Chapter I

INTRODUCTION

1.1 Background of the study

1.1.1 Libraries in the past

The creation of ancient sculpture and drawings were the first attempts in the history of human civilization to establish the foundation of a purposeful and orderly assimilation of human knowledge." From 3000 to 4000 B.C. the Egyptian pictographic writing were found in building stone, known as *hieroglyphics.* During Sumerian-Babylonian-Assyrian civilizations, soft clay, woods etc and other tools was used to sketch out the major events concerning their contemporary mode of existence to their coming generation" (Prasher,1991). Libraries are in existence since long. But emergence of libraries as social institution is a fairly recent phenomenon. Library management in the earlier days was a matter of rule of thumb, and library systems were developed on trial and error basis. Library science emerged as a discipline only towards the end of 19th century.

Till the beginning of the present century, book remained as the chief source of information. The knowledge about different discipline was less. Scientific research was at slow acceleration and scientists were very few to accelerate it. Communications of scientific ideas were supposed for certain premises only. Books were able to carry the entire realm of knowledge of discipline and libraries were only one kind of document preserving house, i.e. book store house.

Gradually its concept was changed with advancement of human knowledge along their witnessed three major revolutions. These are:

a) Agriculture revolution;

b) Industrial revolution; and

c) Technological revolution.

These three revolution brought tremendous changes in mans physical, social, economic, and political environment and behaviors.

Today we are at the threshold of yet another revolution what is described as information revolution. Since the invention of printing, there has been a continuous revolution in the generation, transfer and communication of information. However the role of information has attained new proportions with the acceleration of research; mounting social and population pressure; changing technological environment and increasing needs of planners, decision makers, executives, lawyers, doctors and even the common man. This process of preserving information from stone, micro chips to biometric technology and its wide application for various purpose strain the effective library; i.e. a library where information could be easily encoding, storage, and retrieval to fulfill the desired query. With this varied change and increased need of information a librarian should be able to cater prompt and precise information. For this purpose a library must be managed properly with effectively retrieving technique to meet its objectives. However this is not easy as explained, this needs to be suffered from various phases of processing. Therefore the IRS is the later one step that occupied a great place in library service with direct interaction of end user.

1.1.2 Information retrieval system

The term information retrieval was coined in 1952 and gained popularity in the research community from 1961 onwards. In developing phase its function was seen as management of catalogue and index for use of information or documents. With the technological revolution or specially IT revolution it's important is increased in multiple discipline. From the terminological point of view the concept of IRS is to some extent self explanatory. One may simply denote such a system as one that stores and retrieves information. As a system it is therefore composed of a set of interrogative techniques each of which designed to serve a specific function for a specific purpose.

As IRS occupies a huge era in the knowledge society, an in IRS is designed to retrieve the documents or information required by the user's community. Ideally there would not have been any need for establishing such a system if the number of documents in the universe would not have exceeded a few hundreds or so, and if each author in each subject field would have known the reader of his or her documents. But none of these is true and quite a few reasons are there which have made the knowledge society bound to establish the IRS to facilitate the explosion of knowledge. The industrial

revolution, technological advancement, exploratory and information technology development demands some standard management of knowledge and retrieve of such information or knowledge in precise and prompt way. Findings of research and new invention, conference proceedings, various disciplined abstracts, author's creations etc are the sources of information and IRS aim's collecting and organizing all these documents in order to provide them to their users as soon as required.

After understanding what is IRS and its rationale, there may arises questions in our mind that with which it associates and how it could works? So it is associated with computerized database and online search engine. As information retrieval is the interdisciplinary subject based on computer science, mathematics, library science, information science, cognitive science, psychology, and physics. The information obtained from various sources need to be organized or manage with the help of computerized, mathematical, physical and cognitive science, and psychological techniques. For these there must be a database or search engine in which all the information is organized in such a way that it could be retrieved whenever needed by users. Of course, it is difficult to retrieve the desired information as and when needed from the numerous of information. So these collections should be processed in promptly retrieval manner. The searching and retrieval technique should be managed in such a way that it could be best matched with the navigators needs. Indexing, cataloging, and classification with its effective retrieval model can make this situation user friendly.

With the technological advancement many new subjects have been developed with different mode of subject formation. This development symmetrically increases the complexities in the use of information. Most of the documents comprise more than one subjects; this creates a kind of difficulties in selection and retrieving of documents or subjects. Such a situation may cause loss of time and some times navigators couldn't find his or her required information. Inundated documents may be there in the same subject and same topic where s/he will be lost. So efficient and precise retrieving technique are necessary to support the effective retrieval of document which the navigators browsing. Thus it is a research on Boolean operator an approach in information retrieval system. The concept of information retrieval presupposes that there are some items of information that have to be organized in

some suitable order for easy retrieval. Therefore IR is the science of retrieving the information from books, journals and other documents as well as documents themselves.

"In responding to users queries IRS must achieve a balance between speed, accuracy, cost and retrieval effectiveness in revealing the existence of information items and displaying surrogates or the original items. The effectiveness of retrieval is measured by the pair of measures recall ratio and precision ratio. Rules on bibliographic description usually conform to ISO 2709, with which the range of MARC standards for library catalogue comply".

The subject of an item may be represented by one or more of the following :(Hartley,1996)

1. An abstract;
2. Terms chosen from a thesaurus;
3. A particular database;
4. Themes chosen from natural language;
5. Codes taken from classification scheme;
6. Terms taken from a list of subject headings.

In these approaches it is necessary to understand the subject of the documents and translate those subjects into the appropriate index language for effective retrieval system. Appropriate keywords and the specific subject heading only can't meet the desired retrieve of the particular document. Beyond it one on the most recent uses of Boolean logic also known as Boolean operators is a method for describing a set of object or ideas. It was invented in 19th century by George Boolean, an English math teacher. But it has become part of the foundation for controlling computers, retrieving precise information in search engines and databases. The binary 0 and 1states are naturally related to the true and false logic variables. By inserting operators (OR, AND and NOT) between indexed terms and keywords in a search statement we can describe the relationship among the terms for the precise findings.

1.1.3 Boolean logic

"One of the most recent uses of Boolean logic is in search engines and databases (full text or partial). By using the operators we can effectively communicate exactly what we are looking for to search program. Without a standard language such as this, everyone may have a different way of expressing the same search request, without this precision, searcher less than effective, and retrieving valuable information would be even more difficult. The better we know Boolean logic the more effective our searches will be and the quicker we will be able to find what we are looking for." To access and retrieve the desired information, the users seek the document through the bibliographical databases where, they need the subject heading and keywords about the document.

The user would always like an exact and perfect answer to the query posed to the system. This however is only possible when the query is exactly posed and the data in the database system can be definitely identified as responding to the query or not. The most obvious exact match situation arises with best management of all approachable term, effective coordination of these terms by using Boolean operators. In Nepal no study was done about retrieval techniques, its strength, limitations and user friendliness study of retrieval part. However for indexing there are few studies which hade been deal with the information retrieval system via assigning indexed terms. So this study strive to highlight the need of knowledge and effective use of retrieval techniques to retrieve the exact information what the user always demands with the system. In many times there arises the situation of losing the exact information in spite of effective indexing, due to lack of appropriate retrieving technique. To overcome and finding out such difficulties this study would be good guidelines for the seekers of information.

1.1.4 Database (Goyal, 2000)

Data:

A general term for numerically encoded information, particularly used for information stored in a database. The word however frequently used in a casual way with a sense not especially different from information as for instance, in a phrase like biographical data.

Database:

A database is essentially, any systematically organized collection of information, records, files in whatever form are related to each other. However more usually, information stored on computer files, or on CD-ROM. A database might contain bibliographical data or numerical or statistical materials. Generally in data processing one is interested in collecting similar entities, such as for a book and would be interested in recording information about the associated attributes of each of fields of that particular book. Each attributes is termed as data item or data elements or some times field. Data is generally structured so that it can be sought and retrieved automatically.

Actually a field is a physical space on the storage device, where as a data item is the data stored in the field. For e.g. bibliographical information of certain book may be regarded as an entity. The various attributes of this entity may be: Author name; Title; Publishers; Date of publications; etc. for this we can express on data hierarchy as follows.

Data Item	Book
Record	Bibliography
Field	Author Title Place of publication Name of publisher Year of publication
Data base	Khera, Shiva Kumar You can win Delhi Macmillan India 2007

Fig: 5 Database hierarchies

Thus a database is an organized collection of records and files which are related to each other. In other words a database is a collection of either citations or full-text articles. Each article or citation is called a record. Each record in a database contains the same elements, referred to as fields. Commonly occurring fields include title, author, publisher, data, journal title, keywords, and abstracts, as well the full text of the article (Kent, 1999)

Online database:

Online database offer citation of books; articles in journal; magazines and newspaper; and reports. Ones we have searched the database and found the right information, we can printout or download the bibliographic citations, abstracts and sometimes even full text. Different libraries offer different databases, for e.g. LEXIS focuses on law and NEXIS focuses on business. Some libraries may also offer online services that enable us to access a large number of databases. For e.g. DIALOG contain ERIC (education), PhychINFO (psychology), the MLA International bibliography (literature), News Bank (periodicals) and the government printing office monthly catalog.

1.1.5 Search engine

In today's high-tech world, findings and using search engines is extremely important. Search engines play an important role in providing access to and retrieving key information from Web sites on virtually every subject. A few examples are Google, Alta Vista, Hotbot, and Yahoo. These software-driven spiders or robots as they are called rove through millions of indexed URL's or Web pages-matching sites to the exact word or phrase of your search with amazing speed and accuracy. In short search engines are a research's dream, minimizing both time and effort for what otherwise would be a laborious and painstaking process (Lenburg, 2007).

Thus a search engine is computer software that searches a collection of electronic materials to retrieve citation, documents, or information that matches or answers a user's query. The retrieved materials may be text documents, facts (numerical value) that have been extracted from text, images, or sounds. A query is a question phrased so that it can be interpreted properly by a search engine, Depending on the type of software it may be a collection of commands, a statement in either full or partial sentences, one or more keywords, or in the case of non text searching, an image or sequence of sounds to be matched as per query (Kent, 1999).

Search research on the web has a short and concise history. The World Wide Web Worm (WWWW) was one of the first web search engines. It was subsequently followed by several other academic search engines; there has been a fair amount of work on specific features of search engines. Especially well represented is work which can get results by post-processing the results of existing commercial search

engines, or produce small scale "individualized" search engines. Finally, there has been a lot of research on informational retrieval systems, especially on well controlled collections.

Text:

Search engines are most commonly associated with searching text and data. This is not surprising since collection of electronic text predate image or sound collections in digital form. These collections of text are commonly referred to as databases.

The search engines with which this article is concerned are the foundation of text information retrieval system. They are designed to manipulate large amounts of text. New techniques in searching non text materials, however, are becoming increasingly feasible for widespread use in large collections.

Images:

Image collections, whether of moving or still images, are most commonly searched through textual description that has been attached to the images. Such descriptions are often divided into searchable fields such as name of artist, type of medium, subject of work, and date.

Sound and music:

This area is largely experimental. Matching musical passages has logged behind both text and image searching; through there are now some systems which match typed in musical themes.

When researching and working with libraries or search engines, one very important fact to understand is that organization and retrieval of information with interrogative techniques. Large unorganized collections of information are of minimal use to any one until they have been stored in to a discernible pattern. For this reason, methods for creating access to printed materials were developed. These methods are commonly referred to as indexing, cataloging, and classification. Their purpose is to help users find materials within a collection.

1.1.6 Library catalogues

Libraries create order by sorting information by subject and/or by author. The document to be catalogued can be shelved in only one location. Works that are on more than one topic must be placed in a single spot. To solve these problem libraries created card catalogues. The card catalogue enabled users to fine the same book under

multiple entry points, such as author, subject, title, or series name. The card catalogue or online catalogue acts as a surrogate to the collection. Since library card are easier to duplicate than books, libraries often create multiple access points for each subject within a book as well, thus increasing access to its contents. Standard libraries access points allow the user to find a work by title, author, or subject. This is the brief description of the information retrieval system via catalogue used in early periods.

1.1.7 Indexing

Indexing is the process of analyzing the contents of a documents or collection of documents and translating the results of the analysis into terms for use in an index to allow location and retrieval of information (Bakewell, 1996).

Most of the documents are of composite subjects. So it was required to provide subject approach of information. Relative Index of DDC (1876) brought all the scattered terms together under an approach term (Dewey, 1876). Ranganathan has provided different abbreviations for this purpose such as ‘defined’ by ‘def’, ‘in relation to’ by ‘irt’, ‘referred in relation to’ by ‘rirt’ and so on. (Ranganathan, 1974). Continuous refinement in indexing system has brought important improvement in the field of indexing techniques.

"Cutter was the first to discuss the concept of direct entry in his Rules for Dictionary Catalog in 1876. He advocated the entry should be under its subject heading not under the heading of the class which includes that subject. He also suggested that subject having two or more themes should be provided accordingly composite subject with place, firm, name. This brought in some uncertainty in fixing the order of various components in the subject heading.

Kaiser tried to reduce this uncertainty by fixing the order of significance of the components as ‘concrete’ and ‘process’ (‘systematic indexing’-1911).

Coates brought in further improvement. The ‘concrete’ and ‘proces’ of Kaiser were renamed as ‘thing’ and ‘action’ by Coates. He developed his ideas further and introduced such categories, as ‘part’ and ‘material’. The order was ‘thing’, ‘part’, ‘material’ and ‘action’.

These were ad-hoc solutions without any sound theoretical base. It was Ranganathan who advocated that the order of component should be based on the clear understanding of the concept of specific subject and the vision to formulate it on

scientific basis. For this purpose chain indexing and the use of fundamental categories PMEST was developed.

After chain indexing, Derek Austin developed PRECIS (Preserved Context Indexing System) in 1968. Then G. Bhattacharya propounded POPSI (Postulate-based Permuted Subject Indexing). After that Post Coordinate Indexing System, Keyword Indexing System, Citation Indexing Systems were developed (Prasher, 1991).

1.1.8 Pre-coordinate indexing

Indexing that permanently establishes relationship between tow or more than two subjects/keywords in a single heading is known as pre coordinate indexing. It allows the user to find precise subtopics within a broader category, because in it coordination of terms is occurred at the time of input stage. Chain indexing, PRECIS and POPSI are pre coordinate indexing systems, because in this the coordination of indexed term is done at input stage in anticipation of users approach.

Library subject heading are standardized in controlled vocabularies so that it is not flexible for retrieving information as independently as required by the users. In thesaurus such as the Library of Congress Subject Heading or the Sears List of Subject Heading users should go accordingly its controlled vocabulary. Both these traditional thesauri are examples of pre coordinate indexing.

1.1.9 Post coordinate indexing

Outside the traditional library sphere but simultaneously, we find experimentation with a different information retrieval approach. Post coordinate system were developed in the 1940s and 50s to answer the need for quick access to current and precise topics. Post coordinate indexing assigns single terms to documents, they are not pre coordinated as they are in a library thesaurus or controlled vocabulary. The purpose of post coordinate indexing is to permit any combination of two or more terms with some techniques as Boolean operators. Cross discipline searching is facilitated, and the searcher need not know the established terminology or format in order to all or any terms that were assigned to a document. Post coordinate indexing requires some mean for combining terms.

"The indexing system which is coordinated at the stage of searching or output stage is known as post coordinate indexing system. It means it is done at the output stage by

searcher them selves. Post coordinate indexing system is started to overcome the limitation of the pre coordinate indexing systems (Prasher, 1991).

The information seeker has unrestricted freedom for the free manipulation of the subject at the time of searching in order to achieve whatever logical operations are required. It is designated as manipulative because it permits a greater degree of search manipulation and the index term can be coordinated almost in any combination.

1.1.10 Subject heading

A subject heading is the word or phrase used in the library catalogs to express the main theme or topic of a book. A cataloguer assigns subject heading for a book by selecting from a printed subject heading list or from a locally developed subject authority list. A subject card is prepared by putting the assigned subject heading above the heading of the main card. Series of such subject cards is called subject catalogue. The main purpose of subject catalogue is to list under one uniform word or phrase all of the materials on a given subject heat a library holds.

Classification helps to arrange like document together and to separate unlike where as cataloguing provides subject heading and other elements for the documents. But most of the documents comprise more than one subject. Also in al library books are shelved in classified order according to their call no. but for compound subject there is no provision to assign it in such order. Therefore subject heading is very useful for compound or complex type documents.

Two systems are used for assigning subject heading. One is derived and another is assigned. In the derived typed system all subject terms are taken from the documents itself. In assigned type system the indexer or cataloguer creates the descriptors. It is an intellectual method involving the finding out of specific subject of the document and assigning appropriate subject headings. "All indexing languages with vocabulary control devices such as subject hading list thesauri and classification scheme are assigned term systems. These systems are intellectual and therefore require more time and money at the input stage".

Cataloguer must follow or use certain standard printed subject heading list to maintain uniformity and consistency in subject headings and subject catalogue. There are some

popular subjects heading lists such as Sears List of Subject Heading, and Library of Congress Subject Heading, etc.

"Sears's list of subject heading (SLSH) was first prepared by Minnie Earl Sears in response demand for a list of subject headings suitable to the needs of the small libraries". First edition was published was published in 1923 in the title 'List of Subject Headings for small Libraries', which was based on the subject headings used by nine small libraries. However Minnie Sears early recognized the uniformity, and she followed from the LCSH with few expectations (Kumar, 1994). From its 4th edition, it started to give Dewey Decimal Classification numbers to its subject headings. Now the 19th edition of this list is used.

Library of Congress Subject Heading (LCSH) is the most popular subject heading in the world. "The first edition of the library of congress list called subject headings used in the Dictionary Catalogues of the library of congress was printed in parts between 1909 and 1914. Supplementary list were issued as required, followed by a second edition in 1919. Later editions were published at irregular intervals. The title changed to Library Congress Subject Heading, when the eighth edition was published in 1975.Now the 30th edition (2007) is published containing over 280,000 total headings and references.

1.2 Statement of the problem

Keyword is the most approachable tool to search and browse the information. Without sufficient keywords to represent the document and to support the subject headings, it is impossible to serve its users properly to find the information. So librarians are bound to use uniform, consistent and user friendly keywords and subject heading lists based upon some principles as far as possible. Except sufficient keywords another most important part of the retrieval system is idea about retrieving technique: such as techniques implementation of Boolean operators in between various keyword terms. Specifically the study will focus the idea about Boolean operator among the professional librarians and general users or researchers in retrieving information, and their efficiencies in retrieving information via the operator and the problems faced by them and tackling of problems.

As a service institution, it serves its users. User may face many problems while navigating the information through the data bases and search engines. A search engine is much like the online catalog or database at our local library or university/college library. On the home page of the search engine user have chosen, s/he will find a box where s/he can enter a keyword, when they hit the return key, the search engine retrieves and displays all the web pages and other information in its databases that match with their query. If not so they lose their time in searching the document what they need. Subject approach is the most important in navigating the information but indexed term of related document doesn't cover all the information that may found in that document. Thus checklist of effective searching could perform better than the above one in retrieving environment (Kirszner, 2002).

Checklist: for effective searching comprises;

1. Choosing the right search site/database;
2. Choosing keywords/ indexed term carefully;
3. Narrowing the search via OR, AND,and NOT operators;
4. Checking spelling and permutation of key terms;
5. Consulting the help screen in case of search engine or consulting the manual in case of manual database;
6. Using more than one subject guide or search engine for conformation of information.

