of the Macrothesaurus presents descriptors and their synonyms and sub-topic. The third part arranges descriptors according to their hierarchical relationships, and the final part is permuted descriptor index which supplements the hierarchical classification given in the first part.

The Macrothesaurus was interfaced with the urban planning prototype Micro-CDS/ISIS databases to facilitate retrieval of information via macrothesaurus search system. This is facilitated through the use of THES2 CDS/ISIS pascal program.

The Macrothesaurus has been applied for different purposes. Some of its applications are: in designing a local microthesaurus, by updating it or by taking relevant terms from it, in the specific field of interest and in integrating databases developed with it. NUPI, being one of institutions involved in development and planning fields, can apply this thesaurus in the same way. The thesaurus can help to integrate local databases

need to use a language parser. Hierarchical and other relations can be identified more specifically from the subject context, in the text itself rather than from individual terms collected on separate term cards from texts.

The main purpose of this section is to demonstrate the generation of a thesaurus directly from a text using Micro-CDS/ISIS. The direct thesaurus generation from the text itself also resulted in greater efficiency in the procedures as against the usual semi-manual method. A thesaurus generated from a single text or its well prepared abstract can provide for the majority of the terms and their relations as they are perceived by the experts in the field. This has advantage in saving the time required to design and develop a thesaurus. Users' queries can also be used as a source for generating thesaurus directly. This process of thesaurus generation is particularly helpful as whole text materials or abstracts are available on computer readable electronic media.

A preliminary thesaurus was generated from abstracts of some 12 articles on urban plan studies. Abstracts were prepared by specialists in urban planning. The sentences in the abstract in field 64 were copied on to fields 67

need to use a language parser. Hierarchical and other relations can be identified more specifically from the subject context, in the text itself rather than from individual terms collected on separate term cards from texts.

The main purpose of this section is to demonstrate the generation of a thesaurus directly from a text using Micro-CDS/ISIS. The direct thesaurus generation from the text itself also resulted in greater efficiency in the procedures as against the usual semi-manual method. A thesaurus generated from a single text or its well prepared abstract can provide for the majority of the terms and their relations as they are perceived by the experts in the field. This has advantage in saving the time required to design and develop a thesaurus. Users' queries can also be used as a source for generating thesaurus directly. This process of thesaurus generation is particularly helpful as whole text materials or abstracts are available on computer readable electronic media.

A preliminary thesaurus was generated from abstracts of some 12 articles on urban plan studies. Abstracts were prepared by specialists in urban planning. The sentences in the abstract in field 64 were copied on to fields 67

and 68. Field 67 is for generating hierarchical relations (BT/NT) and synonyms (=) and field 68, for associative relations (RT). [Note: These can be done directly in the original abstract in field 64 itself].

The database name used to generate the preliminary thesaurus is called CDS. The field definition table (FDT), work sheet (FMT), display format (PFT), sample entry of the abstracts and a sample output of the thesaurus are presented below.

Field Definition Table (FDT) Database: CDS

?	Tag	Name	Len	Type	Rep	Del/Patt
	12	Conference main entry	300	X		npdz
	24	Title	500	X		z
	25	Edition	100	X		
	26	Imprint	300	X		abc
	28	Host document	300	X		
	30	Collation	100	X		abc
	44	Series	300	X	R	vz
	50	Notes	500	X		
	60	Class number	25	X		
	64	Abstract	500	X	R	
	67	Strings for hier. rel	500	X	R	xheijsa
	68	Strings for assoc. rel	500	X	R	xerij
	69	Keywords	1000	X		
	70	Personal Author	100	X	R	
	71	Corporate Bodies	300	X	R	
	72	Meetings	300	X	R	npdz
	74	Added title	500	X	R	z
	76	Other language titles	500	X	R	z
	80	ISBN	20	X		
	81	ISSN	20	X		

Worksheet: CDS Page: 1/3 Database: CDS

Conference(s) (012) .....
Title (024) .....
Edition (025) .....
Imprint (026) .....
Host document (028) .....
Collation (030) .....
Series (044) .....
Notes (050) .....
Abstract (064) .....
Structured strings 1 (067) .....
Structured strings 2 (068) .....
Keywords (069) .....
Personal Author (070) .....
Corporate body(ies) (071) .....
Added title(s) (074) .....
Other language titles (076) .....
ISBN (080) .....
ISSN (081) .....

Database Name: CDS Format Name: THES0

if p(v67) then d67 mhu,(v67^e/¦ NT ¦,v67^h,x1,v67^e /v67^h,x1,v67^e/¦ BT ¦,v67^e/if v67^x:'are' or v67^x:' is a' then v67^e/¦ BT ¦,v67^h/v67^h/¦ NT ¦,v67^e/ fi fi v67^s/¦ = ¦,v67^a/v67^a/¦ = ¦ v67^s/) fi, if p(v6 8^x) or p(68^i) then mhu, d68,(68^e/¦ RT ¦,v68^i/c7, v6 8^r/v68^r/¦ RT ¦,v68^j/c7,v68^e/) fi, if a(v68^x) and a(v68^i) then d68, mhu, (v68^e/,¦ RT ¦,v68^r/v68^r/,¦ RT ¦,v68^e/) fi

Sample Data Entry to the Database CDS

ABSTRACT:

The paper exposes the model on the political economy of local governments with different features of local government; i.e. exogeneously determined boundary, land market, housing price and movements of households along jurisdictions. Mathematical models with graphs and tables showing varying cost function parameters on effects of utility function and housing supply parameters are applied. Three goods were assumed in the models: a local public goods, local housing goods and composite private goods, in relation to tax rate of

different assumptions and voters. The existence of intra- and inter-community equilibrium is also indicated.

HIERARCHIC REL :^hMathematical^emodel^hintra-community^e equilibrium^hinter-community^eequilibrium

ASSOCIAT. REL. :^eModel^i(for)^xon the^rpolitical econom y^j(of)^eModel^i(for)^xof^rlocal government ^j(of)^ e Local government^i(features)^rexogeneously determ ined boundary^j(features of) ^eLocal government^i(f e at ures)^rland market^j (features of)^eLocal gove rnment^i(features)^r housing price^j(features of)^e Local government^i(features)^rmovements of househol ds^j(features of)^eMathematical model^i (for) ^xwit h graphs and tables showing varying^rcost function^ i(of)^eUtility function^i(affecting)^xand^rhou sing supply^j(affects)^epublic goods^i(model parame ter) ^xlocal housing goods, and composite private goods, in relation to^rtax rate^j(in relation to) ^xl ocal public goods^elocal housing goods^i(parameter model)^xand composite private goods, in relation to^rtax rate^j(in relation to)^xlocal public goods, local h ousing goods and^ecomposite private goods^i(model parameter)^xin relation to^rtax rate^j(in relation to)^elocal public goods^i(model parameter)^xlocal housing goods and composite private goods, inrelat ion to tax rate of different assumptions and^rvoter s ^j (in relation to)^xlocal public goods^elocal ho using goods^i(model parameter)^xand composite pri v ate goods, in relation to tax rate of different as sumptions and^rvoters^j(in relation to)^xlocal publ ic goods, local housing goods and^ecomposite priva te goods^i(model parameter)^xin relation to tax rat e of different assumptions and^rvoters ^j(inrelatio n to)^ecommunity^i (attribute) ^requilibrium^j(inre lation to)

Sample output of the Preliminary Thesaurus

Hierarchical relations from records in field 67 above.

Model
  NT Mathematical model
Mathematical model
  BT Model
Equilibrium
  NT Intra-community equilibrium
Intra-community equilibrium
  BT Equilibrium
Equilibrium
  NT Inter-community equilibrium
Inter-community equilibrium
  BT Equilibrium

Associative relations from records in field 68 above.

```
Model
  RT (for)
     Political economy
Political economy
  RT (of)
     Model
Model
  RT (for)
     Local government
Local government
  RT (of)
     Model
Utility function
  RT (Affecting)
     Housing supply
Housing supply
  RT (Affected by)
     Utility function
```

The preliminary thesaurus thus generated can be sorted out using CDS/ISIS option P in the main menu (ISISPRT) After sorting, duplicate terms or concepts were checked and eliminated.

Using MTHES pascal program (see chapter 6) each term in the preliminary thesaurus was checked in the classaurus and Macrothesaurus. Those not present in these thesauri were added and appropriate relations established.

8.4 THE INTEGRATED URBAN PLANNING DATABASES

The integrated database is made up of different records describing different entities. These different records can be searched simultaneously for a single query. Based on these possibilities, an integrated urban planning database has been developed.

8.4.1 Reference Records

These are composed of (1) bibliographic and (2) referral records. The bibliographic records are descriptions of documents of various kinds (books, articles in periodicals, offprints, discussion papers, reports, etc.) Retrieval will yield citations of original document(s) dealing with urban planning subjects.

The referral records are profiles of institutions, projects and experts in urban planning and related areas. Such information is important for NUPI in supporting its planning activities.

In the process of planning, NUPI makes contacts with different organizations or institutions for different purposes. Some institutions are sources of information. Some are involved in the planning process in one way or another. Still others are engaged in tasks that have direct or indirect influence on the plan NUPI prepares. Thus, it is necessary for NUPI to have contacts and access to such institutions at the right time and hence the need for the profiling the institutions in a database for easy retrieval. The institutions may be local, national or international; governmental or non-governmental.

