

**BOOLEAN OPERATOR
IN LOGICAL AND
EFFICIENT INFORMATION RETRIEVAL**

**A thesis submitted to
the Central Department of Library and Information Science in
Partial Fulfillment of the Requirements for a Master of
Arts Degree in Library and Information Science
Tribhuvan University**

Submitted by
RAMESH PRAJULI

**Central Department of Library and Information Science
Faculty of Humanities and Social Sciences
Tribhuvan University, Kirtipur,
Katmandu, Nepal
November 2008**

LETTER OF RECOMMENDATION

This is to certify that Mr. Ramesh Prajuli has prepared this dissertation entitled "**BOOLEAN OPERATOR IN LOGICAL AND EFFICIENT INFORMATION RETRIEVAL** ", under my supervision and guidance. I recommend this dissertation for final approval and acceptance.

Date: November 2008

.....

Mr. Bhim Dhoj Shrestha

Thesis Supervisor

LETTER OF ACCEPTANCE

The thesis here to attached, entitled "**BOOLEAN OPERATOR IN LOGICAL AND EFFICIENT INFORMATION RETRIEVAL**", prepared by Mr. Ramesh Prajuli in partial fulfillment of the requirements for the Masters of Arts Degree in Library and Information Science is hereby accepted and approved.

.....
Mr. Bhim Dhoj Shrestha

Thesis Supervisor

.....
Mr. Rudra Prashad Dulal

External Examiner

.....
Dr. Madhusudan Karki

Head of Department

DEDICATION

To my late grandmother **Deu Rupa Parajuli**,

Respected teacher late **Lila Dahal**,

and to my parents for all of your love and support over the years,

This one is in your honor.

PREFACE

BOOLEAN OPERATOR IN LOGICAL AND EFFICIENT INFORMATION RETRIEVAL is the system that widely used in the information navigating and retrieving environment. 21st century is the technological computer based century where a bit of timely information plays a vital role for decision making over the informational, technological industry as well as knowledge based society. With the rapid growth of globalize information and related technologies over the last two decades, information retrieval has become a vital system in maintaining and integrating the timely based information. In today's information society, retrieval technique is moving into the mainstream of the information technology (IT) industry and virtually influencing the way we deal with many of our problems.

Since 1847, when three logical operators had been devised by George Boole, it has been successfully implemented in various fields such as in Mathematics, Electronics, Physics, and in Information Retrieval System. Its dynamic implementation is in information retrieval system operated on the basis of Boolean Logical Statements and Boolean algebra. Therefore large database companies or libraries can't avoid it to facilitating the Boolean Query formulation in their respective database. Hence this is the study of Boolean logical information retrieving techniques based on TUCL online database among the professional librarians of Katmandu valley, general users from some selected libraries, and students from CDLIS.

Current technological advancement caters the human's life towards the interrogative system. Over this advancement Boolean based technique is still used successfully to navigate the complex information. As information technology is taking hold in our society, it is essential to study the retrieving techniques and its awareness among the informatics. Therefore **BOOLEAN OPERATOR IN LOGICAL AND EFFICIENT INFORMATION RETRIEVAL** study highlights the skills and abilities of potential and novice users. For this study data has been collected from four main libraries viz: TUCL, ICIMOD, SSBL, and KUSOM their professionals and users have been taken as the source of information. Total 24 questions were formulated to evaluate their skills of coordinating the keywords and their techniques of implementing Boolean logical operators for successful query formulation. Five search tasks associated with TUCL online database has been presented to fill-up in each questionnaire, on the basis of their responses results were compared with the results as obtained by the TUCL database expertise and thesis instructor. Essential statistical tools such as Rank correlation, Chi square t^2 - Test are applied as per questions for the significant statements. Conclusion and recommendations have been drawn on the basis of this analytical and exploratory study.

ACKNOWLEDGEMENT

As every accomplishment requires help, suggestions, and advice, from many people of different walks, the presented thesis is also an output of combined effort. I would like to take this opportunity to place in record the feelings of gratitude and gratefulness to the deserved persons for their kind support help and cooperation.