If these prerequisite couldn't be followed properly important information will not be retrieved. So the actual problem of libraries and their users is not about collection but surfing the no of information over the existing collection.

Thus problem statement focuses the following main aspects of the retrieval system. Evaluating the abilities or skills of users in exact query formulation via using Boolean operators for database searching, idea about the checklist for effective searching, essence of any orientation for navigation information via using Boolean operators, and efficiency of retrieval technique (Boolean Retrieval Model) in preventing misguiding or false coordination on some kinds of topics.

With the explosion of information the complexities of retrieving information has also became challenging task to every users, with the increasing rate of database and web sites, their different pattern of arranging information creates problem in information retrieval. User always faces problems with the bound keywords regarding the subject

headings and thesaurus. Due to these complexities many of the users can't coordinate keywords appropriately for their desired information. Exact formulation of search statements with Boolean operator is also complex due to its limited approach up to its users.

Reliable and valid results of any problem can't be obtained without any systematic investigation or research study. Therefore comparative study of professionals and general users with any database helps in finding the exact and real problem. Most of the users do not use Boolean operators, the problem is that either it is too complex to use or users have not interest to obtain and increase the orientation about Boolean logical statements. Due to increasing databases, their complexity and inability of the users to use exact keywords, users are unable to formulate successful query formulation, so the users are not so friendly in using Boolean operators.

As TUCL is the largest online library database in Nepal, it has various source of information. Among them online data search is done by many users including professionals. But the user's loss, their time and effort that returned inundated information for their query, because of their insufficient knowledge about Boolean operators use, utility and its effectiveness.

1.3 Objective of the study

(i) To find out the existing proportion of use of retrieval techniques used by the general users or library professionals while navigating information;

(ii) To examine the clear understanding, use and user friendliness of Boolean operators in information retrieving;

(iii) To find out the percentage of navigators who are uses Boolean operators in navigating library databases and other search engines;

(iv) To obtain the no of user who are identified with checklist and their followers;

(v) To make comparative study of retrieve information by the library professionals and general users from the TUCL-online library database and various search engines with the application of Boolean operators;

(vi) To find out whether users are interested or not to obtain orientation of Boolean retrieval technique;

(vii) To suggest/recommend the concerned library database users and search engine users(navigators) to the effective use of Boolean operators and efficiently coordinate their search terms until they didn't get their requirements;

(viii) To measure the efficiency of Boolean operators in information retrieving on basis of above study;

(ix) To find out the problems faced by the users while navigating required information.

1.4 Hypothesis

1. Effectively coordinated search terms with Boolean operators strongly support in information retrieval;
2. Max no of user use Boolean operators while navigating databases/search engines for the precise and prompt information;
3. Efficiency of Boolean operator is high for coordinating keywords and retrieving information.

1.5 Scope and limitation of the study

This study was limited within the four libraries inside the Katmandu valley which have their library database with Boolean search facility and some navigating centers or research centers with their search engine users, whom queried their information with coordinating the terms by Boolean Operators.

Such are:

Tribhuvan University Central Library;

Social Science Baha Library;

Katmandu University School of Management Library; and

ICIMOD Library

It aims to highlights the need of information retrieving technique or in other words need of information retrieval model for effective results. The work is limited to the coordination of index terms with Boolean operator and measure of its efficiency to support the information retrieval. This study helps to examine the existing situation of use of searching techniques among the library databases and search engines.

This study was limited to the subject of library and Information science, however it has broad foundation with the various kinds of users from each disciplines. Mainly two kinds of users (library database users and cyber navigators) were consulted as per requirement of study to attempt the objective of the research/thesis. This study will be completed within three months approval of proposal.

1.6 Significance of the study

Advancement of technology and competence age make explosion of information. Such flood of information in various disciplines has made disorder in their proper management. Just collect the information from market is not the task of library but it is necessary to make them easy available with advanced management, in order to make them maximum utilization. Advance management and max utilization focuses the right information to right user at right time in right form. To cope with quark demand of user we must have need of effective retrieving technique. General user often faces difficulties in approaching exact information especially due to inability of formulating an appropriate search statement. This could be minimized or cost of search can be reduced significantly if an appropriate strategy is followed properly. Search strategy provides the facility to select the optimum path for searching a file or a database or information from the system. Thus three symbolic operators ***, +, -** (AND, OR, NOT), combines the search statements in symbolic command and provides the high degree of retrieving information. Also these operators have their individual character to access the information contained document. Thus this study signifies that efficiency of Boolean operator is high for information retrieval.

1.7 Definition of the terms

Bibliographic database:
It refers to data entered systematically in a defined structure. In a given framework of software, bibliographic elements of bibliographic items, defined by ISBD like title and statement of responsibility, edition, material designation, place and publisher, pagination, series, note, ISBN/ISSN are fed in computer. The programming of such software make possible to retrieve and disseminate the information systematically when required. It can be said as metadata, the data about data. Meta data are structured data provide a short summary about any information resources (Pradhan, 2004).

Catalogue:

Catalogue refers a list of books, maps or other items, arranged in some definite order. It records, describes and indexes (usually completely) the resources of a collection, a library or a group of libraries.

Cataloguing:

Is the process of compiling a catalogue or constructing entries for insertion into a catalogue. In a broad sense, it refers to all the processes connected with the preparation and maintaining of a catalogue, including classification and assignment of subject headings.

Classified catalogue:

A catalogue of subject entries which are arranged in systematic order according to scheme of classification.

Controlled Vocabulary:

A listing of words or terms which must be used as subject heading of descriptors in a particular database.

Data:

A general term for numerically encoded information, particularly used for information stored in a database. The word however frequently used in a casual way with a sense not especially different from information as for instance, like biographical data.

Database:

A database is essentially, any systematically organized collection of information, records, files in whatever form are related to each other.

Index:

A detailed alphabetical list or table of topics, names of persons, places, etc., treated or mentioned in a book or series of books, pointing out their exact positions in the volume, usually by page number (sometimes with an additional symbol indicating a portion of a page) but often by section, or entry, number.

Indexing:

In information retrieval is that which specifies, indicates or designates the information, contents or topics of a document or a group of documents. Also a list of the names or subjects referring to a document or group of documents (IBM).

Index language:

The language that is used in the subject index which is part of an information retrieval system. It may be an alphabetical or classified arrangement of terms, or a variation of these. Each term or heading actual used in the index language, of whatever kind, is called an index term. Also is called ‘Descriptor Language’. Its ‘vocabulary’ is the complete collection of sought terms in the natural language.

Information retrieval:

Finding documents or the information contained in documents, in a library or other collection, selectively recalling recorded information. Methods of retrieval vary from a simple index or catalogue to the documents, to some kind of punched card or microfilm record which required large or expensive equipment for mechanically selecting the material required. Classification, indexing and machine searching are all systems of information retrieval.

Keyword:

In information retrieval systems, it refers to the significant word in a phrase; used for significant word in a title which is describing a document.

Library:

The term used for a collection of books and other library materials which have been kept for reading, study and consultation.

Library service:

Refers to the facilities which are provided by a library for the use of the books and the dissemination of information.

Relative index:

Refers to an alphabetical index to a classification scheme in which all relationship and aspects of the subject have been brought together under each entry.

Search engine:

Search engine is computer software that searches a collection of electronic materials to retrieve citation, documents, or information that matches or answers a user's query.

Subject:

Refers to the theme or themes of book, whether stated in the title or not.

Subject (indexing):

Refers to a unit of concept which is found in or derived from manuscript or published library materials.

Subject cataloguing:

Refers to that part of cataloguing which involves the allocation of subject headings to entries for specific book or other documents.

Subject headings:

Refers to the word or group of words under which books and other materials on a subject have been entered in a catalogue in which the entries have been arranged in alphabetical order.

Subject headings language:

The terms used as subject headings and under which entries have been made, as well as these form, which references have been merely made to other subject terms.

Thesaurus:

A book that is like a dictionary, but in which words are arranged in groups that have similar meanings (Hornby, 2008).

1.8 Organization of the study

The research study has been set up according to the standard format from the department. The first chapter deals with introduction which includes background, statement of the problem, objectives, scopes and limitations, significant of the study, definition of the terms and this heading itself falls. The second chapter deals with relevant studies of the literature i.e. literature review.

The third chapter deals with the subject heading and keywords, subject indexing, subject headings list, their development, keywords and their implementing with different database, Boolean operators and its application the database of different libraries for information retrieval.

The fourth chapter deals with research methodology, research design, population, sampling procedure, data collection procedure, and data analysis procedure. The fifth chapter deals with analysis and presentation of study which evaluates either the set objectives and testing of significance of hypothesis whether positively met or not. The final chapter deals with summaries and recommendations.

REFERENCES

Bakewell, K.G.B. (1996). Indexing. *International Encyclopedia of Information and Library Science* (pp.169-170). London: Rutledge.

Goyal, D.P. (2000*). Management information system: managerial Perspectives* Delhi: Macmillan.

Hartley, R.J. (1996). Information Retrieval. In *International Encyclopedia of Information and Library Science.* (pp.211-212). London: Rutledge.

Hornby, A.S. (Ed.). (2008). *Oxford Advanced Learner's Dictionary*, (7th ed.). Oxford University.

Kent, Allen (Eds.). (2004). Encyclopedia of library and information science (Vol. 64, pp. 218-246). NewYork: Marcel dekker.

Kirszner, L. G. and Mandell, S. R. (2002). *The Holt Handbook* (6th ed.). Fort Worth: Harcourt CP.

Kumar, Giraja (1994). *Theory of Cataloging.* New Delhi: Vikas.

Lenburg, Jeff (2007). *Guide to research.* New Delhi: Viva books.

Parmer, P.P., and Bhuta, B. (Eds.). (1989). *Encyclopaedic dictionary of library and information science* (vol. 1, pp. 1-62). New Delhi: Anmol.

Pradhan, M. R. (2004). Developing digital libraries: technologies and challenges. *Library Herald, 42*(2), 105-108.

Prasher, R.G. (1991). *Information and its Communication.* New Delhi: Medallion.

Ranganathan, S.R. (1974). *Cataloguing Practice* (2nd ed.). (pp.511). Bombay: Asia.

Chapter II

REVIEW OF LITERATURE

Research is the non ending and fact finding process. It starts from the conclusion or findings of the previous study i.e. what is already done, and study of these records helps to obtain detailed knowledge, conclusions or findings and then guide to the present study. In other words what happened in the past is always referable for checking and evaluating the rationality of present study and research. Hence those literature related with this study is mentioned here with what was done, to start further study.

There have been found many studies done on information retrieval system and Boolean operators for retrieving purpose. Information retrieval research has not a long history. The origins of information retrieval research can be traced back to 1953 when separate test were carried in Britain and the United States evaluating the performance of then controversial uniterm system devised by Mortimer Taube, which represented documents by single terms taken from titles or abstracts, against more conventional approach to the subject indexing and retrieval. These two test were the Armed Service Technical Information Agency (ASTIA) – Uniterm test carried out in the United State (Gull, 1956).

And the Cranfield Uniterm test undertaken at the college of Aeronautics, Cranfield in the United Kingdom and described by Throne.

The majority of research carried out in the field of information retrieval system following the Cranfield test did not question the cognitive and behavioral assumptions of the information retrieval model, and for a considerable time most researchers accepted the limiting assumptions of that model rather than questioning their validity. However from the late 1970s and early 1980s this situation began to change and a more eclectic attitude to information retrieval research emerged (Belkin, 1987).

The cataloguer selects appropriate subject headings and keywords for the bibliographic item and a unique classification number, called call number which is used not only for identification but also for the purpose of shelving, placing items with similar subjects near one another, which aids in browsing by library users, who

are thus often able to take advantage of serendipity in their search process (Nayaichyai, 2006).

LCSH is the most popular and widely used subject heading list for large libraries all over the world. The 27th edition of LCSH has contained subject headings and keywords created by cataloguer and are used in the cataloguing at the Library of Congress since 1898.It has given principles for additional subject and keywords and are commonly used sub division (2004).

SLSH is another popular subject heading list made for small and medium size library.SLSH has also followed some principles. It has given rules for subdivision and for coordinating the key words. It is not so big volume but it has contained more subject entries and keywords with controlled vocabulary (Miller and Goodsell, 2004).

S.R. Ranganathan in his Classified Catalogue Code (CCC) has given the rules for chain indexing code. This code is based on normative principles. To derive the subject entries, only the sought headings are accepted for subject headings and keywords.

The basic tasks involved in indexing are to analyze the content of the given document and the representation of this analysis by some content identifiers or keywords (Chowdhaury, 1990).

In subject indexing, however, the basic objective is to match the contents of documents with the users' queries, and thus the product of the conceptual analysis of the subject is represented in natural language form. A number of systems, viz. chain, PRECIS, POPSI, relational indexing, etc., have been developed over the ages for preparing subject index entries of documents. One basic problem involved in the process of subject indexing relates to the choice of appropriate keywords or descriptors through which the index entry is to be represented. The indexer prefers to use such keywords which not only represent the subject clearly, but also are likely to be used by the user while looking for the same subject. In order to standardize the task of choosing appropriate keywords for generation of index entries, a number of vocabulary control devices have been developed. Such devices include thesauri, classaurus, thesaurofacet, etc (Lancaster, 1979).

Historically POPSI is an extension of chain indexing. But they differ on fundamental points. In designing the system Bhattacharyya has successfully dissolved the difficulties of disappearing chain, as complicated the chain indexing. Simultaneously

he has succeeded in setting his system free from the clutches of any classification scheme. It should be emphasized that whilst POPSI is basically designed for manual methods, it can nevertheless be amenable to computers. It may be mentioned in passing that an improved version of software package is in the offing. POPSI does equally well to create an artificial memory and to serve the modern need. But we need to know more about the reactions of the users to the system- so often they are ignored while indexers happily experiment (Chakrabortty and Chakrabortty, 1984).

PRECIS first appearance in the British National Bibliography (BNB) in 1974 and it has touched many responsive chords. It is unquestionably an evidence of vitality of a system that during a decade it has accepted by nearly thirty agencies in British, Australia, Malaysia, and South Africa. Since 1978 Indian Library Science Abstracts (ILSA) published by IASLIC has been using PRECIS for preparing its index. However, PRECIS does not claim to be perfect but it does appear to be one of the best. It has already been tested in Swift's Sociology of Education project, but there are conflicting opinions about the degree of success. Austin claims the soundness of the underlying methodology. On the other hand, Swift and his team affirm that it does not cope with their needs. Bakewell also conducted a small study to evaluate the effectiveness of PRECIS. His conclusion about the relevance of this system to book indexing is worth quoting: "The indexes are not perfect –I detected a number of omissions-but they are very good and do illustrate the feasibility of PRECIS as a system for producing book indexes-provided enough space has been allotted to the task." Such claims and criticisms are liable to deflect the indexers interest away from the many useful suggestions it has to offer. Even in doubt and desire it has already made a mark through MARC projects and is making a march overseas, notably in Canada and Australia (Chakrabortty and Chakrabortty, 1984).

Indexing systems have been designed to assist in the retrieval of documents. It is operated by assigning index terms and keywords to the analyzed subject of each document either manually or automatically. Subject indexing systems have been classified broadly as pre-coordinate and post-coordinate systems. Any indexing system is to represent the contents of documents through keywords or descriptors (Lancaster, 1986).

An exhaustive indexing system is supposed to represent the contents of the input documents fully. However, to attain this objective, the system has to select as many

keywords as possible to represent the idea put forward in the document. In a non exhaustive system, only a few keywords are chosen which represent the subject grossly. Term specificity refers to how broad or how specific are the terms or keywords chosen under a given situation. The more specific are the terms and keywords; the better is the representation of the subject through the index entry.

A good indexing system is to isolate all the documents in a collection from the others in the same collection which do not discuss the desired topic. In other words, one has to choose such words for indexing which can differentiate a given document or a group of documents from all the others in the same collection. Sometimes this is denoted by the term discrimination. In this connection (Harter, 1988) has mentioned:

1. The keywords selected for representing a document should name the subject that is treated in the document.

2. Keywords selected for the index record of a document should name the subjects that are most heavily treated in the document.

3. The keywords selected for the documents should maximize the probability of retrieving the document.

(Lancaster, 1986) has mentioned, "process of subject indexing involves two quite distinct intellectual steps: the ‘conceptual analysis’ of the documents and ‘translation’ of the conceptual analysis into a particular vocabulary. The second step in any information retrieval environment involves a ‘controlled vocabulary’ that is a limited set of terms that must be used to represent the subject matter of documents".

(Aryal, 2005) has mentioned, Keywords are the most essential things to browse the information by the subject approach.

Keyword searching refers to a search type in which you enter terms representing the concepts you wish to retrieve. Boolean operators are not used.

Implied Boolean logic refers to a search in which symbols are used to represent Boolean logical operators. In this type of search on the Internet, the *absence* of a symbol is also significant, as the space between keywords defaults to either OR logic or AND logic. Nowadays, most search engines default to AND.

Implied Boolean logic has become so common in Web searching that it may be considered a de facto standard.

Work on informational retrieval systems goes back many years and is well developed (Witten, 1994). However, most of the research on information retrieval systems is on small well controlled homogeneous collections such as collections of scientific papers or news stories on a related topic; research centers collections; journals abstracts; small library database. Indeed, the primary benchmark for information retrieval, the Text Retrieval Conference (TREC, 1996), uses fairly small, well controlled collections for their benchmark.

A text retrieval system should provide for query formulation by using the Boolean AND, OR, and NOT operators, and also provide nested Boolean searching. Boolean search facilities allows a user to combine search terms (keywords/key terms) in a given search prescription with certain conditions imposed. These conditions specify whether more than one search terms should simultaneously be presented in the desired record, whether any one of some chosen words should be present. Nested Boolean search facilities allow more complex conditions to be imposed along with the search terms.

All text retrieval systems, including online search services like DIALOG and CD-ROM database provide Boolean search facilities.18.

The Boolean query is the de Facto methodology used in information retrieval (Frants , Shapiro, Taksa,and Voiskunskii, 1999).

Studies of web searcher have usually focused on very large search engine logs files (e.g. Jansen and Pooch, 2001; Jansen et al., 1999; Spink et al., 2000). In these studies, the focus has understandably been on the quantitative data (e.g. number of search term used), and not the search tasks the uses were trying to do, the characteristics of the users who were formulating the queries, the successfulness of their searches, or the concepts that the users used in their queries. In general, the log studies have shown that web searchers use short queries (typically from 1to3 terms), Seldom use advanced operators, do not regularly iterate their queries, and only go through a couple of result pages per query.

The familiarity with the topic of the search task also affects the queries the users formulate: as the users becomes more familiar with the topic, the queries they formulate become longer and more detailed (Vakkari, 2000). Nevertheless, the story is not quite that simple. In another study by (Holscher and Strube 2000), users with less topic experience formulated longer queries that the users with more experience.

The authors assumed that the domain experts knew more appropriate terms and thus, needing fewer of them. However, this assumption was not studied in more detail.

has nicely illustrated the challenge the users face in text-based information retrieval: “How to guess what words to use for the query that will adequately represent the person’s problem and be the same as those used by the system in its representation.” For some users, this task is presumably easier than for others and our goal is to study the user characteristics affecting the “guesses” they make. Thus, we designed an empirical study to study the factors affecting initial query formulation (Belkin, 2000).

The classic concerns of research into statistical and probabilistic retrieval technique have been automatic indexing and abstracting, automatic classification, and automatic searching. These concerns are all represented in the experiments carried out by Salton with SMART retrieval system (Ellis, 1996).