A number of development and other projects on urban areas, regardless of their magnitude, exist. These projects may have been completed or ongoing or proposed. In preparing urban plans for a given urban centre, the details of such projects and their status need to be taken into account. NUPI needs the data collected already, contact with personnel involved in the projects, etc. and to integrate the existing projects/plans with the original urban plan.

In addition to knowing about other projects, it is necessary for NUPI to have a record of the projects it prepares, proposes or had completed. This information of NUPI's own projects, could serve the same way as those on projects undertaken by other organizations. Hence, the databases of research projects profiles.

The work of urban planning requires experts of various skill levels and types. NUPI needs short term consultants in some specialized areas for which qualified persons are sought for as the need arises. In addition, in undertaking some special assignments, it subcontracts part of its projects to appropriate consultants. The creation of profiles of experts , both within and outside NUPI, engaged in the urban development and planning areas, helps in making such contacts. The prototype

8.4.2 Specialized Object-Oriented Database

Researchers, planners, executives, decision makers, etc. in the urban planning sector are usually interested in analyzed, synthesized and organized information for their planning tasks. This type of information can be provided by developing specialized object-oriented databases (OODBs) taking into account the interests and needs of the target user group (Neelameghan, 1992).

OODBs are designed to give organized and ready for use information, unlike the reference databases which usually provide only citations or point to other sources.

As part of the database development, a prototype urban OODB has been created. The object about which information is gathered and organized in the database is 'town'. The attributes of the town on which data was collected include finance (income and expenditure distributions), housing, demographic structure, social services and infrastructure, economic activity, physical characteristics, etc. Each of these type of information was classified and presented in detail in the subcategories relevant to each area. For example, the data on demographic structure is divided into such sub-categories as total population, economically-active

population, school age population, population projection, etc. Similarly, economic activity data consists of employment structure, labour force participation, sectoral distribution of employment, etc.

The data in the OODB are organized in the form of tables which contain textual and quantitative tabulated information arranged and presented in a format suitable for ease of use and understanding.

The database has been developed using Unesco software Micro-CDS/ISIS version 3.0.

The FDT, the display format and sample records from the database PLAN is presented below.

Field Definition Table (FDT) Database: PLAN

?	Tag	Name	Len	Type	Rep	Del/Patt
	1	District	100	X		
	2	Town	100	X		
	3	Broad Subject	100	X		
	4	Specific Subject	100	X		
	5	Data	1000	X	R	

Database Name: Plan Format Name: Plan

```
Mhl, 'Record No.',c20,Mfn(4)/"District",c20,v1/"Town", c
20,v2/"Broad Subject",c20,v3/"Specific Subject",c20,v4##
c25,"DATA"##D5,(v5/)#
```

The records were first prepared in table forms in ASCII format using Wordperfect. They were then exported to ISO 2709 and downloaded to CDS/ISIS database.

Fields 1 District, 2 Town, 3 Broad Subject and 4 Specific Subject are indexed to facilitate information retrieval. The sample record retrieved from the database PLAN with the above display format in response to the query "housing" is given below.

Sample Record from PLAN Database

(Query: Housing. Display format: Plan.pft)

RECORD NO.	0002
DISTRICT	Goba-Robe
TOWN	Goba
BROAD SUBJECT	Tenure and Rents
SPECIFIC SUBJECT	Housing Tenure and Rents

D A T A

Housing Tenure and Rents

Tenure	Number	%
Owned	2,903	60.5
Rented	1,896	39.5
Total	4,799	100.0
Distribution of Rents (1988, NUPI Survey)		
Rentr Per Month in Birr		
5	1,164	66.6
5-9	284	16.2
10-29	189	10.5
30+	159	6.7
Total	1,896	100.0

This table data is taken from many records retrieved with the query mentioned above. For more records to the query see Annex 3.

8.4.3 Searches in the Databases

A search expression can be applied successively to each of the databases to retrieve different kinds of information records. For example, the query on:

Urban development planning

applied to the reference records in ABNCD will retrieve bibliographic records, profile of institution, completed or ongoing or proposed projects, and profile of experts in urban development and planning. The same search expression applied to the OODB can retrieve organized information relating to urban development and planning aspects, if any.

The query on: Housing characteristics in Akaki town
using Boolean expression AND (Housing * Akaki)

applied to the reference databases will retrieve bibliographic records, profiles of institutions, projects, and experts, all in or relating to housing in the town Akaki. The same query applied to the specialized

OODB, will retrieve all information relating to housing characteristics in Akaki in the form of data tables. The data from OODB gives organized information for direct use. For sample retrieval see Annex 5.

Information search from databases can be done in different ways: index-based, thesaurus assisted, free term or a combination of them. Use of Boolean and other operators, term adjacency, truncation, etc. are provided for by CDS/ISIS and thus can be applied in searching the databases.

8.4.4 Display of Information

Specialists in the urban planning field may require different forms of information/data display output with respect to the purpose the information is to serve. This can be provided by designing or creating different display formats based on the needs of end-users. For example, if a town planner is interested in housing tenure and rents of a particular town, this can be provided by excluding other fields from the display format or by defining a new one with the needed field(s) only. In our previous sample record from the database PLAN, for example, and with a new dispaly format, the housing tenure and rent information/data for Goba town

can be obtained as follows.

Database Name: Plan Format Name: Tenure

```
Mhl, "Town:",c10,v2##c25,"DATA"##D5,(v5/)#
```

This display format excludes information on the distric, broad subject and specific sabject, and retrieves only the information/data on the town Goba and the housing tenure and rent distribution pattern.

Retrieval Output:

TOWN: Goba

D A T A

Housing Tenure and Rents

Tenure	Number	%
Owned	2,903	60.5
Rented	1,896	39.5
Total	4,799	100.0
Distribution of Rents (1988, NUPI Survey)		
Rentr Per Month in Birr		
5	1,164	66.6
5-9	284	16.2
10-29	189	10.5
30+	159	6.7
Total	1,896	100.0

This makes the user more comfortable with the result of the searches in the databases. Flexibility of outputs of information search enables exclusion of irrelevant or less relevant information output and saves the time of the user.

8.5 USER-INTERFACE FACILITY

User-system interface is an important aspect of information retrieval. Different types of interface facilities to help users to interact with the information retrieval system may be designed. In this work, two user interface systems have been designed to enhance and facilitate retrieval: (1) a thesaurus-like search assistance facility and, (2) a structured query worksheet.

8.5.1 Thesaurus-Like Search Assistance Facility

A thesaurus-like search facility has been provided to facilitate navigating and browsing through the subject areas into which OODB has been divided and sub-divided (see 8.4). This can also be applied to the reference databases if the key terms are matched with the subdivisions of information using the MTHES facility for modifying and editing search expression(s). The MTHES pascal program (see chapter six) has been used to develop the interface. This pascal interface suggests, in the first screen, that the user enters the term

'DATABASE' in lower or upper case. To this, the system wil: respond with the display of the broad categories of field: into which the information content of urban planning ar: divided. This message is displayed below:

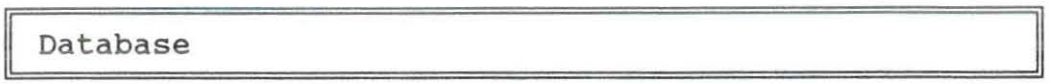
ASSISTED SEARCH FACILITY

USER INTERFACE SYSTEM IN URBAN PLANNING AND RELATED AREAS

1. When you press a key a blank screen will be displayed and you will be asked to enter a search term in the box at the top. Type in the term **DATABASE** in a lower or upper case. The broad areas into which urban planning information has been divided will be displayed. To position the cursor at the desired term, press the **<ENTER>** key as many times as necessary, e.g. **"Demographic.."** and press the **S** key to select it. Sub-divisions and concepts related to Demography will be displayed. Move the cursor to the desired term and press **S** key to select the term. Note the options at the bottom of the screen for formulating search queries.
2. You may also key-in the **term (or a stem of it)** representing your interest and press **<ENTER>** key; a list of terms beginning with the string you have keyed-in will be displayed. You may move the cursor to the desired term and press **S** key to select. Then proceed as indicated in para 1.