Statistical systems usually allow the user to enter queries in plain English without command. They substitute smart programming for some of the knowledge that professional searchers have been required to learn (Tseng, 1996).

Users, particularly professionals, prefer Boolean query models. Boolean queries are precise: a document either matches the query or it does not. This offers the user greater control and transparency over what is retrieved. And some domains, such as legal materials, allow an effective means of document ranking within a Boolean model: Westlaw returns documents in reverse chronological order, which is in practice quite effective. In 2007, the majority of law librarians still seem to recommend terms and connectors for high recall searches, and the majority of legal users think they are getting greater control by using them (Christopher, et.al.,2008).

However, this does not mean that Boolean queries are more effective for professional searchers. Indeed, experimenting on a Westlaw sub collection, (Turtle 1994) found that free text queries produced better results than Boolean queries prepared by Westlaw's own reference librarians for the majority of the information needs in his experiments. A general problem with Boolean search is that using AND operators tends to produce high precision but low recall searches, while using OR operators gives low precision but high recall searches, and it is difficult or impossible to find a satisfactory middle ground (Turtle, 1994).

The performance of the Medical Literature Analysis and Retrieval System (MEDLARS) of the US National Library of Medicine was analyzed between August 1966 and July 1967. The test was conducted on the operational database of MEDLARS, a database of biomedical articles, with index entries being drawn from MeSH, a thesaurus of medical subject headings. The objective of the MEDLARS test was to evaluate the existing system and to find out how it could be improvised. Twenty-one user groups were selected from the user community that would supply some test questions, cover all kinds of subjects in the request, and cover all categories of users (Salton, and McGill, 1986)

The college of Librarianship Wales began an investigation into indexing systems, supported by a grant from the office for Scientific and Technical Information (now the British Library Research and Development Department) in 1968. Various types of indexing languages, including post-coordinate and faceted classification schemes were tested, using a collection of items if the fields of library science and documentation. The conclusion reached was that the languages tested did not often exhibit significant differences in retrieval performance, effectiveness and efficiency, and no really large differences were observed. The uncontrolled languages tested performed overall just as well as the controlled languages. The basic result was, therefore, similar to the Cranfield Project in that no one system was found to be markedly better than any other. One of the incidental discoveries of the Aberystwyth project was that where single terms are concerned, it can be helpful to know the context in which the term is being used (Keen and Digger, 1972).

The search language of CDS/ISIS is based on Boolean algebra, which provides a convenient way of expressing logical operation between classes. Each search term associated with a given record, in fact, can be viewed as representing the class of all those records associate with that term. Thus by expressing logical operations between search terms you can define precisely the class of records to be retrieved in response to your needs (Buxton and Hopkinson, 2001).

REFERENCES

Aryal, Bishnu Prasad (2005). *Subject Heading, its Application for Consistent and Uniform Subject Entries for Literature Published in Nepali Language*. Unpublished master's thesis, Central Department of Library and Information Science, Tribhuvan University, Kirtipur, Katmandu, Nepal.

Belkin, N.J. (2000). Helping people find what they don't know. Communications of the ACM. (pp. 58-61).

Blair, D.C. (1986). *Indeterminacy in the subject access to documents, information processing and management*. (pp. 229-234).

Buxton, Andrew and Hopkinson, Alan (2001). *The CDS/ISIS for Windows handbook*. New Delhi : UNESCO.

Chowdhury, G.G. (1990). *Information Retrieval System*. Calcutta: IASLIC.

Christopher, D. M., Prabhakar Raghavan and Hinrich Schütze (2008). *Introduction to Information Retrieval*. Cambridge University.

Ellis, David. (1996). *Progress and problems in information retrieval*. London: Library association publishing.

Ellis, David. (Eds.). (1996). *Progress and problems in information retrieval*. London:

Frants, V.I., Shapiro,J., Taksa,I. and Voiskunskii, V.G. (1999). Boolean search: Current state and perspectives. *Journal of the American Society of information science, 50*(1), 86-95.

Gull,C.D. (1956). Seven Years of Work on the Organization of Materials in the Special Library. *American Documentation,7,* 320-329.

Chakrabortty, A.R. and Chakrabortty, B. (1984). *Indexing: Principles,processes,and products* (pp.76). Calcutta: World Press.

Chakrabortty, A.R. and Chakrabortty, B. (1984). *Indexing: Principles,processes,and products* (pp.91-93). Calcutta: World Press.

Harter, S.P.(1988). *Statistical Approaches to Automatic Indexing: Document Retrieval System* (pp.81-94). London: Taylor Graham.

Keen, E.M. and Digger, J.A. (1972). Aberystwyth Projects. *Report of an information science index languages text. Aberystwyth College of Librarianship Wales.* Part 1.

Lancaster, F.W. (1979). *Information retrieval system: Characteristics, testing, and evaluation* (2nd ed.). New York: John Wiley and Sons.

Lancaster, F.W. (1986). *Vocabulary Control for Information Retrieval* (2nd ed.). Arlington: Information Resources.

Lancaster, F.W. (1991). *Indexing and Abstracting in Theory and Practice.* (pp.18-42). London: Library Association.
Library association publishing.

Miller, Joseph and Goodsell, Joan (2004). *Sears List of Subject Headings* (18th ed.). New York: The H.W. Wilson.

Nayaichyai, Lila (2006). *Manual v/s computerized cataloging.* Unpublished master's thesis, Central Department of Library and Information Science, Tribhuvan University, Kirtipur, Katmandu, Nepal.

Ranganathan, S.R. (1964). *Classified Catalogue Code: With Additional Rules for Dictionary Catalogue Code.* (pp.44-46). Bangalore: Sarada Ranganathan Endowment for Library Science.

Salton, Gerand, and McGill, M.J. (1986). MEDLARS. *Introduction to modern information retrieval.* New York: McGraw-Hill.

Tillett, B.B. (2004). *Library of Congress Subject Heading* (27th ed., vol.). Washington DC: Library of Congress.

Tseng, Gwyneth. (1996). Boolean Logic. *International Encyclopedia of Information and Library Science*. London: Routledge.

Turtle, Howard (1994). *Natural language vs. Boolean query evaluation: A comparison of retrieval performance.* Otago: ACM Press.

Vakkari, P. (2000). Cognition and changes of search terms and tactics during task performance: A longitudinal case study. *Proceedings of the RIAO 2000 Conference. Paris, France*, pp. 894-907.

Chapter III

FOCUS OF THE STUDY

In the ever growing World Wide Web; large numbers of text retrieval databases and search engines are necessary tools for efficient information access. With the increasing amount of such systems, information searching process becomes a complex process and it requires at least four main steps: problem identification; need articulation; query formulation and result evaluation. This process is affected by environmental (e.g. Database and search topic); searcher (e.g. online search experience); search process (e.g. technique use); and search outcome variables (e.g. precision and recall).

Typically the studies focusing on information search strategies (or specifically query formulation) have studies processional searching from bibliographical database and search engines on the personal experience of librarians and general users. General users of web search engines are a very different population and need to be studied independently for complete understanding of their search strategies. Professionals (especially librarians) have plenty of training and experience in query formulation, where as the web user's often vague and imprecise descriptions of the users underlying information need. Furthermore the lack of organization of the documents / information in the web or database makes it impossible to formulate efficient queries by considering the contents of database, indexing terms, or controlled vocabularies to the same extent as bibliographical database. Thus, even professional searcher might need to use different strategies in web information search as compared to searching from bibliographical database.

3.1 Information retrieval system

The general task of information retrieval (IR) is searching for information in documents. Here “documents” is a general term, which refers to unstructured records in a database and/or search engine. It can be a text document, an image, a video clip, some web pages, and etc. al. The major difference between an IR search and a traditional search is that latter one is usually focus on structured data. Text search perhaps is the most sophisticated area in IR. Its technique usually falls into two categories, statistical approaches and Natural Language Processing (NLP) approaches. The former category usually tokenizes the documents into words, which is the basic

element for statistical processing. Variations to this approach extend the role of words to terms. Terms are not restricted to be the words of the documents; examples are n-gram (consecutive string of n characters). A large corpus is usually needed for statistic purpose. The NLP approach employs rules of syntax and semantic level analysis of documents. And NLP is language sensitive. Of course, the boundary is not sharp and these two categories are often interleaved. Statistical approaches dominate operational IR systems. Detailed models includes Boolean, extended Boolean, probabilistic, and vector space (Jin, French, and Michel, 2005).

3.2 Information retrieval models

There are various retrieval models in an information management system. Among them various models are based on computerized systems along with mathematical procedures. It should remembered that in computerized informational retrieval systems, the judgment as to whether a document is relevant or not to a given query is based on the topicality or lexical similarity between the query terms and the organized set of document. This includes Boolean search model; Statistical or Probabilistic retrieval model; Vector processing model; and Natural Language Processing.

3.2.1 Boolean retrieval model

Most informational retrieval systems work on the principle of text matching, where by a search terms is input and the retrieval system returns a set of records from the database that contain the term in question (Tseng, 1996).

Similarly a Boolean search matches the terms in a document with the terms in a user query. Boolean searching is based on Boolean logic. The searcher is trained to make semi mathematical statements in order to enter a search request or query (Kent, 2004).

The basic Boolean operators coordinate two or more terms in a query. But in practice, however, few search topics can be adequately expressed by a single word or short phrase, and Boolean logic is used as a means of combining brief search terms in order to put a more complex or detailed search expression to the database (Kent, 2004).

Most information retrieval systems offer three Boolean connectors (Boolean operators) to link search terms: AND, OR, and NOT.

AND

And is a Boolean operator used to narrow our search by ensuring that all keywords used appear in the search results. It indicates that the two words connected by the AND operator must both appear in a document in order for that document to be retrieved. It means that it allows the searcher to specify the coincidence of two or more concepts. Since the Web and Database is already huge, it is important to use AND effectively for precise and prompt result.

And is used to link separate concepts to build up a compound search topic. E.g. the 'design' of 'kitchen application' for the 'physically handicapped' (three concepts). The Boolean AND connections is sometimes referred to as intersection ($\cap$) or conjunction ($\wedge$). This terminology and the corresponding notation come from two branches of mathematics which have close links with Boolean logic: set theory and propositional logic, respectively (Tseng, 1996).

Searching or retrieving process

A database being searched may comprise controlled or uncontrolled vocabulary. (Cleverdon, 1988) mentions that a user going to search a database which has controlled index language has to perform the following tasks:

1. To decide the words that might be used by the authors of the relevant documents;
2. To decide which particular database is to be searched;
3. To use the thesaurus of the chosen databases in order to translate the query terms in the appropriate way;
4. To guess which of the chosen terms (or concepts) might have been used by he database indexer;
5. To coordinate the terms (often using Boolean operators) to formulate the search statement;
6. To input the search statement;
7. To repeat the steps 5 and 6 until a desirable set of output is obtained or the search fails together;
8. To identify the actual relevant items from among the retrieved ones;

It may be noted that one major task in the searching process relates to the coordination of terms (step 5 above) in order to formulate the actual search statement. The result of the search depends largely on how adequately the search terms are combined. Sine the beginning of mechanized information retrieval, Boolean search techniques have been used widely. Major aspects of Boolean searching such as idea about coordination and application of Boolean logic can be understood with Venn diagram and Truth tables and is discussed as follows.

Example#1

Query: I'm interested in the relationship between poverty and crime.

Poverty AND *Crime*

In this search, we retrieve records in which both of the search terms are present.
This is illustrated by the shaded area overlapping the two circles representing all the records that contain both the word "poverty" and the word "crime."
Notice how we do not retrieve any records with only "poverty" or only "crime."

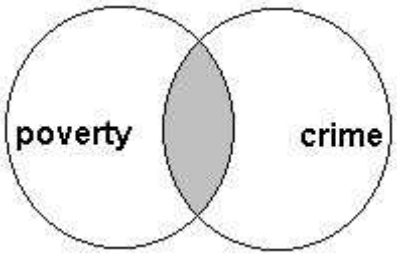
Here is an example of how AND logic works:

Table: 3.1 Example for AND logic

Search terms	Results
poverty	783.447
crime	2,962,165
poverty AND crime	1,677

This can be illustrated in Venn diagram as:

Fig: 3.1 Venn diagram for AND *operator*



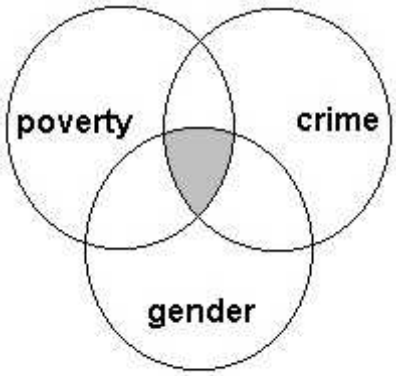
The more terms or concepts we combine in a search with AND logic, the fewer records we will retrieve.

Table: 3.2 Example for more than two keywords AND logic

Search terms	Results
poverty	783.447
crime	2,962,165
poverty AND crime	1,677
poverty AND crime AND gender	76

This can be illustrated in Venn diagram as:

Fig: 3.2 Venn diagram for AND *operator for three keywords*



OR

Or is a Boolean operator used to broaden our search by retrieving any, some, or all of the keywords used in the search statement. It indicates that either of two (or more) terms must appear. Since the Web and Database is already huge and complex arrangements of inundated information day to day, so using OR helps us to make sure we aren't missing anything valuable. But distinguish the relevant one among them is too difficult.

Or is used to link together synonyms, lexical and morphological variants, and terms which are close in meaning in the context of a particular search. The OR connector is often overlooked by novice users of electronic information retrieval systems, yet it can be essential for successful retrieval because the words and phrase used to describe the same subject in different documents can vary enormously. Thus the searcher should anticipate common variants to each search term and join them with the OR connector until successful query formulation or before using the AND connector. The Boolean OR connector is sometimes referred to as union (⋃) (Tseng, 1996).

Example#2

Query: I would like information about college.

College OR *University*

In this search, we retrieve records in which AT LEAST ONE of the search terms is present. We are searching on the terms *college* and also *university* since documents containing either of these words might be relevant.

This is illustrated by :

The shaded circle with the word *college* representing all the records that contain the word "college"

The shaded circle with the word *university* representing all the records that contain the word "university"

The shaded overlap area representing all the records that contain both "university" and "college"

OR logic is most commonly used to search for synonymous terms or concepts.

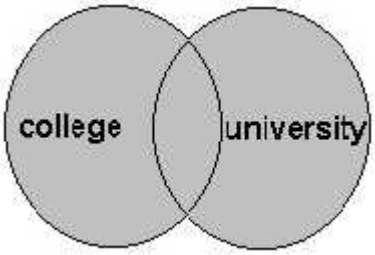
Here is an example of how OR logic works:

Table 3.3 Example for OR logic

Search terms	Results
college	17,320,770
university	33,685,205
college OR university	33,702,660

This can be illustrated in Venn diagram as:

Fig: 3.3 Venn diagram for OR *operator*



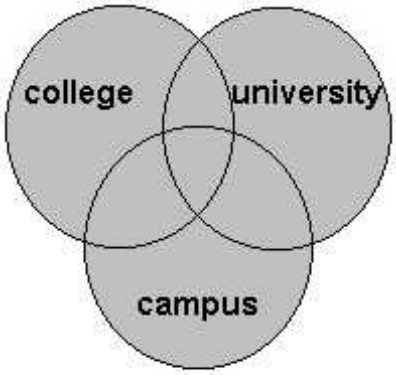
OR logic collates the results to retrieve all the unique records containing one term, the other, or both. The more terms or concepts we combine in a search with OR logic, the more records we will retrieve.

Table 3.4 Example for more than two keywords OR logic

Search terms	Results
college	17,320,770
university	33,685,205
college **OR** university	33,702,660
college **OR** university **OR** campus	33,703,082

This can be illustrated in Venn diagram as:

Fig: 3.4 Venn diagram for OR *operator for three keywords*



NOT

Not is a Boolean operator used to eliminate an unwanted concept or word in our search statement. It directs the system to reject any document that contains the term following the NOT operator (e.g., "A not B": A must be present, but B must not be present in a document in order for the document to be retrieved).

The obvious use of the NOT connector is to avoid retrieving irrelevant documents. There is however some danger in doing this because useful items can be eliminated too. Moore helpfully, NOT can be used to remove from a subsequent set those items which have already been retrieved, to avoid the nuisances and possibly the cost of viewing them a second time. Alternative terms and notation for NOT are complement (^).

Example#2

Query: I want to see information about pets, but I want to avoid seeing anything about cats.

Cats NOT *Dogs*

In this search, we retrieve records in which ONLY ONE of the search terms is present.
This is illustrated by the orange area with the word "cats" representing all the records containing the words "dogs"
No records are retrieved in which the word "cats" appears, even if the word "dogs" appears there, too.

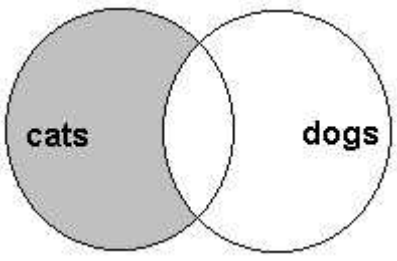
Here is an example of how NOT logic works:

Table 3.5 Example for NOT logic

Search terms	Results
cats	4,556,515
dogs	3,651,252
cats **NOT** dogs	81,497

This can be illustrated in Venn diagram as:

Fig: 3.5 Venn diagram for NOT *operator*



NOTE: NOT logic excludes records from your search results. Be careful when you use NOT: the term you do want may be present in an important way in documents that also contain the word you wish to avoid.

Truth Table

A truth table shows the resulting value when a logical operator is used to join two propositions, forming a new, complex proposition.

Suppose the two propositions being joined are *P* and *Q*. Each of these propositions will have two possible truth values: true, or false. This gives us four possible combinations. These are represented on a table, as follows:

Table 3.6 Format of truth table

P	Q	
T	T	
T	F	
F	T	
F	F	

In the space to the right, a complex proposition is displayed. Beneath the complex proposition are the truth values which result given the four possible truth values of *P* and *Q*. For example, here is the truth table for the complex proposition *P* AND *Q* and *P* OR *Q*

Table 3.7 Results of truth table for AND, OR, and NOT operators

P	Q	P AND Q	P OR Q	P NOT Q
T	T	T	T	T
T	F	F	T	T
F	T	F	T	F
F	F	F	F	F

Notice: that the complex proposition may be true or false depending on the different truth values of *P* and *Q*. Thus, if we know what the truth values of *P* and *Q* are, we know what the truth value of *P* AND *Q* or *P* OR *Q* or *P* NOT *Q* is.

The classic Boolean retrieval model is based on logic predicates of terms. Define P(T) as a predicate which asserts that a term T appears within a document. Then we can connect a group of such predicates by AND, OR, NOT relations. The document's relevance is then calculated on the value of these predicates. In practice, the Boolean syntax can be extended. For example, proximity is supported by the predicate. Sometimes, even more complex relation can be defined, such as the sequence of terms or nested proximity where proximate terms must appear within a specified term distance of other terms (Jin, French, and Michel, 2005).

3.2.2 Vector space model

Vector Space Model (VSM) (Slaton and McGill, 1983) is a widely researched but not generally applied (in practice) retrieval model. The basic idea is first create a vector space, whose dimensionality is equal to the number of terms appearing in the corpus. Each document is mapped to a vector, whose component reflects the corresponding term's weight in that document. This weight can be calculated based on term-frequency in that document (named TF) and the term's important factor (named IDF), which is a global statistic of the corpus. Finally, the query itself is also mapped to a vector and the similarities between query and documents are calculated according to some similarity function. The results are output in a similarity ranked order. There are many variations of vector space model. Different weighting schemes, normalization methods, and similarity functions are proposed within the same frame- work (Baeza-Yates and Ribeiro, 2005).