Press any key

When the user presses any key, the system displays a blank screen, as mentioned above, with an empty box where the term DATABASE should be entered. Entering the term 'database' in the box will display the following:




```
      Database
   SN Information is divided into the following broad
       searchable areas
S  NT Demographic information
   NT Economic activities information
   NT Financial and managerial information
   NT Historical information
   NT Housing information
   NT Maps and plans information
   NT Physical information
   NT Social infrastructure information
   NT Transport and communication information
```

```
◂┘Next [B]ack [F]irst [P]age [S]elect [T]erm select [M]ake file
  add te[R]m(file)    [C]reate term   [A]dd relation [D]elete
  [Q]uery [S]ave  [?] display  Query/saVe  search[/]add  e[X]it
```

Moving the cursor to the term Demographic..., for example, and selecting it by pressing the S key, the following will be the display:


```
      Demographic information
   SN Divided into following searchable areas
   BT Database
   NT Demography
   NT Migration
   NT Place of birth
   NT Population density
S  NT Population distribution
   NT Population dynamics
   NT Population projection
   NT Population structure
   NT Residence
   NT Population size
```

```
◂┘Next [B]ack [F]irst  P]age [S]elect [T]erm select [M]ake file
 add te[R]m(file)    [C]reate term   [A]dd relation    [D]elete
 [Q]uery [S]ave  [?] display  Query/saVe  search[/]add   e[X]it
```

Selection of 'Population distribution' will display the following list of searchable areas:

```
+-------------------------------------------------------------------+
| Population distribution                                           |
+-------------------------------------------------------------------+

     Population distribution
NT   Economically active population
NT   School age population
NT   Male population
NT   Children population
NT   Female population
NT   Youth population
RT   Demography
RT   Fertility
RT   Female
RT   Male

◂┘Next [B]ack [F]irst  P]age [S]elect [T]erm select [M]ake file
 add te[R]m(file)    [C]reate term   [A]dd relation [D]elete
 [Q]uery [S]ave  [?] display  Query/saVe  search[/]add  e[X]it
```

Options at the bottom of the screen enable the user to select the term(s) of interest to prepare search expressions. Moving the cursor to the term(s) of interest, pressing **Q** and after selecting all the terms representing the information need, in this way pressing **X** will **OR** together the selected terms in a search expression for a CDS/ISIS database(s). The terms and/or the operators can be edited/modified/changed before making the search by pressing the **<ENTER>** key.

This thesaurus-like search assistance contains terms/concepts selected from appropriate information

categories needed for the purpose of urban plan preparation as presented in chapter four in detail, and organized in the form of thesaurus with the application of thesaurus construction or development techniques.

The broad categories were based on the major types of information needs of those involved in urban planning. As indicated in para 2 of the first screen , instead of typing in the term '**database**' and going through the thesaurus like assistance, the user can key-in the term or a stem of it representing his/her interest and proceed further as described above for selecting search terms.

8.5.2 Structured User Query

One area of specialist user's difficulty in dealing with an information retrieval system is in filling general unstructured information request form. Users have difficulty in clearly understanding special terms in query worksheet. Concepts, such as, 'search topic', 'descriptor', 'keyword', 'sample citation', etc. are not usually familiar. If the users do not understand these terms, they may not provide a precise expression of their information need. The more specific the expression of user's information need, the more precise will be the retrieved records vis-a-vis the user's information need, that is, less noise or irrelevant information.

It will also be helpful if the users get guidance as to the coverage of the subject by the database, the broad categorization of the information used in the database, the more specific sub-division in each category. Such guidance may be provided by an information intermediary with the help of classification schemes, thesaurus, etc. for the subject field concerned.

Another possibility is to provide a query worksheet which indicates the divisions, sub-divisions, and even more specific topics on which information may be retrieved from the database. The users can simply mark in the query worksheet all the terms (concepts) of interest. This completed worksheet can form the basis for preparing search expression and retrieval.

The query worksheet proposed for use in the urban planning database is presented in Annex 4.

The information need areas of urban planning are given on the left side and the detailed searchable topics are given on the right side in a more or less hierarchical structure. The user can simply mark in the space provided to select the search terms of interest. Since a wide range of search topics are included, the user may comfortably work with the query worksheet without problems. This further gives the user more opportunity to

even select terms which he/she might not have had in mind which, however, may be relevant to his/her information needs. The worksheet is also open-ended so that a user can add other terms not included.

CHAPTER NINE

CONCLUSIONS AND RECOMMENDATIONS

In this thesis, the major focus was computer generation of vocabulary control tools and their application in designing urban planning databases with a suitable user interface facility. This chapter presents the major conclusions drawn from the study and some recommendations for the application of the result of the study.

9.1 CONCLUSION

Since the early 1980s, there has been renewed interest in the development and application of vocabulary control tools, particularly thesauri. This interest in thesaurus making led to the development of different thesaurus softwares that facilitated the design, development, and maintenance of thesaurus and its variations. Such work has been facilitated by the advances in microcomputers which made the tedious and time consuming task of thesaurus development easier and effective.

With the advance in new information technology (NIT) and the availability of various thesaurus programs, it is now possible to establish a complicated and more reliable hierarchical and associative term relationships in the

The principles and postulates of knowledge organization on which analytico-synthetic facetization is based, togehter form the guidelines for developing vocabulary control tools. Therefore, the tools can be used with advantage at the time of database and user interface design, and at the information retrieval stage.

The field urban planning is a specialized sectoral and multi-disciplinary subject field that is concerned with a variety of disciplines. Hence, the tasks of urban planning require experts from different areas of specialization who need a range of information from different sources as explained in chapter four. Urban planning is concerned with a specific geographical area, i.e. urban centre, and the type of information required for that activity is also specific to the area to be planned.

It is not possible to get such a variety and specific information from the international on-line commercial databases and search services in developing countries. Therefore, in order to fulfil the information need of urban planning, there is need to design and develop specialized local or institutional databases on the basis of the information requirements of planners, experts, researchers, executives, etc. engaged in urban planning activities.

With the sophistication of microcomputers, both in hardware and software, as well as the reduction of the cost of such technologies, it has become possible to create integrated database that contains different records, specifically, reference and specialized object-oriented database (OODB) records. Such system of database has been found appropriate to meet the information needs of specialized institutions like NUPI. The study indicated that the experts in NUPI are mostly interested in information specific to the purpose for which they need it. This implies that there is a need to design a database system that provides value added information, i.e. synthesized, analyzed and organized information. This type of information can be provided using specialized object-oriented database.

The design of specialized local databases can be assisted by the application of vocabulary control devices, especially, thesaurus with its variations, as mentioned earlier, and demonstrated in present work.

The design and use of local specialized database permits provision of flexible search outputs. The output can be presented in forms and formats preferred by the end-users.

One area of end-user difficulty is in interacting with information retrieval system. Vocabulary control tools , specifically thesaurus, is found helpful in developing an effective user interface facilities. The thesaurus-like search assistance facility presented is one of such possibilities.

Users of information system have difficulty in understanding the concepts or terms in the query worksheet. It is found that design of structured query worksheet, based on the broad and sub-categorization of information needed in a given field of concern, is helpful in mitigating difficulty of end-users. Users have shown interest in and satisfaction with this type of query in selecting search topics of their interest from the detailed list of search topics provided by the structured query as against the more general query. Such structured query worksheet can also indicate to end-users search topics on which they may need information but might not have had in mind before. Thesaurus systems can help in the design of structured query by providing the substantive terms or subject concepts on which information may be organized and their systematic and hierarchical arrangement.

The function of urban planning is essentially a decision making process and effective decision cannot be made without the use of pertinent information. This necessitates the existence of effective information support system. The result of this study, i.e the generated vocabulary control tools and their application to databases in urban planning, can help in this regard.

In general, the principles and techniques of designing vocabulary control tools and the tools themselves have wider application than what they normally used for. This has been demonstrated in this work.

9.2 RECOMMENDATIONS

The system component designed and developed in this work is for application in NUPI to support its planning activities. In order to apply or implement the results, certain preconditions require to be fulfilled.

(1) The system components (the thesauri, the prototype databases, and the user interface) are developed using thesaurus program specifically designed to work with Micro-CDS/ISIS software. Hence, the application of results of the study needs the acquisition and

installation of the package in NUPI with the thesaurus programs. This package is freely distributed by Unesco and there is no problem in getting it. The latest version 3.0 is recommended.

(2) In this thesis, it was mentioned that NUPI has a computing facility which is underutilized. However, this underutilization of computing resource largely relates to the main frame. The available micros are mostly busy with word processing and they cannot fully support the design of local database system in NUPI. Thus, it is necessary to make available additional microcomputers that can support the automation of the information system of NUPI.

(3) Arrangement for training should be made. This includes training on the use of CDS/ISIS package and the application of the system component developed in this work. Priority should be established so that the information personnel in NUPI, i.e those in DIC and CC are trained first who will then orient end-users on using the system.

(4) The existing information system of NUPI is not integrated. The data processing function and the information provision functions are separately organized and have no meaningful supportive relations. In reality,

these functions are closely related both dealing with information services provision for planning activities of NUPI. Since the activity of urban planning needs a coordinated, effective and well integrated information support system, the existing information system of NUPI should be reorganized and integrated. DIC and CC are recommended to come under a single supervision so as to be able to operate in a manner supporting each other. This unification of the two separate units will create a favourable situation for using the information technology in CC for automating the functions of DIC. It is recommended that the new organization of the information system of NUPI shall take the name Information Service Department or Centre (ISD or ISC). The head of the department, designated information officer (IO), should be responsible to the Deputy General Manger of NUPI or directly to the General Manager. This will place the IO in a position to help with information service the top officials as well as all other users in NUPI.

(5) Together with the integration and reorganization of the information system of NUPI, effort should be made to develop the national urban development information support system (NUDISS) with the use of the system designed and developed in this work. NUDISS shall serve as a central information service system for the country's

urban system.

(6) The value of information is being recognized in an increasing measure in all institutions. It is, therefore, recommended that NUPI makes every effort to market value-added information packages and services on a payment basis to other institutions especially commercial organizations.

(7) An effective information network should be established interlinking NUPI with apprpriate institutions to enable timely and coordinated information exchange and provision of timely and precise data and information to all those need it.

Annex 2: Facetized category of Terms taken from users queries (Note: Terms marked with an asterisk are those added to Thesb database).

Basic Disciplines

Regional planning
.Urban planning
=Spatial planning
=Physical planning
=Town planning
=City planning

Property Terms

Access
Accident
.Traffic accident
Affordability*
.Local affordability*
Age
Altitutde
Climate
.Urban climate
Condition
.Housing condition
Congestion*
Constraint
Culture
Degradation
Ecology
Exchange
.Regionl exchange
.Urban-rural exchange
Feature*
Fertility
Form
.Urban form*
Frequency*
Growth
Hazards
.Natural hazards
Health
Hierarchy
Impact
Intensity*
Methods
Morphology*
Mortality
Policy
Pollution
Poverty*
Property.. Cont'd.
Pressure*
Problem
Relief*
Sanitation
.Urban sanitation
Sex
Slope*
Standard
Suitability*
Technique
Temperature
Topography
Trend

Entity Terms

Airway*
Area
.Rural area
.Slum area
.Urban area
Building
Centre
.Service centre
.Urban center
City
Data
.Demographic data
Design
.Road design
.Spatial design
Environment
Expenditure
.Municipal expenditure
Finance
.Municipal finance
.Public finance
Flood*
Government
.Municipal government
.Local government
Hinterland*
Household
Housing
.Traditional housing
Infrastructure
.Social infrastructure
Junction*
Land
.Nonusable land
Map

.Land cover map
.Social service map
.Thematic map
Transport
.Freight transport
.Public transport
.Topography map
Markets
.Open markets
Model
Network
.Road network
Plan
.Action plan
.Master plan
.Urban plan
Population
Population pyramid*
Property
Railway
Real estate
.Urban real estate
Region
.Planning region
Resource
Revenue
.Urban revenue
.Municipal revenue
Road
=Steet
=Path
.Arterial road*
..Main arterial road*
.One way road*
.Two way road*
.Local road*
Sector
Service
Settlement
.Squatter settlment
Standards*
System
.Cadastral system*
.Sewerage system
Tax
Town
.Shanty town
Urban centre
Utility lines

<u>Action Terms</u>

Activity
.Informal activity*
Adjusment
Administration
Allocation
Analysis
.Slum analysis*
.Suitability analysis*
Application*
.Cadastral application*
Assessment
Conservation
Coordination
.Horizontal Coordination*
.Vertical coordination*
Cultivating*
Development
Designing
.Road designing
.Spatial designing
Distribution
Education
Evaluation
Exchanging
.Regionl exchanging
.Urban-rural exchanging
Function
Growth
Implementation
Improving
Management
Migration
Native planning
Permiting
Planning
Preparation
Protection
Registration
Sampling
Square
Study
Supervision
Supplying
Trading
Transportation
Transportation
Upgrading
Urbanization
Use
.Land Use
..Agricultural use
..Industrial use
..Institutional use
..Residential use
Utilization
Zonning

Annex 3: Sample Records from PLAN Database
(Query: Housing. Display format: Plan.pft)

RECORD NO. 0001
DISTRICT Goba-Robe
TOWN Goba
BROAD SUBJECT Housing
SPECIFIC SUBJECT Housing Units & Physical Characteristics

D A T A

Housing Units & Physical Characteristics

Item	Number	Percent
Housing Units		
Residential	5,963	88.7
Residential and Establishment	763	11.3
Total	6,726	100.0
House Type:		
Single Story Detached	4,640	69.0
Construction Materials		
Walls- Wood and Mud	5,715	85.0
Roof- Corrugated iron sheet	5,796	86.2
Foor-Earth/mud	4,507	67.0
Provision of Infrastructure:		
Water Supply- Tap in-compound	239	5.0
Tap ex-Compound	8	0.2
Other	4,552	94.8
Total	4,799	100.0
Sanitation - No toilet	2,773	57.8
Electricity - Yes	2,428	50.6
Kitchen - Non	1,812	37.8

RECORD NO. 0002
DISTRICT Goba-Robe
TOWN Goba
BROAD SUBJECT Tenure and Rents
SPECIFIC SUBJECT Housing Tenure and Rents

D A T A

Housing Tenure and Rents

Tenure	Number	Percent
Owned	2,903	60.5
Rented	12892	39.5
Total	4,799	100.0
Distribution of Rents (1988, NUPI Survey)		
Rentr Per Month in Birr		
5	1,164	66.6
5-9	284	16.2
10-29	189	10.5
30+	159	6.7

RECORD NO. 0012
DISTRICT Goba-Robe
TOWN Goba
BROAD SUBJECT Households
SPECIFIC SUBJECT Household Paramter

D A T A

Household size and Occupancy Rates

Item	Number	
Average household size	4.4	
Household per housing unit	1.1	
Persons per housing unit	4.8	
Average rooms per housing unit	2.7	
Distribution of housing unit by number of rooms:		
1 room	1,128	23.5
2-3 rooms	2,712	56.5
4+ rooms	959	20.0
Total	4,799	100.0

Record No. 0012 continued

Households with more than 2.4 persons/room	1,509	31.4
Application for rented housing (1984-88 avragre)	311	-
New house completed since 1984	134	-
Newly registered application to build	57	-

RECORD NO. 0015
DISTRICT Goba-Robe
TOWN Goba
BROAD SUBJECT District Information
SPECIFIC SUBJECT Population Density and Land Holding, 1988

D A T A

Population Density and Land Holding, 1988.

Over all density	42 persons/km2 (Total area)
Net Rural density	38 rural population/km2 (developable area)
Average landholding	0.7 hectars cultivated/rural household

RECORD NO. 0020
DISTRICT Goba-Robe
TOWN Goba
BROAD SUBJECT Household (HH) information
SPECIFIC SUBJECT Household data, 1984

D A T A

Household data, 1984

Total Number	5,282 households	
Average size	4.4 persons per household	
Distribution: No. of Persons/HH	Percent of Households	Percent of Population
1-2 persons	30.1	10.5
3-6 persons	51.2	52.2
7+ persons	18.7	37.3
Total	100.0	100.0

RECORD NO. 0021
DISTRICT Goba-Robe
TOWN Goba
BROAD SUBJECT Population and Household
SPECIFIC SUBJECT Population & Household Projections

D A T A

Population & Household Projections

Item\Yr:	1984	1988	1989	1990	1991	1993	1998
Popul'n:	23,054	28,600	30,200	31,900	33,700	37,500	49,000
HHs	5,282	6,500	6,920	7,310	7,720	8,590	1,20:

Assumptions: Growth rates 1984-1993 5.5 per annum;
1993-1998 5.5 per annum
Average household size 4.4 persons/household
(held constant)

RECORD NO. 0027
DISTRICT Akaki
TOWN Akaki
BROAD SUBJECT District Information
SPECIFIC SUBJECT Population Density and Land Holding

D A T A

Table 23: Density and Land Holding, 1984

Over all density	139 persons/km2 (Total area)
Net Rural density	76 rural population/km2 (developable area)
Average landholding	1.4 hrs cultivated/rural household

RECORD NO. 0033
DISTRICT Ziway
TOWN Ziway
BROAD SUBJECT Expenditure
SPECIFIC SUBJECT Expenditure Distribution

D A T A

Table 29: Expenditure Distribution
Average expenditures are distributed as follows:

Item	Percent
Food and Beverages	40.6
Rent/Housing	3.6
Other Shelter costs (Building materials, fuel, electrivity, water	10.9
Tax on land	0.1
Savinds	30.3
Other	14.6

RECORD NO. 0034
DISTRICT Ziway
TOWN Ziway
BROAD SUBJECT Household
SPECIFIC SUBJECT Projected Household Demand

D A T A

Table 30: Household Demand Projection

Item\Year	1984	1988	1993	Total
Households	1143	1490	2,080	2080
Household increase		347	590	937
New Housing Required		324	551	876
Construction, 1984-88 (Units)		626	-	626
Residential Demand (1988-93) (Units)		-	-	250

Annex 4: Structured Query Worksheet

NATIONAL URBAN PLANNING INSTITUTE
INFORMATION SERVICES CENTRE
STRUCTURED INFORMATION SEARCH REQUEST FORM

NOTE: Make a tick mark in the space provided opposite the searchable topic of your interest.