The vector processing model assumes that an available term set, called term vectors, is used for both the stored records and information requests. Collectively the terms assigned to a given test are used to represent text content (Salton, Singhal, and Allan, 1996).

The documents are the objects in the collection and each of which is represented by a number of indexed terms or set of weighted terms (term vectors). The similarity between two objects is normally computed as a function of the number of the properties that are assigned to both objects; in addition, the number of properties that is jointly absent from both the objects may also be taken into account. Substantially similar methods can be used for retrieving information by comparing the query

vectors with the vectors representing the stored items and retrieving items that are found to be similar to the queries.

In the vector space model text is represented by a vector of *terms*. The definition of a term is not inherent in the model, but terms are typically words and phrases. If words are chosen as terms, and then every word in the vocabulary becomes an independent dimension in a very high dimensional vector space. Any text can then be represented by a vector in this high dimensional space. If a term belongs to a text, it gets a non-zero value in the text-vector along the dimension corresponding to the term. Since any text contains a limited set of terms (the vocabulary can be millions of terms), most text vectors are very sparse. Most vector based systems operate in the positive quadrant of the vector space, i.e., no term is assigned a negative value.

To assign a numeric score to a document for a query, the model measures the *similarity* between the query vector (since query is also just text and can be converted into a vector) and the document vector. The similarity between two vectors is once again not inherent in the model. Typically, the angle between two vectors is used as a measure of divergence between the vectors, and cosine of the angle is used as the numeric similarity (since cosine has the nice property that it is 1.0 for identical vectors and 0.0 for orthogonal vectors). As an alternative, the inner-product (or dot-product) between two vectors is often used as a similarity measure. If all the vectors are forced to be unit length, then the cosine of the angle between two vectors is same as their dot-product. If $\vec{D}$ is the document vector and $\vec{Q}$ is the query vector, then the similarity of document D to query Q (or score of D for Q) can be represented as:

$$\text{Sim}\left(\vec{D},\vec{Q}\right) = \sum_{ti \in Q.D} WtiQ.WtiD$$

where $WtiQ$ is the value of the I th component in the query vector $\vec{Q}$ and $WtiD$ is the I th component in the document vector $\vec{D}$ (Since any word not present in either the query or the document has a $WtiQ$ or $WtiD$ value of 0, respectively, we can do the summation only over the terms common in the query and the document.) How we arrive at $WtiQ$ and $WtiD$ is not defined by the model, but is quite critical to the search effectiveness of an IR system. $WtiD$ is often referred to as the *weight* of term I in document D.

Limitation of vector model

In addition to the general problem of dynamically changing database and the effects on weighting factors, there are problems with the vector model on assignment of a weight for a particular processing token to an item. Each processing token can be viewed as a new semantic topic. A major problem comes in the vector model when there are multiple topics bing discussed in a particular item. For e.g. assume that an item bas an in depth discussion on OIL in MEXICO and also COAL in PENNSYLVANIA. The vector model does not have a mechanism to associate each energy source with it's particular geographic area. There is no way to associate correlation factors between terms. Since each dimension in a vector is independent of the other dimensions. Thus the item results in a high value in a search for COAL in MEXICO (Kowalski and Mayburg, 2000).

3.2.3 Statistical and probabilistic retrieval model

The basic underlying tenet of the probabilistic approach to retrieval is that, for optimal performance, documents should be ranked in order in order of decreasing probability of relevance or usefulness to the user.

At the same time that Boolean search systems were being developed, such researchers as Gerard Salton and R.J. Tritschler were experimenting with statistical search engines. These search engines use statistics and probability to predict the similarity of any documents in a database to a query. This methodology can be quite complex, and the algorithm for computing the degree of relevance of any document in a database to a query differs from one search engine to another. The underlying assumption, though, is that the more times a term appear in a document, the more likely it is that the document will be about that subject. This is known as term frequency (TF). The second assumption on which these systems is based is that terms that appear more frequently in a document than they do in the database as a whole further indicate that the term or word in question is a major topic of that document. If the term appears frequently in the document, but infrequently in the database as a whole, the chances are that the document is about that subject. This measure is known as inverse document frequency (IDG) (Kent, 2004).

Since many documents can be said to be somewhat related to a subject, as they are in Boolean OR search, statistical systems often retrieve large sets of documents to

answer a query. Unlike the Boolean OR search, however, they use statistical methods for computing the degree of relevance to a query for each document based on TF/IDF and display the documents retrieved in ranked order, with the most relevant first. In other words, documents that contain more occurrences of a query term will be ranked higher than those with fewer occurrences of that term.

Statistical search engines present two advantages to the user. The first is that relevance ranking is a useful approach for retrieving a fairly large set of documents without confusing or overwhelming the user. The most relevant documents should be in the top group of documents; a user need not wade through all the documents in order to find the ones that are most useful. In addition, statistical systems are not exact match systems.

Advantage

The big advantage to these systems is that they add a degree of flexibility and fuzziness to the search process. It enlarges a search beyond the boundaries that the query originally defined. Since the initial queries rarely describe an information need well, the partially relevant materials can broaden the scope of search in order to find materials that are ultimately of greater use to the user than a Boolean AND operation would be. Statistical systems often enable the user to ask for "more like this one"- using a "perfect" document from an initial list of retrieval results to act as an enriched query for the net search iteration. The statistical systems find all the documents located by a Boolean OR search as well as some that contained misspellings or alternate forms of the query teams. The ranked retrieval set allows the user to find the most relevant documents first. This is an effective method for giving the user precise matches but also approximate matches so that he can enlarge the scope of his query. The technique solves the problem of handling large retrieved sets. The following figure indicates most relevant documents was located outside the Boolean AND set. It contained only two of the three query terms specified (Kent, 2004).

Limitation

Statistical systems have their own shortcomings. The foremost is that they ate not designed to work well no document records that do not contain enough text. Bibliographical records such as those that appear in a typical library catalogue thus not good candidate for statistical systems. In addition, statistical retrieval systems tend

to return large sets which may overwhelm the user. Also since the statistical search engines are based on frequency of words within a document, unless some sort of normalizing algorithms is applied, short documents can get short shrift. Authors who use many synonyms for their writing, even if a document is highly pertinent to a searchers query.

These systems are more computationally complex than the straightforward matching of the Boolean systems, and numerous factors are adjusted differently for each search engine, so that statistical search engines rarely retrieve exactly equivalent sets from the same database, since their algorithms differ (Kent, 2004).

3.4 Natural language processing and information retrieval

Natural Language Processing (NLP) is that area of research and application that explores how natural language text that is entered in to a computer system can be manipulated and transformed in to a form more suitable for further processing. Automatic natural language processing techniques have been identified as a desirable feature of information retrieval. The aim of an information retrieval system is to retrieve documents in response to a users request in such a way that the contents of the documents are relevant to the user's requirement. It is therefore, no wonder that with the help of sophisticated natural language processing techniques, we should be able produce representation of documents and queries for efficient retrieval.

In a natural language environment, the user is quite often unsure of what s/he exactly needs, but they will be able to judge whether a given document is relevant to their information need. The system should therefore be able to accept natural language statements as the expressions of the users needs, or it should make provisions for a man-machine dialogue through which the system can identify the exact requirement of the user and can take necessary measures for searching. These kinds of systems are known as the natural language interface or front-end systems. Most automatic retrieval systems based on natural language processing techniques, converts the content of the document files and users queries in an internal form and the task of matching takes place at that level.

Swartz 4 mentions that the process of building computer programs that understands natural language involves three major problems as thought process, representation,

and world knowledge and three kinds of knowledge as syntactic knowledge, semantic knowledge, and pragmatic knowledge (Lancaster, 1979).

3.5 Boolean retrieval system facilitated libraries

There are various libraries in Nepal which have their own advanced database facilitating the Boolean retrieval system, but this study is limited within the four major libraries situated inside the Katmandu valley naming as Tribhuvan University Central Library (TUCL), International Centre for Integrated Mountain Development Library (ICIMODL), Social Science Baha Library (SSBL), and Katmandu University School of Management Library (KUSOML).

3.5.1 Tribhuvan University Central Library (TUCL)

Tribhuvan University Central Library (TUCL) was established in 1959. It is the largest library in Nepal in terms of collection, services and the number of members. It began with a collection of 1200 volumes of book and now it has more then 3, 00,000 volumes of documents. It serves various types of users mostly university students, professors, researchers etc. In the library electronic databases is put online for local users as a computerized bibliographic database which can be retrieved with 5 different terminals while searching the bibliographic information. 'TUCL has started its electronic database from 1993. CDS/ISIS and WINISIS software are used to manage the electronic database, under Nepal Automation Project Through initiation of IDRC, Canada.

3.5.1.1 TUCL Database with Boolean search facilitation

Since 1995 the library has maintained TUCL Masters Database of the documents processed by the library to allow searching form their materials at computer terminals placed in different location. A database of 42000 documents can be accessed from the library's home page www.tucl.org.np. A database search icon is available in its homepage, where user can get the Boolean search facility to obtain the precise and prompt information. Guided and Expert search provision caters to both novice and expert users respectively. An example of collection database icon with Boolean search operators is presented at *annex (i)*. Except its collection database TUCL has following list of databases:

i. ISBN database.

ii. Article database

iii. Tribhuvan University Archive database

iv. Audio Visual Materials database

v. Serial Database

3.5.2 International Centre for Integrated Mountain Development Library (ICIMODL)

The ICIMOD is the first and only international centre devoted to integrated mountain development. ICIMOD was established because of widespread international recognition, especially in the eight nations of the Hindu Kush-Himalayan (HKH) Region, of accelerating degradation of fragile mountain habitats under severe population pressures, with the consequent increase of impoverishment in mountain communities. Based on a 1981 agreement between the government of Nepal and United Nations Educational, Scientific and Cultural Organization (UNESCO), ICIMOD was inaugurated in December, 1983, and began operating in September, 1984. The primary objectives of the centre are to help promote the development of an economically and environmentally sound mountain ecosystem and to improve the living standards of mountain populations.

3.5.2.1 ICIMODL Database management

ICIMOD has concentrated on the development of a comprehensive documentation centre and library with computerized bibliographic and serial database. ICIMOD also publishes documents on original research, knowledge reviews, workshops, and training courses. Lists of recent publications and other information can be found in its online homepage www.icimod.org/library. ICIMOD also functions as GRID Katmandu, one of the canters of the United Nations Environment Program (UNEP)/ THE Global Resource Information Database (GRID).

ICIMOD books -online provides direct access to all ICIMOD technical and scientific publications. It holds full- text and chapter- wise download options for publications published from 2000 onwards and some selected earlier publications. ICIMOD books online can be searched using full-text contents, title, year of publication, keywords, language, author and broad subjects. If you are looking for a specific book, use advanced search options. Combine multiple entries to make the result more precise. The Boolean search format 'OR', 'AND', 'NOT' results showing all the selected keywords according to their characteristics. Its provision and application is available in *Annex-5*

3.5.3 Social Science Baha Library (SSBL)

The Social Science Baha was set up in January 2002 to foster and facilitate the development of the study of the social science in Nepal. SSBL formally opened to the public in Oct. 31st 2003. The SSBL has been able to develop a good collection covering almost all disciplines of Social Sciences through purchase, gift and permanent loan. It has both: Conventional and Electronic resources. The conventional resource consists of 23099 volumes of books, documents, journals, etc.

3.5.3.1 Social Science Baha Library (SSBL) Database:

There are three bibliographic databases for accessing the available resources: (i). Books, documents, reports, etc, (ii). Journals (iii). Journal articles. Regarding electronic resources, it has online access to a vast treasure of scholarly journal literature in various disciplines through international online database: like JSTOR and AGORA, Blackwell Synergy, EBSCO, EMERALD and Oxford University Press through PERI. Further, the library is connected to DELNET, Delhi that provides access to bibliographic records and inter-library loan service to its users.

Database with sufficient subject headings and keywords have made easier to retrieve the exact documents to their users. Approximately 22,000 recorded documents are available in its database, which caters its local users with Boolean operator search provision. Those types of documents are following:

i. Book, documents etc.
ii. Journal
iii. Journal Article etc.

3.5.4 Kathmandu University School of Management Library (KUSOML)

Kathmandu University is an autonomous, not-for-profit, non-government public institution. KU is an institution of higher learning dedicated to maintain high standards of academic excellence. The main University complex is located in Dhulikhel. KU has the different academic programs, at present it offers various undergraduate, graduates, and postgraduate programs in management, science, engineering, arts, education and medical sciences.

Kathmandu University School of Management library is one of them and it has huge collection of information which are managed systematically through the barcode based software, SOUL. Users can find their required information at the location given in the computer bibliographic records database.

3.5.4.1 Kathmandu University School of Management Library (KUSOML) Database:

Total collections of the documents are recorded in the bibliographic database or online databases, 10,000 volumes of books, documents, journals, etc. Regarding electronic resources, it has online access to a vast treasure of scholarly journal literature in various disciplines through international online database: like JSTOR and AGORA, Blackwell Synergy, EBSCO, EMERALD and Oxford University Press through PERI

REFERENCES

Baeza-yates,Richardo and Ribeiro-netom, Berthier. (2005). *Modern information retrieval.* London: Longman.

Cleverdon, C. (Ed.). (1988) *Optimizing convenient online access to bibliographic databases. In : p. willett, ed. Document retrieval system*, London: Taylor Graham.

Jin, Xiangyu , French, James, and MIchel, Jonathan (2005). *Query formulation for information retrieval by intelligence analysts.* Retrieved June 24, 2008, from Virginia University, Department of Computer Science Web site: http://www.cs.virginia.edu-techrep.cs

Kent, Allen (Eds.). (2004). Encyclopedia of library and information science (Vol. 64, pp. 218-246). NewYork: Marcel dekker.

Kowalski, Gerald J. and Mayburg, Mark T. (20000. *Information storage and retrieval system: theory and implementation*, (2nd ed.). Boston: Kbuwer academic publishers.

Lancaster, F.W. (1979). *Information retrieval system: characteristics, testing, and evaluation*. (2nd ed.). New York: John Wiley and Sons.

Salton,G., Singhal, A. and Allan, J. (1996). Automatic text documentation and structuring. *Information processing and management* ,*32*(2), 127-38.

Tseng, Gwyneth. (1996). Boolean Logic. *International Encyclopedia of Information and Library Science*. London: Routledge.

Chapter IV

RESEARCH METHODOLOGY

Research is the systematic study and investigation of a topic outside our own experience and knowledge, when we are doing research; we have to move from what we know about a topic to what we don't know. Thus it is the skills of the researcher that where to navigate to attempt the actual destination. Mean while the effort, think critically, and understanding capability of researchers and navigating techniques for the topic are the ability of researcher too. Research is rewarding, but it also demanding and time consuming. It requires discipline, strategic planning, careful time management, and a constant willingness to rethink ideas and reshape discussions (Kirszner, 2002).

There are two types of methodologies are accomplish in the social science research, pure research and applied research. The purpose of pure research or basic research is to seek new knowledge and to explore and make advance general scientific understanding. Applied research is conducted specially for the purpose of solving practical problems and improving the quality or skills of life. Applied research focuses on such concerns as methods to improve memory or increase skills of learning (Smith and Mackie, 1995). Therefore the research topic *"Boolean operator in logical and efficient information retrieval "* is more or less *applied research* because it focuses the ability or skills of librarians and general users in retrieving the information via the Boolean operators from various library database and search engines.

It is a careful search of inquiry into subject matter (searching behavior) which is an endeavor to discover or find out valuable fact, which will be useful for finding out the invade state of Boolean operators among the database and search engines and equally among the professional librarians and general users.

4.1 Research design

A research design is the specification of methods and procedures for acquiring the information needed. It is the overall operational pattern of framework of the project that stipulates what information is to be collected from which sources by what

procedure. It ensures that the information obtained is relevant to the research question and that it was collected by objective and economical procedure.

Firstly it is a plan that specifies the sources and types of information relevant to the research question. Secondly, it is a strategy specifying which approach will be used for gathering and analyzing the data. Finally, since most research studies have time and economic constraints, both time and cost budgets are typically included.

Information retrieval part is an important era of the library studies. All the professionals and user of libraries are associated with the information retrieving techniques. Therefore this topic and its sources of information are obtained from these associates on the basis of their daily experiences with the help of questionnaire and interview. For the comprehensive understand this study is divided in to professional group and user group. There is no value of comparisons among them except that the classification of study helps in generalizing and implementing the problems and their solutions individually and effectively.

Questionnaire and few interview type methods are applied for data collection. All the relevant questions related to the title are accompanied in questionnaire as for as possible. However, these approaches were not sufficient so frequently field visit and direct observation method was also used.

For analyzing part the focus is based on each question as response by the respondents, mean while rank correlation and significance (hypothesis) testing strategy is used. This strategy i.e. significance testing of hypothesis provides the validity for research, their results and their respective methods which is used for data collection and analysis etc. For each questions testing of hypothesis (chi-square test) is applied as for as possible on the basis of respective data and their analysis.

4.2 Population

The population of study was libraries which have the Boolean search facility with their database. Libraries were of either academic or public or government types and their professionals and users are represented as population. For search effectiveness whether they used Boolean operator or not, if they used, up to what level they using it, all of these information are categorized for the convenience of the study.

Therefore all government, public, academic and private libraries their professional staffs and users, located inside the Katmandu valley except that other professionals those haven't the Boolean query formulation in their respective database but identified with TUCL database are also accompanied as population for the study.

4.3 Sampling method

Social science research can focus on a specific population or complete set of units being studied (for e.g. all academic libraries inside the Katmandu valley) when time, cost, and accessibility often prohibit the collection of data from every member or about every item. In this situation it is necessary to select a "representative sample" of the population, one in which the same range of characteristics or attributes can be found in similar proportions. It is only with a truly representative sample that we can "generalize" the research findings to the whole populations. So judgments have to be made to ensure that the sample is as representative as possible adopting one of a number of different "sampling strategies" to go some way towards overcoming potential limitations. Therefore questionnaire were distributed to the professional of all libraries located inside the Katmandu valley, M. Lib Science students and TUCL library users as for as possible. But the focus provides towards the TUCL, ICIMOD, SSBL (Social Science Baha Library), and KUSOM (Katmandu University School of Management Library) with their professional staffs and users.

4.4 Data collection methods

Questionnaire:

A set of structured questionnaire was developed for the purpose of collecting data from the library users and professionals. The questionnaire contained subjective as well as objective. The questionnaire was distributed through the personal contact in which they were requested to fill up the questions. Same questions were distributed for professional staffs and general users on the assumptions that those whom can respond the questionnaire in better way they have the god knowledge or skills about Boolean retrieving technique whether they are professionals or general users or vice versa. But the analysis of the questions and data is made distinctly and ability were checked on their respective field only.

Altogether 195 questionnaires were distributed and the individual numbers according to the libraries is presented as follows.

Table: 4 Sample respondents profile

Name of libraries	TUCL	ICIMOD	SSBL	KUSOML	Others	Total
Professionals	15	4	3	3	45	70
General Users	45	15	15	15	35	125
Total	60	19	18	18	80	195

NOTE: Others represent the professional library staffs associated with the different academic, public, organizational, and governmental libraries situated inside the Katmandu valley. Except above four libraries the column for others corresponding to general users i.e.35 are the M. Lib. Science students were taken from Central Department.