Category of Information	Search Topics
1. Physical Characteristics Information	[ ] Location [ ] Altitude [ ] Topology [ ] Geology [ ] Topography [ ] land use pattern [ ] Meteorology ( ) Temperature ( ) Humidity ( ) Rainfall [ ] Natural Resources ( ) Water ( ) Forest ( ) Mineral ( ) Soil Other: _____ Other: _____
2. Demographic Information	[ ] Population size [ ] Population Distribution ( ) By age ( ) By sex [ ] Population Density [ ] Population dynamics ( ) Fertility ( ) Birth rate ( ) Death rate [ ] Population projection [ ] Population Structure [ ] Migration ( ) Rural-urban ( ) Urban-urban ( ) Attraction (pull/push factors) [ ] Birth place ( ) Rural ( ) Urban ( ) Region [ ] Length of Residence data Other: _____
3. Housing Information	[ ] Housing size [ ] Age of housing units [ ] Housing tenure ( ) Owned ( ) Rented [ ] Rent distribution [ ] Purpose of housing ( ) Residential ( ) Business ( ) Residential and Business [ ] Typology ( ) Unit type ( ) Building type ( ) No. of Rooms ( ) Service quarter

Continued on page 2

Category of Information	Search Topics
(Housing... cont'd.)	[ ] Housing structure/material ( ) Foundation ( ) flooring ( ) Walling ( ) Kitchen ( ) Roofing ( ) Ceiling [ ] Housing services ( ) Water ( ) Bathing ( ) Toilet ( ) Lighting ( ) Telephone ( ) Fuel/energy ( ) Garbage collection [ ] Housing Demand [ ] Housing supply [ ] Construction plan [ ] Household size ( ) Total size ( ) Average size ( ) Occupancy rate Other: _____
4. Transport and Communication Information	[ ] Origin-destination [ ] Private transport [ ] Public transport [ ] Road network [ ] Pedestrian [ ] Route capacity [ ] Mode of transport Other: _____
5. Social Infrastructure Information	[ ] Education ( ) School type ( ) Enrolment ratio ( ) Student class ration ( ) Eligible population size ( ) Demand for educational facilities [ ] Health facilities ( ) Hospitals ( ) Clinics ( ) Health stations ( ) Pharmacies [ ] Religious facilities [ ] Libraries [ ] Recreational facilities ( ) Sports centres ( ) Cinema/theatre halls Others (specify): _____ Other: _____
6. Maps and Plans Information	[ ] Maps ( ) Original ( ) Semioriginal Subject _____ Scale _____ [ ] Plans Subject _____ Scale _____ Type _____ Other: _____

Continued on page 3

Category of Information	Search Topics
(Housing... cont'd.)	[ ] Housing structure/material ( ) Foundation ( ) flooring ( ) Walling ( ) Kitchen ( ) Roofing ( ) Ceiling [ ] Housing services ( ) Water ( ) Bathing ( ) Toilet ( ) Lighting ( ) Telephone ( ) Fuel/energy ( ) Garbage collection [ ] Housing Demand [ ] Housing supply [ ] Construction plan [ ] Household size ( ) Total size ( ) Average size ( ) Occupancy rate Other: _____
4. Transport and Communication Information	[ ] Origin-destination [ ] Private transport [ ] Public transport [ ] Road network [ ] Pedestrian [ ] Route capacity [ ] Mode of transport Other: _____
5. Social Infrastructure Information	[ ] Education ( ) School type ( ) Enrolment ratio ( ) Student class ration ( ) Eligible population size ( ) Demand for educational facilities [ ] Health facilities ( ) Hospitals ( ) Clinics ( ) Health stations ( ) Pharmacies [ ] Religious facilities [ ] Libraries [ ] Recreational facilities ( ) Sports centres ( ) Cinema/theatre halls Others (specify): _____ Other: _____
6. Maps and Plans Information	[ ] Maps ( ) Original ( ) Semioriginal Subject _____ Scale _____ [ ] Plans Subject _____ Scale _____ Type _____ Other: _____

Continued on page 3

Annex 5: Sample Records fron Urban Integrated Database (ABNCD+)

*** PROFILE OF INSTITUTION ***

TRANS. NAME	National Urban Planning Institute (NUPI), Addis Ababa, Ethiopia
INSTITUTION	Biherawi Ye Ketemoch Plan Institute, BKPI, Addis Ababa, Ethiopia
START DT	1987-06-06
LOCATION	K15K30Kaz.
ADDRESS	Addis Ababa 2405 Addis Ababa Ethiopia
PHONE	15-19-68
TELEX	21551 NUPI ET
WORK.LANG.	EngAmharic
HEAD	Mekonnen, Mulat (Eng.), General Manager
GEOG.COVERAGE	National
INST. TYPE	Governmental, Planning and Research
ASSOC.ENTITIES	UDPO; Special Housing Project; City Councils; UNDP; ILO
OBJECTIVES	(1) Prepares Urban plans of various kinds, (2) Conducts research and studies that are necessary for the preparation of urban plans and (3) Training manpower required for the preparation and implementation of urban cenre, metropolitan and regional plans.
ACTIVITIES	Various Urban planning and development projects
DESCRIPTORS	Urban; Planning; Research; Ethipia; Urban Planning; Training; Consultancy
INFO. SERVICE	Data processing and dissemination; Library and documentation Services
DATABASES	Databases of various data collected through field surveys and processed for the purpose of urban planning and other related studies are available
CLASSIFIC.	DDC
NOTES	The National Urban Planning is being reorganized to adjust itself to the current Structural Adjustment Policy of the current Ethiopian Government.
FINANCE	Government budget
EQUIPMENT	HP Mainframe Computer; IBM Compatible Micros; Telex and Telefax Equipment; and Other office equipment

*** PROFILE OF INSTITUTION ***

TRANS. NAME	Agency for the Administration of Rented Houses (AARH)
INSTITUTION	Ye Kiray Betoch Astedader Dirijjit
START DT	1975-10-04
LOCATION	K15K30Kaz.
ADDRESS	Addis Ababa 299 Addis Ababa Ethiopia
WORK.LANG.	Amharic
HEAD	Yaa'ii, Tullama, General Manager
GEOG.COVERAGE	National
INST. TYPE	Governmental, Service Rendering
ASSOC.ENTITIES	Ministry of Works and Urban Development, National Urban Planning Institute, Urban Development Project Office, Dweling Houses Construction Enterprize, City Councils, C onstruction Materials Supply Enterprize,
OBJECTIVES	(1) To rent houses, (2) To render housing maintenance services, (3) To collect housing rents, (4) to administer government houses rented for Birr 100 and above, (5) To get housing construction projects studied, (6) to estimate and establish housing rents, and (7) To get built houses for rent
ACTIVITIES	Special housing construction project; Bole Road Appartment Construction; Construction of different economic houses; Ground plas one (G+1) projects, etc.
DESCRIPTORS	Urban; Rented houses; Houses; Agency; Administration; urban houses; Ethiopia
INFO. SERVICE	Library Services
CLASSIFIC.	DDC
FINANCE	Housing Rents

*** PROFILE OF INSTITUTIONS ***

INSTITUTION	Kenya Industrial and Development Institute (KIRDI)
START DT	1979
LOCATION	Nairobi
ADDRESS	Kenya Industrial Research and Development Institute P.O.Box 30650 Nairobi
PHONE	22265
PERSONNEL	Specialists in various science and technology disciplines
HEAD	Arunga, Director
INST. TYPE	Research institute
OBJECTIVES	1) to conduct research and development in industrial and allied technologies including engineering, commodity technologies, mining and power resources development
ACTIVITIES	1) identification and developing appropriate process and product technologies to serve the local market and export potential; 2) designing, developing and adapting machinery, tools, equipment and instruments and process suitable for introduction in the rural areas; 3) developing suitable treatment/recovery processes and services to reduce environmental hazards created by industrial wastes and effluents; 4) setting up pilot plants where necessary to demonstrate the efficacy of industrial technology5)provision of industrial information and technical services
DESCRIPTORS	Industrial information; industrial development; industrial technology; industrialists; research; environmental hazards
INFO. SERVICE	Consultancy services to industrialists and potential industrialists
FINANCE	Kenya Government Donors
PERIODICALS	Annual report. KIRDI

*** PROFILE OF EXPERT ***

NAME	HADGU, BARIAGABIR, DR.
NATIONALITY	ET
QUALIFIC.	Statistics. B. A. Addis Ababa University. 1971. University of Ghana (UN Centre). M. A. Demography. 1975. Demography. Ph. D. University of Ghana (UN Centre). 1992.
SPECIALIZATION	Population and Social Affairs Planning and Development.
LANG. COMP.	Tigrigna. Amharic. English.

Employment Record

CURR. EMPLOY.	Acting Deputy General Manager, National Urban Planning Institute.
LAST EMPLOY.	Teacher, Ministry of Education, 1963-66/1971-72. Team leader, C. S. A., 1975-1983. Senior Expert, Natianal Urban Planning Institute, 1984-1992
HONOURS:	Fellowship, UN Regional Institute for Population Studies, 1973-75. Fellowship, UN Regional Institute for Population Studies, 1973-75.

*** PROFILE OF EXPERT ***

NAME	EYOB, DOLICHO
NATIONALITY	ET
QUALIFIC.	Social Work. B.A. Addis Ababa University. 1978. Housing. Post Graduate Diploma. 1985. Management. Certificate. 1990.
SPECIALIZATION	Housing.
LANG. COMP.	Amharic. English. Hadiya. Sidama. Guragee, etc.

Employment Record

CURR. EMPLOY.	Head of Housing Cooperative division, Ministry of Urban Development and Housing, Full time.
LAST EMPLOY.	Teaching (Teacher), National Public School, 1972-73. Housing Expert, Ministry of Urban Development and Housing, 1974-present.
HONOURS:	Further Studies, Institute of Housing Studies (Rotterdom, the Netherlands), Jan. 1993.

*** PROFILE OF EXPERT ***

NAME	MATHEWOS, ASFAW, (ATO)
NATIONALITY	ET
QUALIFIC.	Architecture and Town Planning. M.Sc. Addis Ababa University. 1986.
SPECIALIZATION	Town Planning.
LANG. COMP>	Amharic. Oromiffaa. English.

Employment Record

CURR. EMPLOY.	Town Planning, National Urban Planning Institute, Full time.

*** PROFILE OF EXPERT ***

NAME	TILAHUN, FEKADE, (ATO)
NATIONALITY	ET
QUALIFIC.	Agricultural Economics. B. Sc. Alemaya University of Agriculture. 1987. Regional Development Planning. M. A. institute of Development Studies. 1991.
SPECIALIZATION	Regional Development Planning. Economics.
LANG. COMP	Amharic. English.

Employment Record

CURR. EMPLOY.	Research and planning in urban planning and and related areas.
LAST EMPLOY.	Researcher and Planner, National Urban Planning Institute, 1988 to present.

*** PROFILE OF EXPERT ***

NAME	BIRKE, YAMI, (WEIZERO)
NATIONALITY	ET
QUALIFIC.	Architecture and Town Plnning. B. Sc. Addis Ababa University. 1987. Town Plnning. M. Sc. Helsinki University of Technology. 1992.
SPECIALIZATION	Urban design, Town Planning and Architecture.
LANG.COMP.	Oromiffaa. Amharic. English.

Employment Record

CURR. EMPLOY.	Architect-Town Planner. cNational Urban Planning Institute, 1987 to present.

*** PROFILE OF EXPERT ***

NAME	DANIEL, LEGESSE, (ATO)
NATIONALITY	ET
QUALIFIC.	Sociology. Addis Ababa University. 1982.
SPECIALIZATION	Social Affairs Studies.
LANG. COMP.	Amharic. English.

Employment Record

CURR. EMPLOY.	Social Service Coordinator, UDPO, full time.
LAST EMPLOY.	Social Work Expert, Hararghee Prison Administration, 1982-1984.

*** PROFILE OF EXPERT ***

NAME	FELEKE, ENGIDASHET, (ATO)
NATIONALITY	ET
QUALIFIC.	Law. L. L. B. Addis Ababa University (Lae Faculty). 1987.
SPECIALIZATION	Legan Advise.
LANG. COMP.	Amharic. Oromiffaa. English. Russian.

Employment Record

CURR. EMPLOY.	Legal Advise, National Urban Planning Institute, full time.

*** DEVELOPMENT PROJECT ***

INSTITUTION	National Urban Plan Institute, NUPI, Ethiopia, PreparationDire Dawa City Council, DDCC, Implementer
ADDRESS	Addis Ababa Kazabchis Addis Ababa 2405
PHONE	15-19-68
WORK.LANG.	English
GEOG.COVERAGE	Local
INST. TYPE	Governmental, Planning And Research
OBJECTIVES	To prepare Master plan of the Town particularly Land use plan
DESCRIPTORS	Dire Dawa; Master plan; Urban plan; Ethiopia; Urban planning;
FINANCE	Government budget

*** DEVELOPMENT PROJECT ***

INSTITUTION	National Urban Planning Institute, NUPI, Preparation

START DT	1992-10-01
COMPLETED	1993-03-31
DURATION	6 Months
ADDRESS	Addis Ababa Kazabchis Addis Ababa 2405
PHONE	15-19-68
WORK.LANG.	English
GEOG.COVERAGE	Socal/Regional
INST. TYPE	Governmental, Planning and Research
OBJECTIVES	To assess the short term economic and social impact of privatization, to learn from the experience of that particular region and propose policy recommendations for national privatization of rented houses in Ethiopia.
DESCRIPTORS	Economics; Social; Rented houses; housing; Privatization; Tigray; Mekele; urban Planning; Ethiopia

Record no. 95

Social welfare and urbanization in Africa / ECA, Addis Ababa, ET Standing Committee on Social Welfare and Community Development, 1st session, Addis Ababa ET, 5-13 Feb 1962 7 Feb 1962. 25p.

Document no.: E/CN.14/SWCD/8 T, 00-80-01055.
Descriptors: /SOCIAL SERVICES*/, /URBANIZATION*/, /WOMEN/, /UNEMPLOYMENT/, /YOUTH/, /CULTURAL CHANGE/, /AFRICA/; A10

Points out some of the social problems that are aggravated by urban life in Africa and the activities of the social welfare services that are trying to solve them.

Record no. 96

Social aspects of economic development / ECA, Addis Ababa, ET Standing Committee on Social Welfare and Community Development, 1st session, Addis Ababa ET, 5-13 Feb 1966, 9 Jan 1962. 28p.

Document no.: E/CN.14/SWCD/2 T, 00-80-01057.

Descriptors: /SOCIAL ASPECTS*/, /ECONOMIC DEVELOPMENT*/, /SOCIAL STRUCTURE/, /AGRARIAN STRUCTURE/, /URBANIZATION/, /OBSTACLES TO DEVELOPMENT/, /RURAL DEVELOPMENT/, /INDUSTRIALIZATION/; F40

Traces certain problems in broad outline with a view to the subsequent development and organization of social research.

Record no. 98

Use of population data in development planning in Africa / ECA, Addis Ababa, ET. Joint conference of African planners, statisticians and demographers, 2nd session, Addis Ababa ET, 8-17 Mar 1982 10 Dec 1981. 33p. : tables, illus.

Document no.: ST/ECA/PSD.2/3 T, 00-81-00430.
Descriptors: /DEVELOPMENT PLANNING*/, /DEMOGRAPHY*/, /POPULATION/, /DATA ANALYSIS/, /EDUCATIONAL PLANNING/, /HEALTH PLANNING/, /URBAN PLANNING/, /STATISTICAL DATA/, /AFRICA/; A10

Reviews the use of population data in African development plans and the integration of demographic variables in African development planning and analysis. Examines the crucial role population factors play in educational, health and urban planning with illustrations on the application of population variables.

Record no. 100

Report on the general situation concerning the establishment of multinational interdisciplinary development advisory teams / ECA, Addis Ababa, ET (1972). 6+24p.
Document no.: E/CN.14/ECO/45 T, 00-82-00103.
Descriptors: /DEVELOPMENT PROJECTS*/, /TECHNICAL ASSISTANCE*/, /WORK PROGRAMME*/, /EXPERTS/, /TRAINING/, /EMPLOYMENT/, /RURAL MIGRATIONS/, /URBAN DEVELOPMENT/, /ROAD CONSTRUCTION/, /MAINTENANCE AND REPAIR/, /FINANCING/, /REGIONAL COOPERATION/, /AFRICA/; C10

Evaluates the general picture of the projects for the establishment of Multinational Interdisciplinary Development Advisory Teams in Africa. Indicates the responsibilities to be assigned to the ECA sub-regional offices and the working relations which must exist between these offices and the other technical assistance missions, in particular the United Nations Resident Representatives.

Record no. 101

Demographic disparities among five major urban areas in Ethiopia / Hadgu Bariagaber; Addis Ababa Master Plan Project Office, Addis Ababa; ET Workshop on Urban Basic Services, Nazareth ET, 4-8 Oct 1983 In: Report on the Workshop on Urban Basic Services / Ministry of Urban Development and Housing. UNICEF, New York, ETUS 1983. p.103-131 : tables.

Descriptors: /DEMOGRAPHIC ANALYSIS*/, /URBAN AREAS*/, /URBANIZATION/, /URBAN POPULATION/, /FERTILITY/, /POPULATION DYNAMICS/, /MORTALITY/, /INFANT MORTALITY/, /POPULATION INCREASE/, /AGE-SEX DISTRIBUTION/, /LITERACY/, /ETHIOPIA/; A30

Describes the level of urbanization, taking the five major urban areas (Addis Ababa, Dire Dawa, Dessie, Nazareth and Jimma) in Ethiopia; assesses disparities in the dynamics of population, the disparities in general demographic characteristics, age-sex structure, economic activity, marital status, literacy status and educational attainment of the population.

Record no. 102

Summaries of the major features of the four selected towns and programme areas in basic services: (based on reports of the city councils) / Dessie City Council, Dessie, Nazareth City Council, Nazareth, Dire Dawa City Council, Dire Dawa, Jimma City Council, Jimma, ETETETET 1983. p.133-136

Descriptors: /URBAN AREAS*/, /SLUMS/, /DAY CARE CENTRES/, /HEALTH SERVICES/, /ORPHANAGES/, /FLOODS/, /FLOOD CONTROL/, /SLAUGHTERHOUSES/, /HOSPITALS/, /PHARMACISTS/, /FEMALE MANPOWER/, /ETHIOPIA/; A10

Record no. 1089

Call No.: 63 "19" SEL/Sel
Development and its impact on the environment; ONC / Tekheste Ahderom, National Urban Planning Institute, Addis Ababa ET, 22-25 May 1990, Con.