Interview/Observation

Interview and observation method was specially focused for the general or the novice users of library database and search engines. Researcher had requested to coordinate their information need with the Boolean logical operators. Many of the users were not familiar with the Boolean logical operators and some of them were familiar but they rarely applied it in especially for search engine query formulation. This method was done because lack of proper knowledge among the various users, so researcher firstly tried to make understand about it and then only made it applied in their need.

4.5 Data analysis procedure

In the process of analysis data in the form of questionnaire have been collected, edited, coded, tabulated, and classified or categorized according to the respondent's responses. Simple percentage analysis has been used along with the statistical tools. The result of analysis has also presented in tables and different kinds of figures are used to express findings and conclusions.

4.6 Use of Spears man Rank Correlation

Some times the variables, which are to be correlated, can not be measured quantitatively. In other words the qualitative characteristics (attributes) such as ability,

intelligence, honesty, beauty, sociability cannot be measured quantitatively but can be arranged serially, we can however rank them in some order or assign sort of rating to measure these variables. In such a situation Charles Edward Spearman, a British psychologist, developed a formula in 1904 which consist in obtaining the correlation coefficient in between the ranks of n individuals in the two attributes under study. This is known as spearman rank correlation. Usually denoted by ° (Rho) is given by formula $° = 1 - \frac{6\sum d^2}{n(n^2-1)}$

Where, d = is the difference between the pair of ranks of the same individual in the two attributes or characteristics.

n = is the number of pairs.

4.7 Use of Chi-Square (¡ ^{2}) test

Chi square test is used to test whether more than two populations can be considered equal. Actually chi-square test allows us to do a lot more then just test for equality of several proportions. If population are classified into several categories with respect to two attributes. (For example: Key terms coordination and information retrieval), one can then use chi square test to determine if the two attributes are independent to each other.

Steps for the computation of Chi-Square (¡ ^{2}) test (Gupta, 1996)

Step1: Complete the expected frequencies $E_1, E_2, \ldots\ldots, E_n$ corresponding to the observed frequencies $O_1, O_2, \ldots\ldots\ldots O_n$ under some theory or hypothesis.

Step 2: Compute the deviation (O-E) for each frequency and then square them to obtain (O-E) ^{2}

Step 3: Divide the square of the deviations $(O-E)^2$ by the corresponding expected frequency to obtain $\frac{(O-E)^2}{E}$.

Step 4 : Add the values obtained in step (3) to compute $¡^2 = \sum\left[\frac{(O-E)^2}{E}\right]$.

Step 5 : Under the null hypothesis that the theory first the data well, the above statistic follows, ¡ ^{2} - distribution U= (n-1) d.f.

Step 6: Lookup the tabulated (critical) values of ¡ ^{2} for (n-1) d.f. at certain level of significance, usually 5% from the table. (Significance values of ¡ ^{2} is given in the table).

If the calculated value of ¡ ^{2} obtained is less than the corresponding tabulated value obtained in step (6) than it is said to be non significant at the required level of significance. This implies that the discrepancy between observed values (experiment) and the expected values (theory) may be attributed to change, i.e. fluctuations of sampling. In other words, data of not provide any evidence against the null hypothesis [given in step (5)], which may therefore, be accepted at the required level of significance and we may conclude that there is good correspondence (fit) between theory and experiment.

On the other hand, if calculated value of ¡ ^{2} is greater than tabulated value, it is said to be significant. In other words, the discrepancy between the observed and expected frequencies can not be attributed to change, and in the situation, null hypothesis can be rejected. Thus, one can conclude that the experiment does not support the theory.

REFERENCES

Gupta, S.K. (1996). *Fundamental of statistics.* (4th ed.). New Delhi: Himalayan.

Kirszner, L. G. and Mandell, S. R. (2002). *The Holt Handbook* (6th ed.). Fort Worth: Harcourt CP.

Chapter V

ANALYSIS AND PRESENTATION OF FINDINGS

The previous chapters incorporated introduction of the study review of literature, focus of the study, and the research methodology employed in the study respectively. This chapter incorporates analysis and interpretation of data. The data and information collected from the respondents are presented analyzed and interpreted in this chapter for attaining the stated objective of the study. The data and information collected from respondents (professionals and general users) are presented, interpreted and analyzed according to the research questions formulated for the study.

Out of 195 respondents only 160 of them respond to the questionnaires. Especially data were collected from the five libraries viz TUCL (Tribhuvan University Central Library), ICIMOD, SSBL (Social Science Baha Library), and KUSOML (Katmandu University School of Management Library). Besides the professional staffs of the libraries mentioned above professional staffs from other libraries of Katmandu valley were also taken for study purpose, to make widen the area of research and are denoted as (others) in the table below.

According to the questions in the questionnaire different suitable statistical tools were applied for the significance of the study. It is assumed that the tables, graph, charts, and statistical tools sufficiently and correctly represents those all responses which are classified on the basis of questions as obtained in the questionnaire with their relevancy. 70 questionnaires were distributed to professional staffs, 90 were distributed to general users and 35 questionnaires were distributed to the M lib science students of central department of library and information science.

Table No: 5.1 Respondents profile with their responses

Name of library	*Professionals*		*General users*		*Percentage*		*Total Returned*	
	Distributed	Returned	Distributed	Returned	P.S.	G.U.	No.	%
TUCL	15	14	45	30	93.33	66.66	44	100
ICIMOD	4	4	15	14	100	93.33	18	100
SSBL	3	3	15	8	100	53.33	11	100
KUSOML	3	3	15	8	100	53.33	11	100
OTHERS	45	36	35(M lib)	30	80	85.71	76	100
Total	70	60	125	90	94.66	84.72	120	100

According to this table, total 70 questionnaires were distributed, of which only 6o (93.33%) were returned with full information as required by the questionnaire. Out of total 25 questionnaires were distributed to four libraries as described in the table and 45 were distributed to other academic, government, and nongovernmental libraries situated inside the Katmandu valley, and researcher requested them to fill up these questions on the basis of their knowledge's, skills, and daily experience for information retrieving purpose. Hence, total 60 respondents were responded from the professional group. As professionals are associated with service providing agencies, thus any results obtained from them will not be limited up to them selves, but the results will indicate how proficient they are in providing the precise and prompt information to their consumers (users) in the respective fields.

Similarly 90 users were participated in this study from above four libraries among them 30 users were students of master levels in library and information science from central department of library and information science. Users were encouraged to fill up all these questions by themselves and only technical assistance was provided by the librarians of respective libraries. Most of the users were from master's level, and remaining were form research background in social science sector. So it was hoped that they have the good skills in searching and retrieving information via Boolean logical operators.

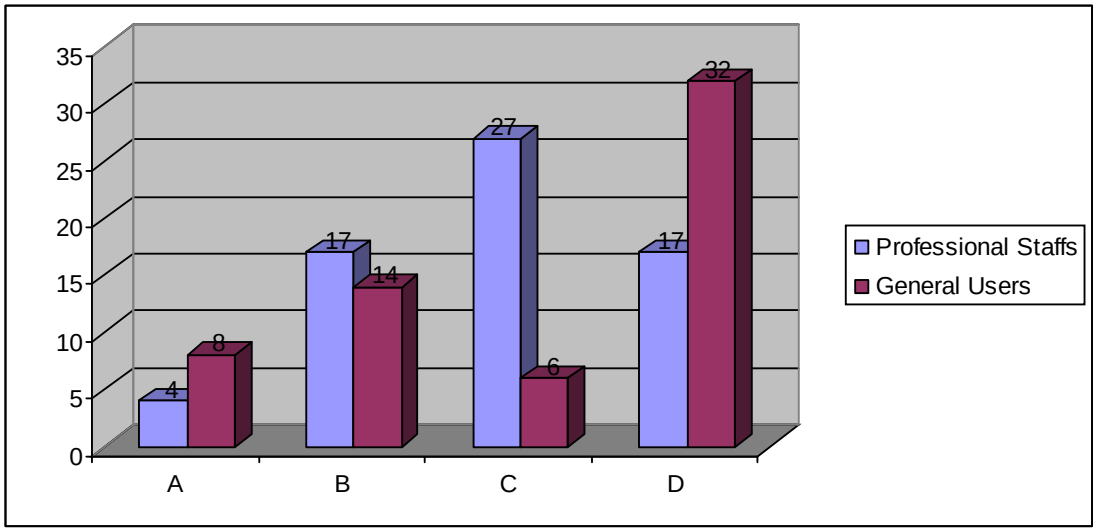
5.1 Respondents frequent requirement

In question no 2 researcher facilitated the users to rank the allocated options independently on the basis of their frequent requirement. It was based on the use of journals, books/abstracts, library database, and search engine. Among them top rated sources are ranked as 1 and lowest as 4 in decreasing order as their individual usefulness. The detailed description in table is as follows:

Table No: 5.2 Respondents ranked to their sources of information

Ranked	A	B	C	D	Total
Professional Staffs	4	17	27	17	60
General Users	13	19	11	47	90

Source field survey



This indicates that the there is high degree of negative correlation between the general users and professional staffs, (see appendix-1) which shows that sources for obtaining information varies with the general users and professional staffs. Since in question first two options were related with manual sources of information and later options were based on computerized search, so it is concluded that those persons involved in computerized search might have good knowledge coordinate searching and Boolean searching.

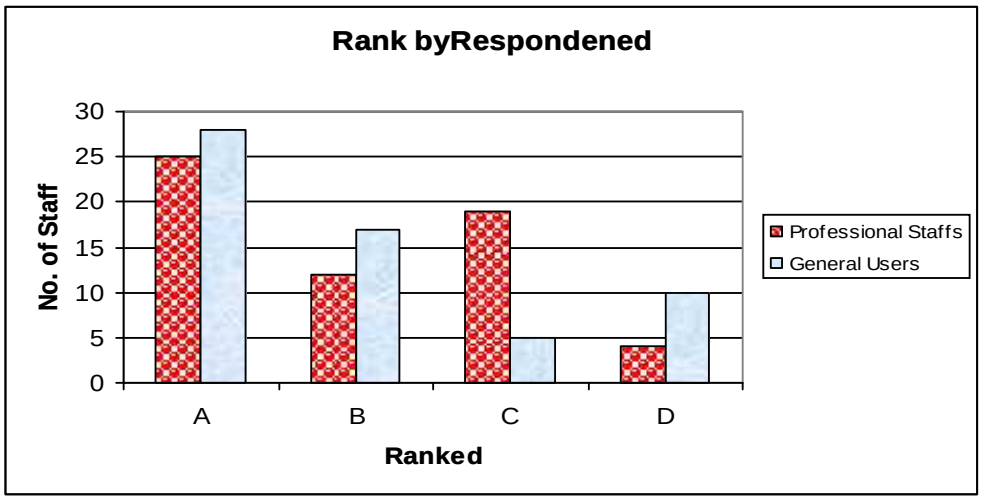
5.2 Factors affecting for successful query formulation

The same, statistical tool has been used for question 17 associated with the effectiveness of query formulation. The same tool was applied because this question exposes the abilities or skills of searching technique, which is difficult to measure via direct interpretation. The question for search and retrieval purpose was about experience, familiarity, and case dependency of information retrieval. Table representation is as follows:

Table No: 5.3 Respondents ranked to affecting factors for successful query fomulation

Ranked	A	B	C	D	Total
Professional Staffs	25	12	19	4	60
General Users	38	25	11	16	90

Source field survey



Since the ability of successful query formulation differs from individual to individual, its quantitatively measurement is not possible hence it is calculated according to the spearman rank correlation method. Recorded data indicates that there is low degree of positive correlation in between the techniques used or followed by the general users and professional staffs. Both the groups had gone top priority to experience for searching. But the professional ranked the Boolean as their second choice while the general users make it their last option. Familiarity with database and search engine was third option to professional staffs while it was second to general users. Professional staffs reported that the type of search task was used as last option while for general user it was third option.

5.3 Respondents searching habits

For question no 3 and 4 it was found that respondents used both kinds of searching i.e. manual and online whichever necessary and available. For manual type library (i. e. traditional library which have not providing computer facility) users used library

catalogue instead of library index. Similar response was obtained from all the respondents. There for it is concluded that respondents searching habits depends on types of library whether it is traditional or automated.

5.4 Preference for search engine

In question no 5, question was associated with the type of search engine that respondents used for online searching and it was found that most of the respondents used the Google search engine for obtaining the information they required.

5.5 Preference for library database

In question no 6, question was related with the type of library database that respondents used for online or remote or local login and it was found that maximum no of users and staffs used their respective libraries database. Some times professionals consulted the PEERI and JOSTER etc. database as per need, besides their respective libraries database.

5.6 Kinds of approach for navigation

For question no 7, question was associated with technique of navigating or approach of searching that respondents surrogates for their required information and it was found that library professional prefer keyword approach and general users focused on word/ term approach. This can be shown in following table according to individual libraries.

Table 5.4 Respondents profile for approach of information

Name of library	*Professionals*				*General users*			
	Keywords approach	%	Term approach	%	Keywords approach	%	Term approach	%
TUCL	10	71%	4	29%	5	17%	25	83%
ICIMOD	3	75%	1	25%	4	29%	10	71%
SSBL	2	66%	1	34%	3	37%	5	63%
KUSOML	3	100%	-	-	5	63%	3	37%
OTHERS	25	69%	11	31%	23	73%	7	23%
Total	43	76%	17	24%	40	44%	50	80%

Source field survey

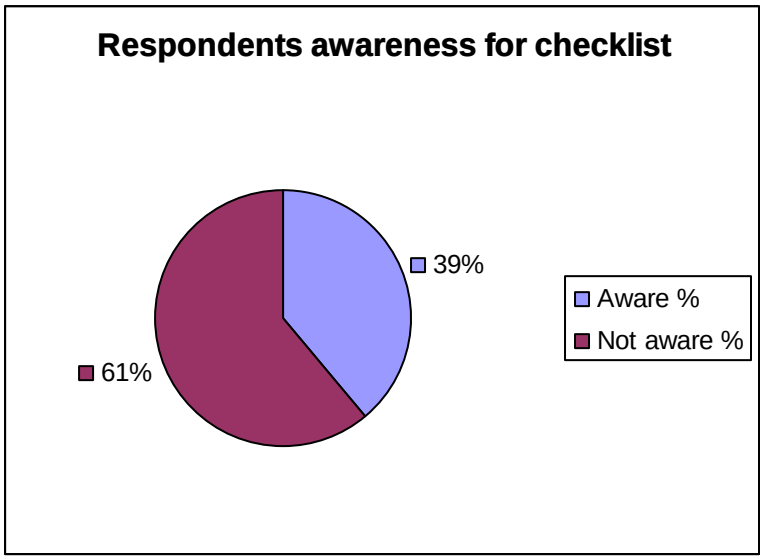
5.7 Awareness of checklist

Question no 8 was dealt whether users are identified of not the checklist for effective search results and it was found that total 34 no of professional i.e. 57% are identified or aware about checklist out of total 60 professionals. Whereas only 25 general users are aware out of 90 and this prefers only 27% general users are identified rest 73% not so, their responses can be shown as follows:

Table 5.5 Respondents awareness for checklist

Name of library	*Aware*				*Not aware*			
	Professional Staffs	General User	Total	%	Professional Staffs	General Users	Total	%
TUCL	10	4	14	31%	4	26	30	69%
ICIMOD	2	2	4	22%	2	12	14	78%
SSBL	2	3	5	45%	1	5	6	55%
KUSOML	1	5	6	54%	2	3	5	46%
OTHERS	19	11	30	31%	17	19	36	69%
Total	34	25	59	39%	26	65	91	61%

Source field survey



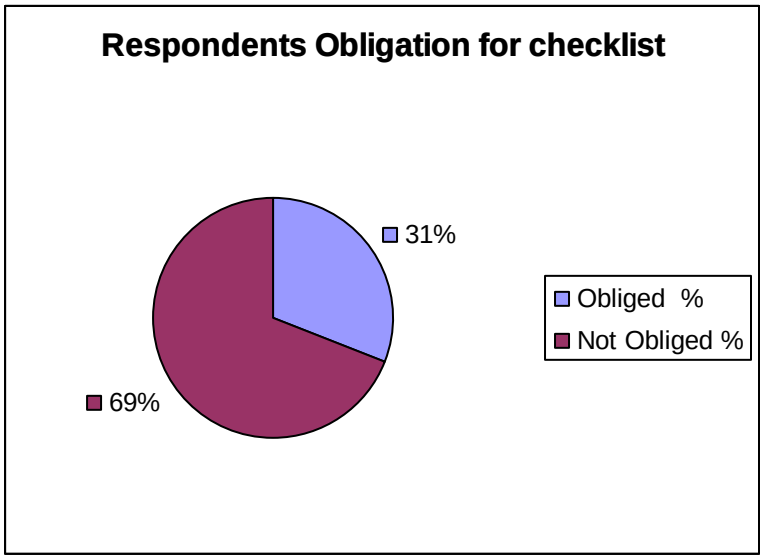
5.8 Respondents Follow-up of checklist in navigating process

Question no 9 was associated with checklist that the follow-up by respondents while searching their required information and it was found that only 31% serious information seekers follow-up it properly while 69% not obliged. It can be shown as follows:

Table 5.6 Respondents obligation for checklist

Name of library	*Follow-up*				*Not Follow-up*			
	Professional Staffs	General User	Total	Obliged %	Professional Staffs	General Users	Total	Not obliged %
TUCL	6	4	10	22%	8	26	34	78%
ICIMOD	2	6	8	44%	2	8	10	56%
SSBL	1	3	4	36%	2	5	7	64%
KUSOML	1	2	3	27%	2	6	8	73%
OTHERS	12	10	22	33%	24	20	44	67%
Total	22	25	47	31%	38	65	103	69%

Source field survey



5.9Necessity of retrieving technique

Question 10 and 11 dealt with the necessity of any kinds of information retrieving technique and their type, the respondents replied that they needed the technique but type of technique used varied according to the users. The professional librarians prefer Boolean retrieving technique while the general users preferred natural language techniques.

Table 5.7 Respondents preference for retrieving model/technique

Name of library	*Professionals*				*General users*			
	Boolean preference	%	N.L.P. Preference	%	Boolean preference	%	N.L.P. preference	%
TUCL	9	64%	5	36%	12	40%	18	60%
ICIMOD	4	100%	-	-	4	29%	10	71%
SSBL	2	66%	1	34%	2	33%	6	73%
KUSOML	3	100%	-	-	3	38%	5	62%
OTHERS/ M Lib Sc.	20	55%	16	45%	22	74%	8	26%
Total	38	63%	22	37%	43	48%	47	52%

Source field survey

5.10 Respondent using the Boolean operator

In question no 12 researcher was interested to know how many numbers of users have been using the Boolean logical operators, and it was found that 63% and 48% of professionals and general users respectively were used it. This question was associated with the research hypothesis that the "maximum no of users use Boolean operators for information retrieval", so this hypothesis was tested and associated tables and results are presented as follows.

Table 5.8 Respondents profile for using the Boolean operators

Name of library	*Professionals*				*General users*				*Total*	
	Yes	%	No	%	Yes	%	No	%	Yes	No
TUCL	9	64%	5	36%	12	40%	18	60%	19	23
ICIMOD	4	100%	-	-	4	29%	10	71%	8	10
SSBL	2	66%	1	34%	2	33%	6	73%	4	7
KUSOML	3	100%	-	-	3	38%	5	62%	6	5
OTHERS/ M Lib Sc.	20	55%	16	45%	22	74%	8	26%	42	24
Total	38	63%	22	37%	43	48%	47	52%	79	69

Hypothesis

H_0 : Max no of user use Boolean operators while navigating in databases/search engines.

H_1 : Limited no of users use Boolean operators while navigating in database/ search engine.

Results:

¡ ^{2} Value at 5% level of significance for 4 degree of freedom is 9.488.

Calculated value of ¡ ^{2} is 5.8

Since calculated value< tabulated value so null hypothesis (H_0) is accepted.

Since the null hypothesis formulated in this case was accepted implying a maximum no of user use Boolean operator in retrieving information, but their efficiency would be measured with the help of question no 15, which is presented ahead.

5.11 Respondent's skills for coordinating the key words

Question no.13, and 14 were formulated to find out whether the respondents had the idea of coordinating keywords or key terms, their skills and familiarities to formulate exact search statements by the combination of AND, OR, NOT and their success in finding the exact information through the keywords using the Boolean logic. A majority of respondents were found to have no idea about the use of Boolean operator and hence their coordination, except some of the professional's librarians.

5.12 Evaluating the skills for successful Boolean query formulation

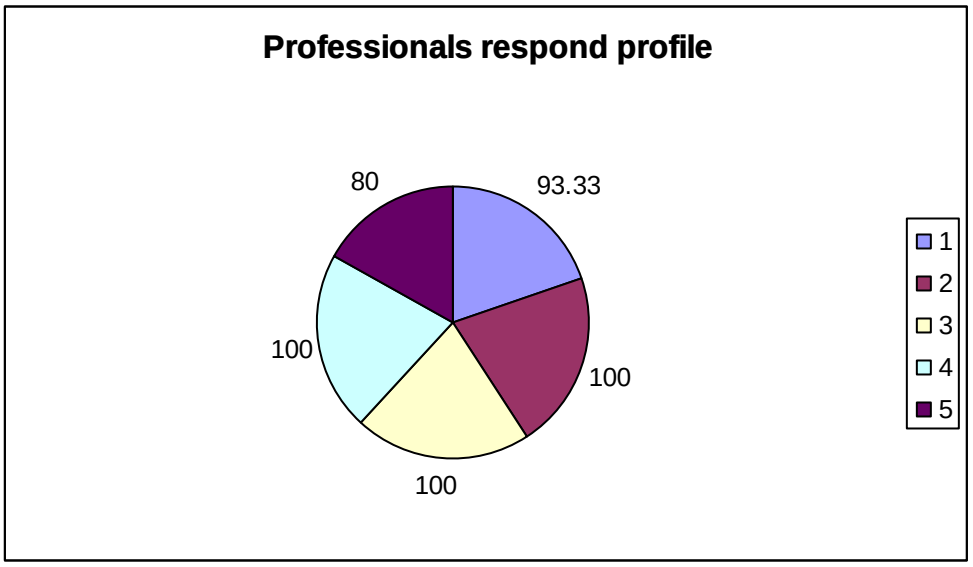
Question no 15is the question made to meet the exact objective of the research purpose. This question is based upon the skills, ability knowledge of individual professional staffs, general users, and M Lib science students in using the Boolean logic and finding the information of necessity. Broadly two major user groups, professional librarians and general users were selected from the TUCL database users. M Lib Science students also categorized in general users group. Here each of the respondent were provided with a set of five test questions or titles, cover all kinds of subject in the request as for as possible and were requested to retrieve the information using the online TUCL database and fill the form with the number of information the individual got. These five titles were set in the assumptions that they cover the wide range of information retrieving problems in the daily life of information retrieval and were selected according to the suggestions provided by thesis instructor and data base expertise. These titles or problems have the implicit title, subject identification, and etc. characteristics. Since the expected number of observation in this case would vary with the individual as different individual had different way of coordinating keywords or key terms. To avoid this and to save the expected theory the no of observation as observed by the TUCL Database specialist, Chief Librarian Mr. Krishna Mani Bhandari, deputy librarian Mr. Chiranjivi Neupane and thesis supervisor as well as faculty member of central department Mr. Bhim Dhoj Shrestha was taken as the expected frequency. Deriving the study it was assumed that there was no significance difference in record of retrieving information by the professional staffs and standard expected set by above two specialists. Using the expected frequency obtained as above, different observed frequencies from individual respondents (professional staffs and general users along with M L Isc students) were collected.

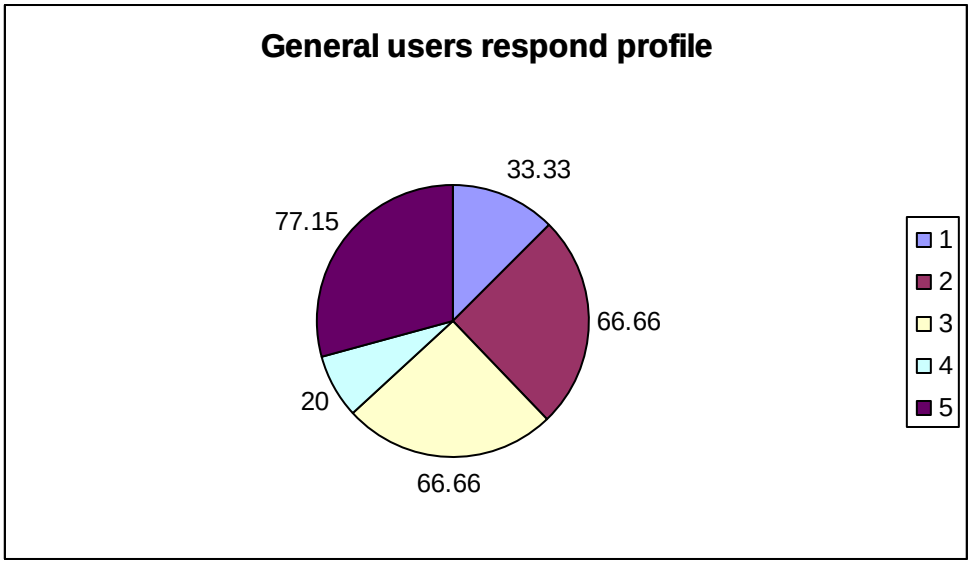
Sample respondents profile for question no 15 is allocated as follows:

Table 5.9 Respondents profile for use of Boolean operators in TUCL database

Name of library	*Professionals*		*General users*		*Percentage*		*Total Respondents*	
	Distributed	Returned	Distributed	Returned	P.S.	G.U.	No.	%
TUCL	15	14	45	15	93.33	33.33	29	100
ICIMOD	4	4	15	10	100	66.66	14	100
SSBL	3	3	15	10	100	66.66	13	100
KUSOML	3	3	15	3	100	20.00	6	100
OTHERS	45	26	35(M lib)	27	80	77.15	53	100
Total	70	50	125	65	94.66	52.00	115	100

Source field survey





For question no 15, only 115 respondents were participated out of 150. The respondents so selected provided $115 \times 5 \times 3 = 1725$ search results. After completion of search the output results as obtained by two broad user groups is presented in following *Table no.11-15* were compared with the result as obtained by database expertise for the relevance assessment. Total value obtained by two (2) broad user groups, professionals and general users to their five (5) titles, corresponding to three (3) Boolean operators i.e. total $2 \times 5 \times 3 = 30$ Chi-Square **(¡ ²)** test was used as the significance of the study.

Table 5.10 Retrieved results of problem no 15 by specialist and thesis instructor

No of problems	Observed values by TUCL database specialist and thesis instructor		
	OR	AND	NOT
a.	403	2	13
b.	793	1	280
c.	3657	111	947
d.	17563	35	441
e.	337	8	227

Source direct observation

Table 5.11 Retrieved results for problem no. (a)

No of respondents	Observed values by 50 professional staffs problem no. (a)		
	OR	AND	NOT
5	398	5	29
8	392	8	20
9	402	9	7
11	405	11	11
17	403	17	13
No of respondents	Observed values by 65 general users for problem no. (a)		
	OR	AND	NOT
5	279	0	29
8	392	4	20
14	174	6	7
17	167	2	11
21	403	8	13

Source field survey

Table 5.12 Retrieved results for problem no. (b)

No of respondents	Observed values by 50 professional staffs problem no. (b)		
	OR	AND	NOT
7	758	12	295
10	807	1	265
13	810	26	272
20	793	121	280
No of respondents	Observed values by 65 general users for problem no. (b)		
	OR	AND	NOT
7	793	12	240
10	807	1	335
18	810	26	280
28	758	121	265

Source field survey

Table 5.13 Retrieved results for problem no. (c)

No of respondents	Observed values by 50 professional staffs problem no. (c)		
	OR	AND	NOT
3	3664	111	941
4	3650	111	893
43	3657	111	947
No of respondents	Observed values by 65 general users for problem no. (c)		
	OR	AND	NOT
3	3664	111	941
4	3650	111	893
58	3657	111	947

Source field survey

Table 5.14 Retrieved results for problem no. (d)

No of respondents	Observed values by 50 professional staffs problem no. (d)		
	OR	AND	NOT
2	441	16	424
7	632	48	426
9	16713	42	445
14	17536	29	454
18	17563	35	441
No of respondents	Observed values by 65 general users for problem no. (d)		
	OR	AND	NOT
2	441	16	424
7	632	48	426
10	16713	42	445
18	17536	29	454
28	17563	35	441

Source field survey

Table 5.15 Retrieved results for problem no. (e)

No of respondents	Observed values by 50 professional staffs problem no. (e)		
	OR	AND	NOT
6	337	8	232
2	337	8	219
42	337	8	227
No of respondents	Observed values by 65 general users for problem no. (e)		
	OR	AND	NOT
6	337	8	177
7	337	8	219
52	337	8	327

Source field survey

Hypothesis

H_0**:** There is no significance difference between the no. of data as retrieved by specialist to that retrieved by professional staffs and general users.

H_1**:** There is significance difference between the no. of data as retrieved by specialist to that retrieved by professional staffs and general users.

Table 5.16 Decision table for problem no. 15

Decision table Problems	¡ ^{2} results for 50 professionals at 5% level of significance			¡ ^{2} results for 30 M. Lib. Sc. students and 35 general users at 5% level of significance		
	OR	AND	NOT	OR	AND	NOT
a.	H_0 Accepted	H_0 Rejected	H_0 Rejected	H_0 Rejected	H_0 Rejected	H_0 Rejected
b.	H_0 Accepted	H_0 Rejected	H_0 Accepted	H_0 Accepted	H_0 Rejected	H_0 Rejected
c.	H_0 Accepted	H_0 Accepted	H_0 Accepted	H_0 Accepted	H_0 Accepted	H_0 Accepted
d.	H_0 Rejected	H_0 Rejected	H_0 Accepted	H_0 Rejected	H_0 Rejected	H_0 Accepted
e.	H_0 Accepted	H_0 Accepted	H_0 Accepted	H_0 Accepted	H_0 Accepted	H_0 Rejected
Accepted H_0	4	2	4	3	2	2
Accepted H_1	1	3	1	2	3	3

Source appendix 3 and 4

From the calculations made as shown in the appendix (3-4). It can be deducted that, the professional staffs were proficient in retrieving information by coordination the key words or key terms using Boolean OR operator and NOT operator. But their proficiency of retrieving information using the AND Boolean operator was moderate, implying less skills in coordination terms using AND Boolean operator.

From the ¡ ^{2} test as applied to the data retrieved by the M. Lib. Sc. Students, general users, and professional staffs of different libraries, it was found that their skill, knowledge, and ability to retrieve information using Boolean logic was unsatisfactory. As from the results, it can be deduced that they were not good enough in coordinating the terms using AND and NOT Boolean operator, but had a moderate coordinating skills using OR operator.

Since, OR operator, retrieves inundated information so the user or information consumer find difficulties to distinguish the precise and prompt information, which also consumes lot of time and money, and is also against the Dr. Ranganathan's fourth law of library science. So the users or navigators should be proficient in obtaining information using AND and NOT operator.

5.13 Ability of distinguishing relevant and non relevant one

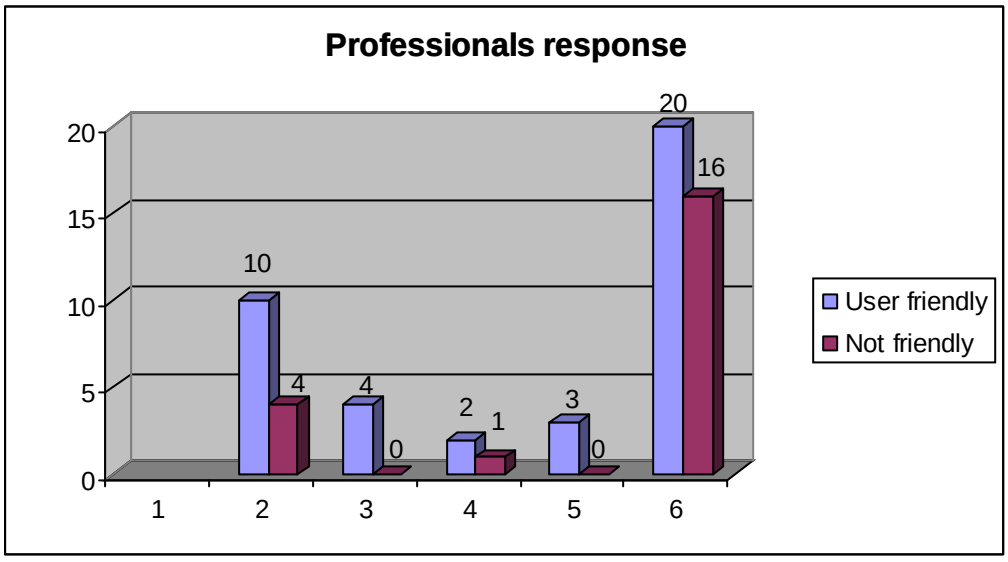
Through question no 16 it was intended to find out if the respondent were able to distinguish relevant and non relevant information by retrieving it via the Boolean operators. Their ability to distinguish relevant and non relevant information depended upon their knowledge and experience such that professional librarians were able to distinguish while the general users had a dilemma.

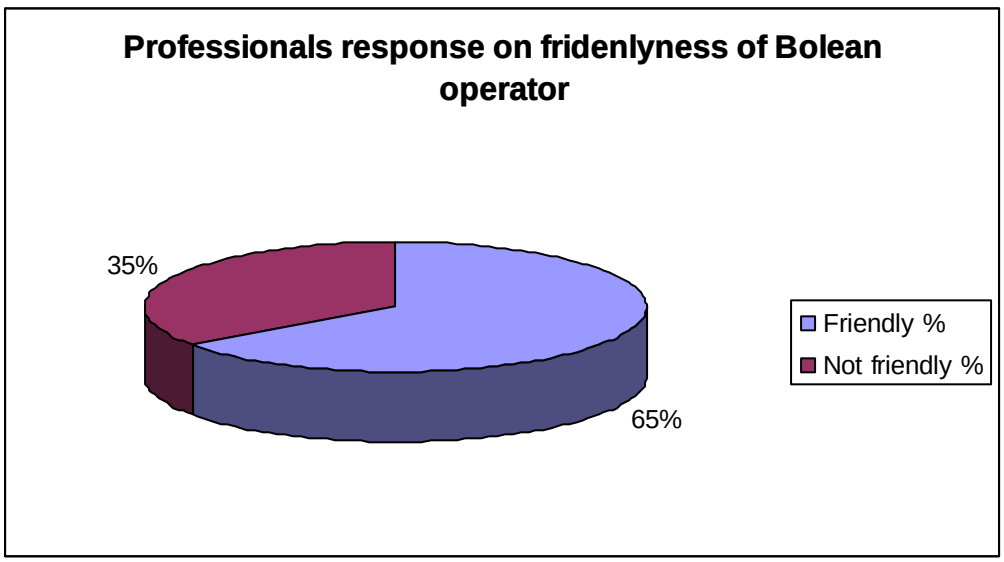
5.14 User friendly of Boolean operator

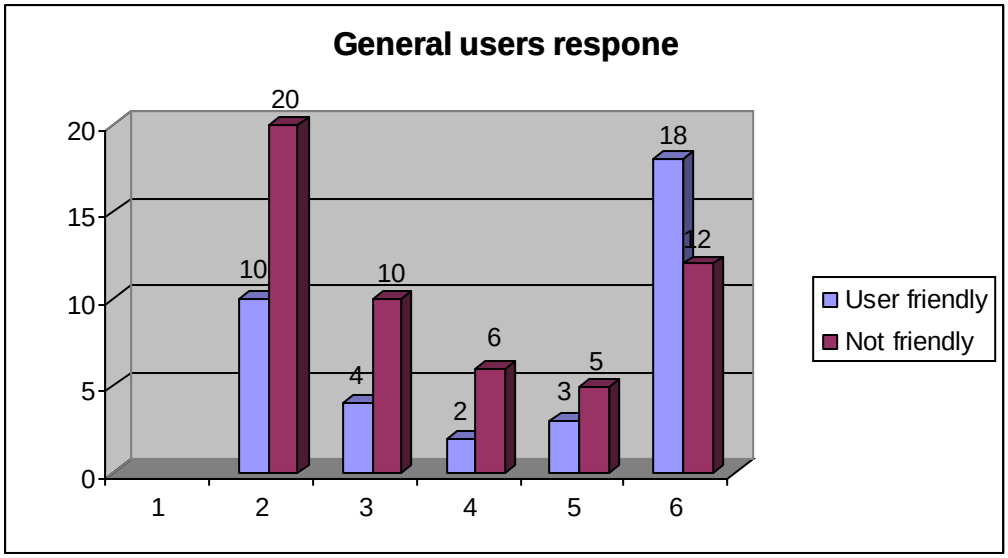
Through question no 18 it wan intended to find out the user friendliness of Boolean operators. It was found that Boolean operators are user friend to some professional librarians and not for general users. This can be understood more clearly from the following table and pie-chart, where 65% professionals feel it user friendly 35% not feel so. Similarly only 41% general users feel it user friendly whereas 59% general user not so.

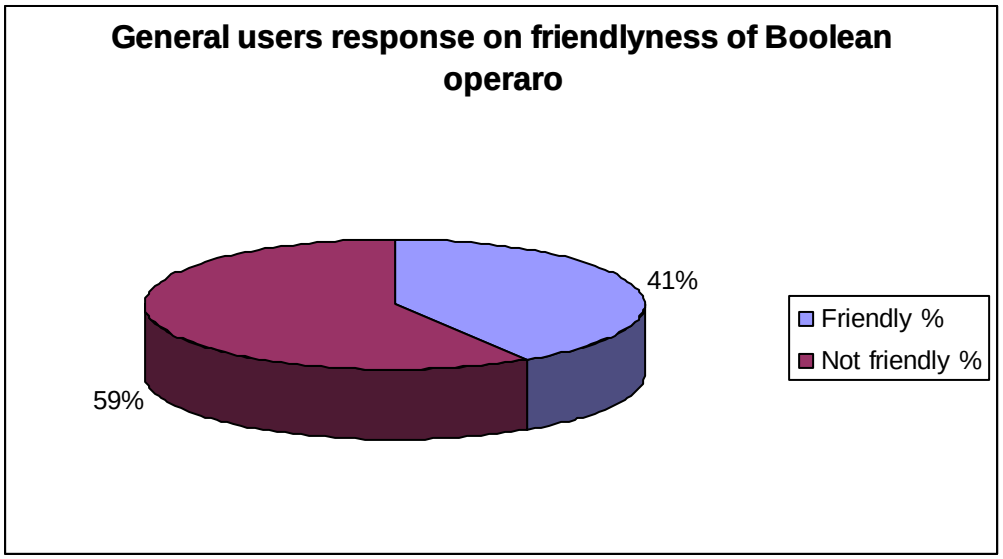
Table 5.17 User friendliness of Boolean operator

Name of library	*Professionals*				*General users*			
	User friendly	%	Not friendly	%	User friendly	%	Not friendly	%
TUCL	10	71%	4	29%	10	33%	20	66%
ICIMOD	4	100%	-	-	4	29%	10	71%
SSBL	2	66%	1	34%	2	33%	6	73%
KUSOML	3	100%	-	-	3	38%	5	62%
OTHERS/ M Lib Sc.	20	55%	16	45%	18	60%	12	40%
Total	39	65%	21	35%	37	41%	53	59%









Source field survey

5.15 Efficiency of Boolean operator

To know the proficiency of keywords or key terms and if Boolean operators were helpful to retrieve the exact information question no 19 and 20 were asked and it was found that Boolean operators were helpful to professional librarians only, not to general users, and most of the respondents (excluding a few professional librarians) replied retrieval of information using more than two key words was more difficult as compared to the use of one or two keywords. Thus it is analyzed that efficiency of Boolean operator is limited up to some of the professional librarians only.