Descriptors: /URBAN ENVIRONMENT//WASTES//ENVIRONMENTAL EFFECTS//ETHIOPIA/#TH

Paper focuses on the trends of Urban environmental issues in present day Ethiopia. It touches on one aspect of Environmental planning and Urban management tools, while it dwells a great deal on Urban waste disposal and pollution problems with special emphasis on air pollution, sewerage, water quality, agricultural waste and noise pollution. The Urban-rural interaction is briefy discussed with a view to demonstrating the linkages. A few examples related to Environmental impact factors are given, based on the experience of the Addis Ababa Master Plan Project as well as that of the recently established Ethiopian National Urban Planning Institute (NUPI). The paper strongly recommends a rigorous application of Urban Environment Impact Assessments under large significant projects, especially in environmentally sensitive areas of the country.

Annex 6: Information Search Request Form

NOTE: Query information is among the set of data required for the preparation of my Master Thesis. Even though other alternative ways of securing requests for information search are available, I decided to use the actual query from the actual information users for reasons of objectivity and reliability. Therefore, I kindly request you, as one of the actual user of information in the field of urban planning and related areas, to fill out this search request form the type of information you want searched for you. The search result will be sent back to you for relevance judge vis-a-vis your information need satisfaction. Your kind co-operation is **invaluable** for the success of my thesis project.

Kindly, Teshome Bula
SISA, Addis Ababa University

---

1. **Requester's name:** _____
2. **Organization:** _____
3. **Mailing address:**
   - 3.1 Business: _____
   - 3.2 Residence: _____
4. **Educational qualification:** _____
5. **Field of specialization:** _____
6. **Title of post:** _____

7. **Search topic:** (State, in narrative form, the topic or subject on which you want a search)

_____
_____
_____
_____
_____
_____

8. **Sample citations:** (Give below at least two document known to you that are relevant to the topic or subject of the search)

_____

_____

_____

9. **Key words (Descriptors):** (List or give key terms, phrases, or concepts that describe your topic or subject of the search)

_____

_____

_____

10. Please state the **purpose** that the search will serve (e.g. Research project, for Planning, problem identification, etc.)

_____

11. Please tick one of the following that describes your **expectations** from the search on your topic or subject:

[ ] Want my search to be quite comprehensive in its coverage. This means that I would like to get output which will contain some items that may not be very relevant to the search topic.

[ ] I want my search output to be quite specific in its coverage. This means that I would like to get output that is mostly specific to the search topic.

Annex 7: A questionnaire for collecting Documents Descriptions

INFORMATION ON DOCUMENTARY MATERIALS IN THE FIELD OF URBAN PLANNING AND RELATED AREAS

1. Country of origin: _____
2. Record entered on: _____ Changed on: _____
3. Bibliographic level (M for monograph; A for Analytic): _____
4. Bibliographic level of parent (M for Monograph; S for serial): _____
5. MFN of Parent: _____ 6. MFNs of parts: _____
7. MFNs of other language versions: _____
8. Langauge of analysis: English
9. Language of text: _____
10. Language(s) of summaries/abstracts: _____
11. Title (in Language of Original):

    _____
12. Parallel Title;

    _____
13. Translated Title - English (If different from 11):

    _____
14. Personal Author(s) (a. Surname, Forename; b) Role):

    _____
15. Corporate Author(s) (a. Parent body, b. Sub body, c. Place, d. Country code, e. Role):

    _____

    _____

    _____
16. Affiliation of Personal Author(s) (a. Parent body, b. Sub-body, c. Place, d. Country code)

    _____

    _____

    _____

17. Associated Institutions (a. parent body, b. Sub-body, c. Place, d. Country code, . Role):

_____

_____

18. Meeting (a. Name and No., b. Place, c. Country code, d. Dates, e. Dates in ISO form:

_____

_____

19. Edition: _____

20. Publisher(s) (a. Place, b. Publisher, c. Country code): _____

21. Date of Publication/Issue: Free form: _____

ISO Form: _____

22. Collation (M/C) (a. No. of pages, b. Description, c. Dimension): _____

23. Part statement: (a. Vol./issue, b. pagination);

_____

24. Series statement (a. Title, b. Part no.):

_____

25. Thesis (a. Thesis, b. Degree, c. Institution/course, d. Year): _____

_____

26. Related project:

_____

27. Notes: _____

28. ISBN: _____ 29. Document No. _____

30. Availability: _____

FIELD FOR DESCRIBING PARENT ITEM

31. Title of serial:

_____

32. ISSN: _____

33. Title of parent Monograph:

_____

34. Personal Author(s) of Parent (see 14):

35. Corporate Author(s) of parent (see 15);

36. Meeting (as author of parent) (see 18):

SUBJECT ANALYSIS FIELD

37. Primary descriptors (e.g. <PLANNING> <URBAN ANALYSIS>):

38. Secondary descriptors (e.g. BIBLIOGRAPHY);

39. Geographic descriptors (e.g. Southern Showa, Ethiopia):

40. Local descriptors:

41. Proposed descriptors:

42. Abstract (Tick here if abstract is attached):

43. Broad subject headings:

Annex 8: A Form to Collect Information on Profile of Experts

INFORMATION ON EXPERTS IN THE FIELD OF URBAN PLANNING AND OTHER RELATED AREAS

1. Name of person: _____
2. Sex: _____
3. Current Employer: _____
4. Mailing Address:
   a) Region: _____ b) Locality: _____
   c) Town: _____ d) Box: _____
   e) Country: _____
   f) Phone: i) Home_____ ii) Office_____
   g) Telex: _____ h) Cable: _____
   i) Fax: _____ J) E-mail: _____
5. Designation: _____
6. Nationality: _____
7. Formal educational qualification:
   (d. degree/diploma, i. institution, y. year)

	Institution	Year obtained
Diploma	_____	_____
1st Degree	_____	_____
2nd Degree	_____	_____
3rd Degree	_____	_____

8. Language Competence (List names of the Languages):
   _____
   _____
9. Main working Language(s): _____

10. Work experience (in years of service to the nearest year):

a) In the current employer: _____

b) In previous employers:

i) Employer (Organization) _____

ii) Title/post: _____

iii) Duration:

From: _____ To: _____

11. Current work: _____

12. Project title(s): _____

13. Project recommended by: _____

14. Honours and Awards obtained:

i) Granting Institution: _____

ii) Date obtained: _____

iii) Type of Grant: _____

15. Membership in Associations:

_____

_____

16. Marital status: _____

17. Field of specialization: _____

_____

(To be filled by the investigator)

18. Person(s) entering data: _____

19. Remarks/Notes: _____

_____

_____

Annex 9: A Form to Collect Information on Profile of Institutions

INFORMATION ON INSTITUTIONS INVOLVED IN URBAN PLANNING AND RELATED FIELDS

1. Name of Organization:
   a) In original Language: _____
   _____
   b) In Official English Translation: _____
   _____
2. Affiliation (Parent body): _____
   _____
3. Permanent Mailing address:
   a) Region: _____
   b) Locality: _____
   c) Town: _____ d) Box: _____
   e) Country: _____ f) Phone: _____
   g) Telex: _____ h) Cable: _____
   i) Fax: _____ j) E-mail: _____
4. Former name (if applicable): _____
   _____
5. Principal Officers:
   a) Surname/forename: _____
   b) Designation: _____
6. Date of establishment (YYYY-MM-DD): _____
7. Nationality (Registered in country): _____
8. Location of main office (Place): _____
9. Type of organization (Tick the appropriate one(s)):
   [ ] NGO [ ] Governmental [ ] International
   [ ] Commercial [ ] Other (specify):
   _____
10. Specific field(s) of interest in its area of operation: (e.g. planning, research, consultancy, etc.)
    _____
    _____

22. Personnel (a. Type, b. No.; e.g. a. Researcher b.15):
    a. _____ b. _____
    a. _____ b. _____
    a. _____ b. _____
    a. _____ b. _____
    a. _____ b. _____
    a. _____ b. _____
23. Equipment/Facilities, etc.:

    _____
    _____
24. Financial support (Optional):

    _____
25. Membership in Associations:
    a) Name of Association: _____
    b) Type of membership: _____
    c) Year of membership (YYYY-MM-DD): _____
26. Honours and Awards:
    a) Name of Award: _____
    b) Granting institution: _____
    c) Year of Award (YYYY-MM-DD): _____
27. Description (Any other information, e.g. Historical)

    _____
    _____
    _____

Annex 10: A Form to Collect Information on Profile of Projects

INFORMATION ON PROJECTS IN URBAN PLANNING AND RELATED FIELD

(A Questionnaire on Project profile)

1. Project status:
   [ ] Proposal [ ] Ongoing [ ] Completed
2. Title of Project:
   a) In original language:

   _____

   _____

   b) English translation of title:

   _____

   _____
3. Project No. _____
4. Contract No. _____
5. Implementing organization:

   _____

   _____
6. Location of project (place, country):

   _____

   _____
7. Affiliation (Parent body(ies)):

   _____

   _____

   _____
8. Mailing address:
   a) Box: _____
   b) Locality: _____
   c) Town: _____ d) Province _____
   e) Country _____ f) Phone_____
   g) Telex _____ h) Cable _____
   i) Fax _____ j) E. mail _____