5.16Respondent's problems for retrieving information

Question no 21 was asked to know the problems faced by the respondents while retrieving information. The problem faced were the lack of subject knowledge and insufficient keywords, not proper idea about Boolean operators and its retrieving technique.

5.17Respondent's response to overcome their retrieving problems

Question no 22 and 23 were asked do they require any orientation or/and their approach in solving the problem while they retrieve their information. Among them maximum numbers of respondents preferred for orientation, and their approach to solve the problem it was found that the general users preferred other techniques (such as Common Language Processing.) while the professional were developing more knowledge about the Boolean operators.

Finally, it can be concluded that though maximum of users have increasing inclination towards the keywords, key terms, using AND, OR, NOT. It was also found that maximum no of users could not retrieve the exact information in a short period using this logic.

Chapter VI

SUMMARY CONCLUSION AND RECOMMENDATION

6.1Summary of the findings

In the knowledge based society of today information professionals provide a vital role across all areas of industry. This requires a firm grasp of the theories and principles of management, informational retrieval, and information technology. Effective retrieving technique and effective communication is fundamental to this and must be combined with the skills required in the management of information.

Information professionals provide essential expertise in our knowledge-based society. There is a growing demand for individuals with the specialized skills required to manage, retrieve, and propagate information effectively. Advance the intelligent application of information technologies and the principles, models of management and retrieving various kinds of information is life blood of informationist to provide an effective service. Plan, implement, and evaluate information services to meet the needs of current and future users is also an essential skills of information manager.

Based on the above fact, the information seeking behaviors of our professionals and their effective services can be evaluated with their skills of retrieving precise and prompt information. Libraries and information centers are the service providing agencies; meanwhile professionals associated with these institutions are the service provider, so they must have to be proficient in their respective field. It was considered that knowledge of Boolean operators could help them to cater the right information to right person at right time in precise and prompt way.

On the other hand, huge area is occupied by the users inside or outside the libraries. Some times they navigated their needed information by themselves and some times they get the information through the help of professional staffs. For the former case it is difficult to say whether they are able to use Boolean operator or not to retrieve their needed information in precise and prompt way. Therefore this study was based on these circumstances and hence it can be said that this study is based on information retrieving behaviors of professional staffs and general users using Boolean operators. In other words this study was based to know the proficiency of the professional staffs and general users in retrieving the needed information using Boolean operators especially in library, databases and search engines.

Total 195 questionnaire were distributed among the professional staffs, general users and M. Lib. Science students, out of this only 150 i.e. 76.92% respondents were taken for the study. The four libraries viz TUCL, ICIMOD, SSBL, and KUSOML were taken as the main source for data collection among them TUCL database serves a great effort in thesis study. Different data were collected from the field survey via questionnaire and direct observation. Spearman rank correlation and ¡ ^{2} - Test was used for testing the significance of the research work. Rank correlation was used for question no 2 and 17, and total 25 ¡ ^{2} - Test was used to the questions no 12 and 15 for the significance of the study. This study implies Boolean operators are the most important information retrieving model. Based upon the responses given by professional staffs, general users, and direct observations of database users and database expertise, researcher had applied the different statistical tools, to make the advance analysis and obtain the following.

1. Professionals consulted information firstly available in library database, then books/ abstracts, search engines and journals respectively.
2. General users obtained their required information from search engines, books/abstracts, journals, library database with respective priorities.
3. Both groups (i.e. professional staffs and general users) used both kinds (i.e. manual and online) retrieval technique.
4. Google search engine is the best search engine for their information need.
5. Most of the professionals and users besides using their respective library databases used PEERI, JOSTER etc.
6. All kinds of general users preferred to use term approach (i.e. common language approach) while retrieving the information, where as professional staffs used the keywords for the same purpose.
7. Most of the information seekers need some kinds of retrieving technique, for this professional staffs used Boolean operators where as general users used common language processing technique.
8. Most of i.e. 57% professionals were found that they are aware with the checklist of successful results, where 73% of general users are not aware.

9. Out of 39% aware respondents only 31% serious respondents obliged to checklist for successful results, rest 8% not follow-up it properly while navigating.

10. Majority of general users were found to have no idea about the use of Boolean operators. Among them only 48% respondents used Boolean operator but they were not able to coordinate exact keywords and also failed to make successful query formulation.

11. 63% professional staffs had idea about Boolean operators and they used it while searching or retrieving information form database. But data and their significance test (see. *appendix 4-5* and *table 5.8*) indicated that they were not able to coordinate three and/or more than three keywords and hence this affected their successful query formulation results.

12. Professionals were proficient in retrieving information by coordinating the key words using Boolean OR and NOT operator. But their proficiency of retrieving information using the AND Boolean operator was moderate, implying less skills in coordination terms using AND operator.

13. General users and M. Lib. Sc. Students were not proficient for coordinating keywords via Boolean operators. It was found that their skill, knowledge, and ability to retrieve information using Boolean logic were unsatisfactory.

14. Professional librarians and few general users were able to distinguish relevant and non relevant information while few general users had a dilemma.

15. 65% professionals find keywords and Boolean operators helpful in retrieving precise information, and equally 65% professionals prefer it is user friendly too; whereas in case of general users only 41% feel it so.

16. Most of the users were found that they had not sufficient subject knowledge to make interactive search process to get their exactly needed information; also they hadn’t the sufficient idea about Boolean operators.

17. Most of the professional staff's recommended that the basic preconditions and requirements are necessary to become a talented expertise.

6.2 Conclusions

The meaning of the term information retrieval can be very broad. Just getting a credit card out of our wallet so that we can type in the card number is a form of information retrieval. However, as an academic field of study, information retrieval might be defined thus:

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers). As defined in this way, information retrieval used to be an activity that only a few people get engaged in: reference librarians, paralegals, and similar professional searchers. Now the world has changed, and hundreds of millions of people engage in information retrieval every day when they use a web search engine or search their email. Information retrieval is fast becoming the dominant form of information access, overtaking traditional database-style searching

IR can also cover other kinds of data and information problems beyond that specified in the core definition above. Information retrieval systems can also be distinguished by the scale at which they operate, and it is useful to distinguish three prominent scales. In web search, the system has to provide search over billions of documents stored on millions of computers. Distinctive issues need to gather documents for indexing, being able to build systems that work efficiently at this enormous scale, and handling particular aspects of the web, such as the exploitation of hypertext and not being fooled by site providers manipulating page content in an attempt to boost their search engine rankings, given the commercial importance of the web.

Developing a good search strategy requires knowledge about the nature and organization of target databases and also the exact need of information seekers or the user. Subject knowledge and their keyword as well as various thesaurus terms can make interactive search formulation and re-formulation until their desired free text query. Use of Boolean operators could become a landmark in information retrieval system and make a high precision rate for the same, if they have the successful keywords permutation knowledge. Skills required to access the remotely located database with interrogative techniques is the current requirement of the navigators.

6.3 Recommendations

In study of *Boolean operator in logical and efficient information retrieval* in the scenario of library database and their users various types of observations are obtained and all are already presented above. In practice, however, few search topics can be adequately expressed by a single word or short phrase, and Boolean logic is used as a means of combining brief search terms in order to put a more complex or detailed search expression to the database or to the Internet.

The Internet is a vast computer database. As such, its contents must be searched according to the rules of computer database searching. Much database searching is based on the principles of Boolean logic. Few search engines nowadays offer the option to do full Boolean searching with the use of the Boolean logical operators. It is more common for them to offer simpler methods of constructing search statements, specifically implied Boolean logic and template language. So some recommendations may be useful for the implementation in information retrieving strategies among the users and rewarding for beyond study to other researcher. From this experience point of view in the libraries mentioned as above the following are this recommendations to the professionals, general users, M. Lib. Science students, and who use or want to use Boolean logic in information retrieving purpose.

1. All library professionals should be able to retrieve complex information via Boolean logical operators.
2. To retrieve precise and prompt information all users should use computerized or online based navigation for their frequent requirement.
3. To be proficient in retrieving information short term training and demonstrated orientation should be provided to all professionals and general users respectively.
4. For complex statements they should be able to distinguish appropriate keywords/ key terms and should be able to coordinate these terms accurately.
5. Though user's who are not identified with checklist for successful query formulation, they must have aware about it and though who are identified with checklist they must have to follow-up while navigating their requirement.

6. The OR connector is often overlooked by novice users of computer database retrieval systems, yet it can be essential for successful retrieval because the words and phrases used to describe the same subject in different documents can vary enormously. Thus the searcher should anticipate common variants to each search term and join them with the OR operator BEFORE using the AND operator.

7. AND operator is main to obtain the precise and prompt information and it also increases the precision and recall rate, so all the users should be able to make successful query formulation via AND operator.

8. Boolean logic is not intuitive and is not so easy for most people to use it correctly so they should be able to understand logical statements. For e.g. *"*query *find the names of the Oxford* AND *Cambridge colleges"* and its equivalent Boolean search expression "Oxford OR Cambridge".

9. Similarly Boolean logic is unable to accurately represent some queries, leading to imprecise retrieval, so users must have to be clear on their subject matter. For e.g. *"Teaching AND Nepali* AND *Schools"* can retrieve items concerning *'teaching Nepali in school'* as well as *'teaching in Nepali school'*, and more…

10. If a search statement contains two or more different Boolean operators it almost certainly needs brackets to ensure that the logic is unambiguous. So the users should be able to make successful query statements.

11. To satisfy a query, search terms simply have to be present anywhere in the matching database records according to the specified Boolean relationship. So the users must be able to permute the keywords until their desired results.

12. Each library should maintain their own library database with Boolean search facility along with orient classes to their users is time to time.

13. The Boolean retrieval model contrasts with *ranked retrieval models* such as the vector space model, in which users largely use *free text queries* , that is, just typing one or more words rather than using a precise language with operators for building up query expressions, and the system decides which documents best satisfy the query. So the libraries should incorporate these

system based technologies or databases. This avoids the limitation and rigidness of Boolean operators.

14. A Boolean retrieval model does not have a built-in way of ranking matched documents by some notion of relevance: e.g. doesn't say that a document that satisfies N+1 clause in the query is more relevant than a document that satisfies N clauses. So the system should be integrated with ranking and other advanced strategy based technologies. This helps to increase the degree of matching the precise and prompt need.

REFERENCES

Aryal, Bishnu Prasad (2005). *Subject Heading, its Application for Consistent and Uniform Subject Entries for Literature Published in Nepali Language*. Unpublished master's thesis, Central Department of Library and Information Science, Tribhuvan University, Kirtipur, Katmandu, Nepal.

Aula, Anne (2000). *Query formulation in web information search.* Retrieval July 4, 2008, from University of Tampere, Department of Computer science Web site: http://www.cs.uta.fi/~aula/questionnaire.pdf

Baeza-yates,Richardo and Ribeiro-netom, Berthier. (2005). *Modern information retrieval.* London: Longman.

Bakewell, K.G.B. (1996). Indexing. In *International Encyclopedia of Information and Library Science* (pp.169-170). London: Rutledge.

Belkin, N. J. and Croft, W.B. (1987). Retrieval Technique. *Annual Review of Information Science and Technology, 22,* 109-145.

Belkin, N.J. (2000). Helping people find what they don't know. Communications of the ACM. (pp. 58-61).

Blair D.C. (1986). *Indeterminacy in the subject access to documents, information processing and management*. (pp. 229-234).

Buxton, Andrew and Hopkinson, Alan (2001). *The CDS/ISIS for Windows handbook*. New Delhi : UNESCO.

Chowdhury, G.G. (1990). *Information Retrieval System*. Calcutta: IASLIC.

Christopher, D. M., Prabhakar Raghavan and Hinrich Schütze (2008). *Introduction to Information Retrieval*. Cambridge University.

Cleverdon, C. (Ed.). (1988) *Optimizing convenient online access to bibliographic databases. In : p. willett, ed. Document retrieval system*, London: Taylor Graham.

Ellis, David. (1996). *Progress and problems in information retrieval*. London: Library association publishing.

Ellis, David. (Eds.). (1996). *Progress and problems in information retrieval*. London:

Frants, V.I., Shapiro,J., Taksa,I. and Voiskunskii, V.G. (1999). Boolean search: Current state and perspectives. *Journal of the American Society of information science, 50*(1), 86-95.

Goyal, D.P. (2000*). Management information system: managerial Perspectives*. Delhi: Macmillan.

Gull,C.D. (1956). Seven Years of Work on the Organization of Materials in the Special Library. *American Documentation,7*, 320-329.

Gupta, S.K. (1996). *Fundamental of statistics.* (4th ed.). New Delhi: Himalayan.

Harter, S.P.(1988). *Statistical Approaches to Automatic Indexing: Document Retrieval System* (pp.81-94). London: Taylor Graham.

Hartley, R.J. (1996). Information Retrieval. In *International Encyclopedia of Information and Library Science.* (pp.211-212). London: Rutledge.

Hornby, A.S. (Ed.). (2008). *Oxford Advanced Learner's Dictionary*, (7th ed.). Oxford: Oxford University.

Jansen, B.J. and Pooch, U., (2001). Web user studies: A review and framework for future work. *Journal of the American Society for Information Science and Technology*, *52*(3), 235-246.

Jansen, B.J., Spink, A., and Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing and Management*, *36*(2), 207-227.

Jin, Xiangyu , French, James, and MIchel, Jonathan (2005). *Query formulation for information retrieval by intelligence analysts.* Retrieved June 24, 2008, from Virginia University, Department of Computer Science Web site: http://www.cs.virginia.edu-techrep.cs

Kent, Allen (Eds.). (2004). Encyclopedia of library and information science (Vol. 64, pp. 218-246). NewYork: Marcel dekker.

Kirszner, L. G. and Mandell, S. R. (2002). *The Holt Handbook* (6th ed.). Fort Worth: Harcourt College.

Kowalski, Gerald J. and Mayburg, Mark T. (2000. *Information storage and retrieval system: theory and implementation*, (2nd ed.). Boston: Kbuwer academic publishers.

Girja Kumar (1994). *Theory of Cataloging.* New Delhi: Vikas.

Lancaster, F.W. (1979). *Information retrieval system: Characteristics, testing, and evaluation* (2nd ed.). New York: John Wiley and Sons.

Lancaster, F.W. (1986). *Vocabulary Control for Information Retrieval* (2nd ed.). Arlington: Information Resources.

Lancaster, F.W. (1991). *Indexing and Abstracting in Theory and Practice*. (pp.18-42). London: Library Association.

Lancaster, F.W. (1979). *Information retrieval system: characteristics, testing, and evaluation*. (2nd ed.). New York: John Wiley and Sons.

Lenburg, Jeff (2007). *Guide to research.* New Delhi: Viva books.

Library association publishing.

Miller, Joseph and Goodsell, Joan (2004). *Sears List of Subject Headings* (18th ed.). New York: The H.W. Wilson.

Nayaichyai, Lila (2006). *Manual v/s computerized cataloging.* Unpublished master's thesis, Central Department of Library and Information Science, Tribhuvan University, Kirtipur, Katmandu, Nepal.

Parmer, P.P., and Bhuta, B. (Eds.). (1989). *Encyclopaedic dictionary of library and information science* (vol. 1, pp. 1-62). New Delhi: Anmol.

Pradhan, M. R. (2004). Developing digital libraries: technologies and challenges. *Library Herald, 42*(2), 105-108.

Prasher, R.G. (1991). *Information and its Communication.* New Delhi: Medallion.

Ranganathan, S.R. (1964). *Classified Catalogue Code: With Additional Rules for Dictionary Catalogue Code.* (pp.44-46). Bangalore: Sarada Ranganathan Endowment for Library Science.

Ranganathan, S.R.(1974). *Cataloguing Practice* (2nd ed.). (pp.511). Bombay: Asia.

Salton,G., Singhal, A. and Allan, J. (1996). Automatic text documentation and structuring. *Information processing and management ,32*(2), 127-38.

Smith, Eliot R. and Mackie, Diane (1995). *Social psychology.* New York: M. Worth.

Spink, A. et al. (2000). Searching the web: The public and their queries, *Journal of the American Society for Information Science and Technology, 52*(3), 226–234.

Tillett, B.B. (2004). *Library of Congress Subject Heading* (27th ed., vol.). Washington DC: Library of Congress.

Tseng, Gwyneth. (1996). Boolean Logic. *International Encyclopedia of Information and Library Science*. London: Routledge.

Turtle, Howard (1994). *Natural language vs. Boolean query evaluation: A comparison of retrieval performance.* Otago: ACM Press.

Vakkari, P. (2000). Cognition and changes of search terms and tactics during task performance: A longitudinal case study. *Proceedings of the RIAO 2000 Conference. Paris, France*, pp. 894-907.

APPENDICES

Appendix: 1 Rank correlation used for question no 2 and 15.

Appendix – 1A

Professional Staffs (R1)	General Users (R2)	d= R1-R2	d2
4	3	1	1
2	2	0	0
1	4	-3	9
3	1	2	4
Total			$\sum d^2 = 14$

So calculation is obtained as follows:

$$\circ = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)}$$

$$\circ = 1 - \frac{6 \times 14}{4\left(4^2 - 1\right)}$$

$$\circ = 1 - \frac{84}{60}$$

$$\circ = -0.4 \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

Appendix – 1 B

Professional Staffs (R1)	General Users (R2)	d= R1-R2	d2
1	1	0	0
3	2	1	1
2	4	-2	4
4	3	1	1
Total			$\sum d^2 = 6$

$$\circ = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)}$$

$$\circ = 1 - \frac{6 \times 6}{4\left(4^2 - 1\right)}$$

$° = 1 - \frac{36}{60}$

$° = 0.4$ …….. (ii)

Appendix: 2 ¡ ^{2} *- Test for use of Boolean operators.*

Appendix- 2 A

Name of library	*Use of Boolean operators*		*Row marginal*
	Yes	No	
TUCL	21	23	44
ICIMOD	8	10	18
SSBL	4	7	11
KUSOML	6	5	11
OTHERS/ M Lib Sc.	42	24	66
Column marginal	81	69	Total N=150

Note: Value drawn from *Table no.5.8*

Calculation of expected frequency E= $\frac{RT \times CT}{GrandTotal(N)}$

Calculation	*Name of library*	*Observed value (o)*	*Expected Value (E)*	*O-E*	$\frac{(O-E)^2}{E}$
YES	TUCL	21	23.76	-2.27	0.32
	ICIMOD	8	9.72	-1.72	0.30
	SSBL	4	5.94	-1.94	0.63
	KUSOML	6	5.94	1.94	0.63
	OTHERS/ M Lib Sc.	42	35.64	1.13	1.13
NO	TUCL	23	20.24	2.76	0.37
	ICIMOD	10	8.28	1.72	0.35
	SSBL	7	5.06	1.94	0.74
	KUSOML	5	5.06	-0.06	7.11×10^{-4}
	OTHERS/ M Lib Sc.	24	30.36	-6.36	1.33
Total					$\sum\left[\frac{(O-E)2}{E}\right]$ =5.8

Degree of freedom (d.f.) = (r-1) (c-1)

d.f. = (5-1) (2-1)

$= 4\times1$

$= 4$

¡ ^{2} Value at 5% level of significance for 4 degree of freedom is 9.488.

Calculated value of ¡ ^{2} is 5.8

Since calculated value< tabulated value so null hypothesis (H_0) is accepted.

Appendix: 3 ¡ ^{2} - Test of 50 Professional staffs for question no 13.

Appendix – 3 A

Tabulated value of ¡ ^{2} for 0.05 at 4 d.f. is 9.488

¡ ^{2} - *Test of OR for problem (a)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
5	398	400	-2	4	0.01
8	392	400	-8	64	0.16
9	402	400	2	4	0.01
11	405	400	5	25	0.062
17	403	400	3	9	0.02
Total=50					$\sum\left[\frac{(O-E)^2}{E}\right]$=0.262

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

¡ ^{2} - *Test of AND for problem (a)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
5	0	4	-4	16	4
8	4	4	0	0	0
9	6	4	2	4	1
11	2	4	-2	4	1
17	8	4	4	16	4
Total=50					$\sum\left[\frac{(O-E)^2}{E}\right]$=10

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

¡ ^{2} - *Test of NOT for problem (a)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
5	29	16	13	169	10.56
8	20	16	4	16	1
9	7	16	-9	81	5.06
11	11	16	-3	9	0.56
17	13	16	-5	25	1.56
Total=50					$\sum\left[\frac{(O-E)^2}{E}\right]$=18.74

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

Appendix – 3 B

Tabulated value of ¡ ^{2} for 0.05 at 3 d.f. is 9.488

¡ ^{2} - *Test of OR for problem (b)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
7	758	792	-34	1156	1.45
10	807	792	15	225	0.28
13	810	792	12	144	0.18
20	793	792	1	1	0.0012
Total=50	3138				$\sum\left[\frac{(O-E)^2}{E}\right]$=1.97

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

¡ ^{2} - *Test of AND for problem (b)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
7	12	40	-28	784	19.6
10	1	40	-39	1521	38.05
13	26	40	-14	196	4.9
20	121	40	81	6561	164.02
Total=50	160				$\sum\left[\frac{(O-E)^2}{E}\right]$=226.52

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

¡ ^{2} - *Test of NOT for problem (b)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
7	295	278	17	289	1.03
10	265	278	-13	169	0.60
13	272	278	-6	36	0.13
20	280	278	2	4	0.01
Total=50	1112				$\sum\left[\frac{(O-E)^2}{E}\right]$=1.77

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

Tabulated value of ¡ ^{2} at 5% l.s. for 2 d.f. is 5.991

Appendix – 3 C

Tabulated value of ¡ ^{2} for 0.05 at 2 d.f. is 5.991

¡ ^{2} - *Test of OR for problem (c)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
3	3664	3657	7	49	0.013
4	3650	3657	-7	49	0.013
43	3657	3657	0	00	0000
Total=50	10971				$\sum\left[\frac{(O-E)^2}{E}\right]$=0.026

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

¡ ^{2} - *Test of AND for problem (c)*

For problem *ECONOMICS* AND *DEVELOPMENT* all 50 professionals obtained (111) the same results as obtained by specialist. So this problem doesn't need test via ¡ ^{2} tool.

¡ ^{2} - *Test of NOT for problem (c)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
3	941	927	14	196	0.21
4	893	927	-34	1156	1.24
43	947	927	20	400	0.43
Total=50	2781				$\sum\left[\frac{(O-E)^2}{E}\right]$=1.88

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

¡ ^{2} - *Test of OR and AND for problem (e)*

For OR, AND operator all the general users obtained the same results as prescribed or obtained by the specialist, i.e. 337 and 8 respectively. So applying ¡ ^{2} - Test was no logic.

¡ ^{2} - *Test of NOT for problem (e)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
6	232	226	6	36	0.15
2	219	226	-7	49	0.21
42	227	226	1	1	0.004
Total=50	678				$\sum\left[\frac{(O-E)^2}{E}\right]$ = 0.36

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

Appendix – 3 D

Tabulated value of ¡ ^{2} for 0.05 at 4 d.f. is 9.488

¡ ^{2} - *Test of OR for problem (d)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
2	441	10577	-10136	102738496	9713.38
7	632	10577	-9945	98903025	9350.76
9	16713	10577	6136	37650496	3559.65
14	17536	10577	6959	48427681	4578.58
18	17563	10577	6986	48804196	4614.18
Total=50					$\sum\left[\frac{(O-E)^2}{E}\right]$ =27695.55

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

¡ ^{2} - *Test of AND for problem (d)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
2	16	34	-18	324	9.52
7	48	34	14	196	5.76
9	42	34	8	64	1.88
14	29	34	-5	25	0.73
18	35	34	1	1	0.02
Total=50					$\sum\left[\frac{(O-E)^2}{E}\right]$ =17.91

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

¡ ^{2} - *Test of NOT for problem (d)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
2	424	438	-14	196	0.44
7	426	438	-12	144	0.32
9	445	438	7	49	0.11
14	454	438	16	256	0.58
18	441	438	3	9	0.02
Total=50					$\sum\left[\frac{(O-E)^2}{E}\right]$ =1.47

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

Appendix: 4 ¡ ^{2} - Test of 65 General users for question no13.

Appendix – 4 A

Tabulated value of ¡ ^{2} for 0.05 at 4 d.f. is 9.488

¡ ^{2} - *Test of OR for problem (a)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
5	279	283	-4	16	0.056
8	392	283	109	11881	41.98
14	174	283	-109	11881	41.98
17	167	283	-116	13456	47.54
21	403	283	120	14400	50.88
Total=65					$\sum\left[\frac{(O-E)^2}{E}\right]$=182.43

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

¡ ^{2} - *Test of AND for problem (a)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
5	0	4	-4	16	4
8	4	4	0	0	0
14	6	4	2	4	1
17	2	4	-2	4	1
21	8	4	4	16	4
Total=65					$\sum\left[\frac{(O-E)^2}{E}\right]$=10

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

¡ ^{2} - *Test of NOT for problem (a)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
5	29	16	13	169	10.56
8	20	16	4	16	1
14	7	16	-9	81	5.06
17	11	16	-3	9	0.56
21	13	16	-5	25	1.56
Total=65					$\sum\left[\frac{(O-E)^2}{E}\right]$=18.74

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

Appendix – 4 B

Tabulated value of ¡ ^{2} for 0.05 at 3 d.f. is 9.488

¡ ^{2} - *Test of OR for problem (b)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
7	793	792	1	1	0.06
10	807	792	15	225	0.28
18	810	792	12	144	0.18
28	758	792	-34	1156	1.45
Total=65	3138				$\sum\left[\frac{(O-E)^2}{E}\right]$=1.97

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

¡ ^{2} - *Test of AND for problem (b)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
7	12	40	-28	784	19.6
10	1	40	-39	1521	38.05
18	26	40	-14	196	4.9
28	121	40	81	6561	164.02
Total=65	160				$\sum\left[\frac{(O-E)^2}{E}\right]$=226.52

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

¡ ^{2} - *Test of NOT for problem (b)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
7	240	280	-40	1600	5.71
10	335	280	55	3025	10.8
18	280	280	0	0	000
28	265	280	-15	225	0.80
Total=65					$\sum\left[\frac{(O-E)^2}{E}\right]$=17.31

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

Appendix – 4 C

Tabulated value of ¡ ^{2} for 0.05 at 2 d.f.. is 5.991

¡ ^{2} - *Test of OR for problem (c)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
3	3664	3657	7	49	0.013
4	3650	3657	-7	49	0.013
58	3657	3657	0	00	0000
Total=65	10971				$\sum\left[\frac{(O-E)^2}{E}\right]$ =0.026

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

¡ ^{2} - *Test of AND for problem (c)*

For problem *ECONOMICS* AND *DEVELOPMENT* all 65 students obtained (111) the same results as obtained by specialist. So this problem needs not to be test via ¡ ^{2} tool.

¡ ^{2} - *Test of NOT for problem (c)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
3	941	927	14	196	0.21
4	893	927	-34	1156	1.24
58	947	927	20	400	0.43
Total=65	2781				$\sum\left[\frac{(O-E)^2}{E}\right]$ =1.88

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

¡ ^{2} - *Test of OR and AND for problem (e)*

For OR, AND operator all the general users obtained the same results as prescribed or obtained by the specialist, i.e. 337 and 8 respectively. So applying ¡ ^{2} - Test was no logic.

¡ ^{2} - *Test of NOT for problem (e)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
6	177	241	-64	4096	16.99
7	219	241	-22	484	2.00
52	327	241	86	7396	30.6
Total=65	732				$\sum\left[\frac{(O-E)^2}{E}\right]$ = 49.67

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

Appendix – 4 D

Tabulated value of ¡ ^{2} for 0.05 at 4 d.f. is 9.488

¡ ^{2} - *Test of OR for problem (d)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
2	441	10577	-10136	102738496	9713.38
7	632	10577	-9945	98903025	9350.76
10	16713	10577	6136	37650496	3559.65
18	17536	10577	6959	48427681	4578.58
28	17563	10577	6986	48804196	4614.18
Total=65					$\sum\left[\frac{(O-E)^2}{E}\right]$=27695.55

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

¡ ^{2} - *Test of AND for problem (d)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
2	16	34	-18	324	9.52
7	48	34	14	196	5.76
10	42	34	8	64	1.88
18	29	34	-5	25	0.73
28	35	34	1	1	0.02
Total=65					$\sum\left[\frac{(O-E)^2}{E}\right]$=17.91

Since calculated value>tabulated value. So H_0 rejected and H_1 accepted.

¡ ^{2} - *Test of NOT for problem (d)*

Respondents	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
2	424	438	-14	196	0.44
7	426	438	-12	144	0.32
10	445	438	7	49	0.11
18	454	438	16	256	0.58
28	441	438	3	9	0.02
Total=65					$\sum\left[\frac{(O-E)^2}{E}\right]$=1.47

Since calculated value<tabulated value. So H_0 accepted and H_1 rejected.

Appendix: 5 ICIMOD Library Database Search

ICIMOD Library Database Search

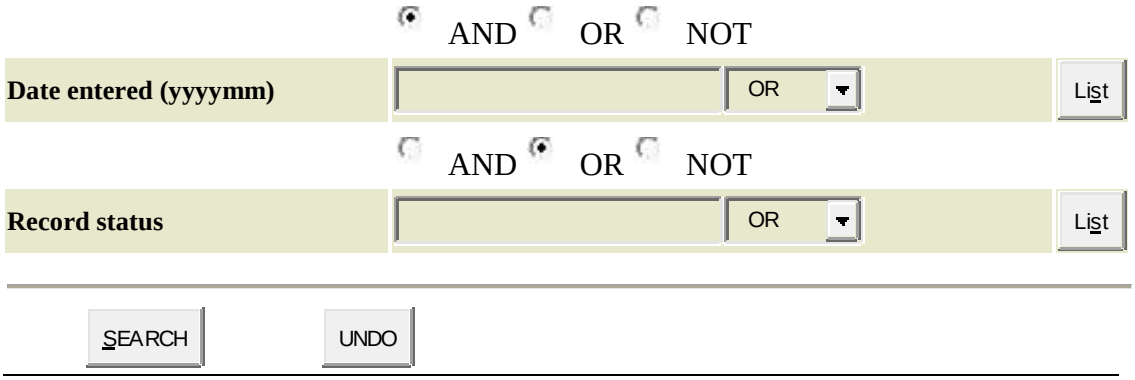
SEARCH UNDO

Exact match of words: ☐

Sort: by Author

Format: Short format

Keywords from record	OR	List
	◉ AND ○ OR ○ NOT	
Keywords from Title	OR	List
	○ AND ◉ OR ○ NOT	
Keywords from Serial Title	OR	List
	○ AND ○ OR ◉ NOT	
Personal Author	OR	List
	◉ AND ○ OR ○ NOT	
Keywords from Corporate Author	OR	List
	○ AND ◉ OR ○ NOT	
Keywords from Conference	OR	List
	○ AND ○ OR ◉ NOT	
Descriptors	OR	List
	◉ AND ○ OR ○ NOT	
Language	OR	List
	○ AND ◉ OR ○ NOT	
TRN	OR	List
	○ AND ○ OR ◉ NOT	
Subcentre code	OR	List



Search Tips

- Enter search term(s) or click on **List** button to select terms from the dictionary.
- Press **SEARCH** button to execute the query.
- When entering more than one search terms in a field, separate the terms with a semicolon (;), which will be interpreted by the system according to the choice of boolean operator.
  You may also type your ISIS logical operators:
  + for OR
  ^ for AND NOT
  * for AND
  In this case, the operators in the boolean selection box are not active and they do not affect operators specified in the query.
- Use **UNDO** to delete last input and **CLEAR** to empty the form.
- Use **SAVE FORM** button to save your query as a bookmark and to reuse it in the future.
- If the box "Exact match of words" is marked, you can also use the $ mask for root searching.

See also examples

QUESTIONNAIRE

Central Department of Library and Information Science

Date: …. / …. / ….

Dear Sir, Madam,

Boolean operator: an approach in information retrieval system

Your assistance is required in providing a quick response to the attached questionnaire, which has been designed to gather pertinent information of the above mentioned topic. I am currently conducting research on the above topic during the course of this 2nd year's masters program as the partial fulfillment for the master's degree in respective discipline.

In order to arrive at realistic and pragmatic solutions if any I require frank and honest feedback from those who are actually involved in the management and operation of information system and information seekers within the library or internet world. It is for this reason that you have been approached to respond to the attached questionnaire. Your response would be invaluable in reflection the actual invade state of Boolean operator in information retrieving system.

For purpose of clarification the concept of Boolean operator as it is applied to retrieve information from the libraries or internet world is concerned with the information seeking behaviors of the various kinds of professionals or users.

Due to the limited time set for the study your cooperation in responding quickly to the questionnaire would be most appreciated. Thank you for your time in filling out the questionnaire.

Full name: …… Designation: ……...

Name of Institution/Library……...

Qualification: …….

Please choose any one answer:

1. You are the part of this library as:

   a. Professional staff b. Member c. Others

2. What type of information do you need frequently? [Please rank the along options]
   a. Based on journals [ ]
   b. Based on books/ abstracts [ ]
   c. Based on library database [ ]
   d. Based on search engine [ ]
3. Which retrieval technique do you use to search your information?
   a. Manual b. Online c. Both
4. For manual searching what kind of sources of techniques do you follow?
   a. Library catalog
   b. Library index
   c. Some other kinds of sources
5. For online searching which search engine do you follow frequently?
   a. Google search engine
   b. Yahoo search engine
   c. Others (please specify) …….
6. For database searching which database do you search frequently?
   a. TUCL
   b. ICIMOD
   c. SSBL
   d. Others (please specify)……..
7. Among search engines what kind of approach do you follow?
   a. Keywords approach
   b. Term approach
   c. Others (please specify) ……
8. Do you identified/informed with checklist for effective search results?
   This comprises:
   i. Choosing the right search site/database;
   ii. Choosing keywords/ indexed term carefully;
   iii. Narrowing the search via OR, AND,and NOT operators;
   iv. Checking spelling and permutation of key terms;
   v. Consulting the help screen in case of search engine or consulting the manual in case of manual database;
   vi. Using more than one subject guide or search engine for conformation of information.

   a. Yes b. No
9. Do you follow such checklist while you navigate your required information?
   a. Yes b. No
10. Do you feel necessity of some kinds of information retrieving technique?
    a. Yes b. No

11. If so which one retrieving technique would you like to use frequently?
    a. Boolean retrieving technique.
    b. Probabilistic and statistical technique
    c. Natural language processing technique
    d. Any others (please specify) ……
12. Do you use Boolean logical retrieving technique?
    a. Yes b. No
13. Do you have some idea about coordinating the key words/terms? or Are you able to formulate exact search statement by the combination of AND, OR, NOT?
    a. Yes b. No

14. Have you got the exact information through the key words/terms with application of Boolean logic?
    a. Yes b. No
15. Could you search the following problems from the TUCL database www.tucl.org.np with using the Boolean operators? Please fill up the following table with your retrieved results.
    a. Quantum theory of physics
    b. Anthropology and human civilization
    c. Economics and development
    d. Society and culture of Nepal
    e. Organization and leadership

    Yours Results:

Problems	a	b	c	d	e
OR					
AND					
NOT					

16. Are you able to distinguish relevant and non relevant information after retrieving it via the Boolean operators?
    a. Yes b. No c. Confusion

17. What factors affecting for your successful query formulation? [Please rank the along options]
    a Experience of information retrieval . [ ]
    b. Familiarity with database and search engine [ ]
    c. Familiarity with the Boolean operators [ ]
    d. Type of search task [ ]
18. Do you feel Boolean operators are friendly to retrieve exact information?
    a. Yes b. No
19. Could you predict a priory exactly how many items are to be retrieved to satisfy your query with your search terms?
    a. Yes b. No

20. What do you think about which one is better for information retrieval with application of Boolean operators?
    a. Only one keyword.
    b. Two keywords.
    c. More than two key words.
21. What problems do you face in retrieving the information?
    a. Lack of subject knowledge
    b. Insufficient keywords
    c. Others (please describe)
    ……
    ……
    ……
22 Do you need any kinds of orientation/demonstration about Boolean retrieving technique?
    a. Yes b. No

23. How have you overcome your problems of searching and retrieving the information?
    a. Developing knowledge or idea about Boolean operator?
    b. Choosing any other techniques?
    c. Others (please describe)
    ……
    ……
    ……

24. What do you think; the basic preconditions or requirements are necessary to achieve precise and prompt information? (Please describe)

    ……
    ……..……
    ……
    ……
    ……
    …….

Thank you for your time.

Researcher: Ramesh Prajuli
Thesis Year. Central
Department of Library
and Information Sience.
T.U. Kirtipur.

BIO-DATA

Name	Ramesh Prajuli
Address	Kirtipur-18, Katmandu (Temporary) Alamdevi-8, Syangja (Permanent)
Telephone	01-4333321
Cell	9841469695
Email	parajuli256anooj@yahoo.com / rameshprajuli@gmail.com
Nationality	Nepali
Date of Birth	13 August 1983

Education

2005-2007	Tribhuvan University: Masters in Library and Information Science (Distinction)
2004-2005	Tansen Multiple college: Bachelors of Education in Mathematics (Second Division)
2001-2004	Tribhuvan Multiple college: Bachelors of Science in Physics (Second Division)
2000-2001	Board of Secondary Education Madhya Pradesh Bhopal: 10(+2) in Science (First Division)
1999	Board of Secondary Education Madhya Pradesh Bhopal: S.L.C. (First Division)

Employment

2008- present	Librarian, GEMS Institute of Higher Education, Dhapakhel Lalitpur.
Skills	Computer literate: WIN/ISIS, CDS/ISIS, Data Communication and Networking, and expert in library automation software i.e. LIBRA, MIDAS, and SOUL.
Interest	Social welfare, Tracking & Travel, and Swimming.