9. Former title of project (if applicable):

_____

_____

10. Project series:

a) Series name _____

b) Number/year _____

11. Working language: _____

12. Objective(s) of the project:

_____

_____

_____

_____

_____

13. Project description:

_____

_____

14. Principal officers (a. Surname/Forename,
b. designation/role)

_____

_____

_____

15. Project team members (a. Surname/forename,
b. Function)

_____

_____

_____

_____

_____

_____

16. Associated organizations:
(a. Name, b. Parent body, c. Country, d. Role)

_____

_____

_____

_____

_____

_____

17. Duration of project: _____

18. Date proposed/approved (a. YYYY-MM-DD, b. YYYY-MM-DD)

_____

29. Starting date (YYYY-MM-DD): _____

20. Expected completion date (YYYY-MM-DD): _____

21. Date of actual completion (YYYY-MM-DD): _____

22. Date of termination (YYYY-MM-DD): _____

23. Type of organization:

[ ] NGO [ ] Governmental

[ ] International [ ] Commercial

[ ] Other (Specify): _____

24. Specific field of project (e.g. Housing, development, etc.)

_____

25. Geographic area of coverage:

[ ] Local [ ] National

[ ] Regional [ ] International

Mention specific area if possible (e.g. Housing project, Kotobe, Southern Ethiopia):

_____

26. Local descriptors: _____

27. Broad subject areas: _____

38. Publications of the project: (Please attach list of publication) [ ] Lists [ ] No publication

29. Type of research undertaken by the project organization:

[ ] Fundamental [ ] Applied

[ ] Developmental [ ] Survey

30. Patent taken: (a. Patent No., b. Title, c. Year, d. Country registered in)

_____

_____

31. Related project(s) in urban planning fields:

_____

_____

_____

32. Equipment/Facilities owned by the project organization:

_____

_____

_____

_____

33. Financial resources (Optional):

_____

_____

34. Honour and Awards:

a) Name of award: _____

b) Granting Institution: _____

c) Year (YYYY-MM-DD): _____

35. Other information on the project:

_____ _____

_____

_____

_____

BIBLIOGRAPHY

1. Abebe Rorissa, et al. 1992. "ABNCD+: A Prototype for an Integrated Information Storage and Retrieval System". Microcomputers for Information Management. **9**(3): 137-159.

2. Aitchison, J. 1968. "Practical Application of Facet Classification with Special Reference to the English Electric Faceted Classification for Engineering".In Classification for Information Retrieva. London: Clive Bingley: 43-72.

3. Aichison, Jean and Gilchrist, Alan. 1972. Thesaurus Construction: A Practical Manual. ASLIB. London.

4. Alshuler, A. 1965. "The Goal of Comprehensive Planning and the New Comprehensiveness". Journal of the American Institute of Planners. **31**(3). (August): 178-189.

5. American National Dictionary for Information Processing. 1977. Washington D. C. Computer and Business Equipment Manufacturing Association (CBEMA), (Report No. X3/Tr-1-77).

6. Ayeni, Bola. 1979. Concept and Techniques in Urban Analysis. Croom, Helm.

7. Backwell, K. G. B. 1968. Classification for Information Retrieval.London: Clive Bingley.

8. Bhattacharyya, G. 1981. "Some Significant Result of Current Classification Research in India". International Forum on Information and Documentation. Moscow, January, **6**(1): 11-18.

9. Brodie, M. L., et al. 1984. On Conceptual Modelling Perspective from Artificial Intelligence: Database and Programming Language. NY: Springer Verlag.

10. Castelles, M. 1977. "Towards a Political Urban Sociology". In Harloe, M. (ed.) Captive Cities. London: Wiley: 61-78.

11. Chadwick, G. E. 1971. "The Alternative Future of Alternative Futures". In Planning Outlook. New Series **10** (Spring): n.p.

12. Codd, E. R. 1970. "A Relational Model of Data for Large Shared Databases". Communications of the ACM. **13**(6): 377-387.

13. Cripps, E. L. and Hall, P. 1969. "An Introduction to the Study of Information".In: Information and Urban Planning. Centre for Environmental Studies (CES) **8**.

14. Davindoff, P. 1965. "Advocacy and Pluralism in Planning in Planning". Journal of the American Institute of Planners. **31**(4): 331-338.

15. Davis, B. 1980. Database in Perspective. H. Charlesworth and Co. Ltd.

16. Debons, Anthony; Esther, Horne and Scott, Cronenwelt. 1988. Information Science: An Integrated View. Boston: G. K. Hall.

17. Dyckman, J. W. 1969. "Guest Editor's Introduction to the Practical Uses of Planning Theory". **35**(5) (September): 300-315.

18. Ethiopian Government. 1987. A Proclamation to Provide for the Establishment of the National Urban Planning Institute. Proclamation No. 317/1987. Negarit Gazeta. **46**(15): 135-138

19. Fenichel, C. H. 1980-81. "The Process of Searching On-line Bibliographic Databases: A Review of Research". Library Research. **2**(3): 107-27.

20. Forkett, D.J. 1964. London Educational Classification. London Institute of Education.

21. _____. 1982. The Subject Approach to Information. 4th ed. London: Clive Bingley.

22. Friedman, J. A. 1985. "Response to Alshuler: Comprehensive Planning as a Process". Journal of the American Institute of Planners. **31**(3) (August): 191-205.

23. Ganzmann, Jochen. 1990. "Criteria for the Evaluation of Thesaurus Software". International Classification. **19**(3/4): 148-157.

24. Gerrie, Brenda. 1983. On-line Information Systems: Use and Operations characteristics, Limitations and Design Alternatives. Arlington, Virginia: Information Resource Press.

25. Gilleson, Mark L. 1985. Database: A Step-by-Step. John wiley, 1985.

26. Guinchat, C. and Menou, M. 1983. General Introduction to the Techniques of Information and Documentation Work. Paris: Unesco.

27. _____. 1990. Sciences et Techniques de l'Information et de la Documentation: Introduction generale. Deuxieme edition. Paris: Unesco.

28. Hansen, W. B. 1968. "Metropolitan Planning and the New Comprehensiveness". Journal of the American Institute of Planners. **34**(5) (September): 295-311.

29. Harrington, J. L. 1989. Making Database Management Work. Dryden Press, Reinehart and Winston Inc.

30. Hartley, R. J., et al. 1990. On-line Searching: Principles and Practice. London: Bowker-Saur.

31. Heaps, H. S. 1978. Information Retrieval: Computational and Theoretical Aspects. London: Academic Press Inc. (London) Ltd.

32. Ingwerson, Peter and Wormell, Irene. 1992. "Ranganathan in the Perspective of Advanced Information Retrieval". Libri. **4**(3): 184-201.

33. Korth, Henery F. and Silberschatz, Abram. 1986. Databases Systems Concepts. New York: McGraw-Hill.

34. Lancaster, F. W. 1986. Vocabulary Control for Information Retrieval. Washington D. C.: Information Resource Press.

35. Macleod, Jan A. and Reuber, A. R. 1987. The Array Model: A Conceptual Approach to Document Retrieval. JASIS. **38**(3) (May): 162-170.

36. Maltby, A. 1968. "Faceted Classification". In: Classification for Information Retrieval. London: Clive Bingley: 31-42.

37. Massey, D. B. and Cordey, Hayes, M. 1973. "The Use of Model in Structure Planning". Town Planning Review. **42**: 28-44.

38. Meyriat, Jean. 1991. "Information for Socio-Economic Planning". A Paper submitted to the Inter-Regional Euro-Arab Technical Seminar on Information Systems and Networks. Granada, Spain: 25-29 december, 1991: 1-6 (Unpublished).

39. McLoughlin, J. B. 1965. Urban and Regional Planning: A Systems Approach. London: Faber and Faber.

40. _____. 1965. "The PAG Report Background and Prospect". Journal of Town Planning. **57**(7), (July/August): 257-261.

41. Molla Hunegnaw. 1993. Enhancement of User-Friendliness of Micro-CDS/ISIS: Search and Retrieval Phases. Master Thesis. AAU, SISA. Addis Ababa.

42. Neelameghan, A. 1992. "Application of Ranganathan's General Theory of Knowledge Classification in Designing a Specialized Databases". Libri. **42**(3): 202-206.

43. _____. 1992. "Home Grown Object-Oriented Databases". In: New Information Technology (NIT) 5th International Conference. November 30-December 2, 1992. Kowloon, Hongkong: 259-284.

44. _____. 1991. "Concept Categorization and Knowledge Organization in Specialized Databases: A Case Study". International Classification. **18**: 92-97.

45. _____. 1975. Presentation of Ideas in Technical Writing. Delhi: Vikas Publishing House.

46. _____. 1992. On-line Database Searching and Retrieval: Strategies, Procedures, Commands, and Problems: A Brief Guide in **3 Parts. Part 1:** On-line Database Services and Information Retrieval: An Overview. SISA, Addis Ababa University, Addis Ababa, Ethiopia.

47. OECD. 1985. Macrothesaurus for Information Processing in the Field of Economic and Social Development (3rd ed.). New York: United Nations.

48. Ralson, Anthony (ed.) and Reilly, Jr., Edwin D. (Ass. Ed.). 1983. "Data". In: Encyclopedia of Computer Science and Engineering. 2nd ed. Van Nostrand Reinold Co. Inc.: 443.

49. Ratcliffe, J. 1981. Introduction to Town and Country Planning (The Built Environment). 2nd ed. Anchor Press, Ltd.