At first I 'm very grateful to Mr. Bhim Dhoj Shrestha- my supervisor, Mr. Rudra Prasad Dulal- my external examiner, and to the Head of Department Dr. Madhusudan Karki who spent his valuable time in my favor, giving me his worthy suggestions, guidance and support to carry out this work.

I always respect and admire the valuable help and instruction provided by former Head of Department Mrs. Nirmla Shrestha, faculty members Dr. Mohan Raj Pradhan, and Mr. Bishnu Prashad Aryal. Their direct, indirect teaching suggestions, inspirations will always be saluted. My sincere thanks go to all the professional staffs, M. Lib. Science Students and general users of libraries of different libraries.

I am greatly pleased to have guidance, help and cooperation from Chief Librarian Mr. Krishna Mani Bhandari and Deputy Librarian Mr. Chiranjivi Neupane in filling the no. of results of ' search task'. I am thankful to the librarians as well as my college Mr. Jagadish Chandra Aryal, of Social Science Baha Library, Mr. Ram Saran Thapa, Librarian of ICIMOD, Mr.Prem Raj Adhikari, Librarian of the KUSOM.

My special thanks go to my all colleges, my seniors, and my juniors who helped me time to time. My humble gratefulness to my brother Mr. Nabin Parajuli for making a sound environment and also I'm thankful to Mr. Mahesh Khanal, Mr. Manoj Bhandari, and Mr. Damodar Parajuli for their kindly cooperation.

Lastly my whole hearted obligation goes to each and every members of my family, especially my parents whose encouragement , inspections brought me at this position.

I have used my sincere effort to publish original information from the research. I have understood I apologies and take full responsibility if any mistakes have been committed knowing or unknowing in course of my work.

ABSTRACT

Entitled **BOOLEAN OPERATOR IN LOGICAL AND EFFICIENT INFORMATION RETRIEVAL** database searching and Web searching became the most popular activities on the local and remote login, information retrieval (IR), which was performed mainly by professionals, researchers, now is an activity that every user on the Web is engaged in. The growing interest in general user's database searching and Web searching ability naturally leads to the awareness that some kind of guidance is needed. Since IR on the database or Web has many similarities with traditional IR, IR skills of retrieving information have been adopted to teach general users how-to search or about retrieving information from the Web. Two major concerns regarding users' behavior are frequently found: the queries tend to be short, and Boolean operators are insufficiently used in structuring queries.

Drawing on the social informatics view of IR and the semantic Web research, this thesis first argues that short queries have unique values in database searching and Web searching because OR and NOT operators shows greater effectiveness upto two keywords among the both user groups. It also argues that using Boolean operators is an effective strategy in database searching and Web searching. Because Web searching is generally motivated by knowledge discovery purpose, high recall and precision is not as important as it is in small databases, so database searching become more sophisticated with Boolean operators. The study also discusses the tradeoffs between the benefits of operator use with enhanced search interfaces and the time and effort that users need to exert to use the features.

The study mainly focuses the information seeking behavior of library professionals, M. Lib. Sc. Students, general users of university level, and researcher. Their behavior or skill measured with the help of questionnaires on the basis of their personal knowledge and experience. Therefore **BOOLEAN OPERATOR IN LOGICAL AND EFFICIENT INFORMATION RETRIEVAL** study highlights the skills and abilities of potential and novice users. For this study data has been collected from four main libraries viz: TUCL, ICIMOD, SSBL, and KUSOM their professionals and users have taken as the source of information. Total 24 questions were formulated to evaluating their skills of coordinating

the keywords and their techniques of implementing Boolean logical operators for successful query formulation. Total 63% and 48% of professionals and general users respectively were used Boolean operators while they navigating and it meets its hypothesis. In similar investigation total 65% professionals feel Boolean operators are user friendly and rest 35% not feels so. Similarly only 41% general users feel it user friendly whereas 59% general user not so. Five search tasks associated with TUCL online database has been presented to fill-up in each questionnaire, 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 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 (χ^2) test was used as the significance of the study. Conclusion and recommendations has been drawn on the basis of this analytical and exploratory study.

It finally discusses the carryover of user's information behaviors across different kinds of the digitized library or automated library database or Web, and the effects of the strong commercial forces which spur the need for sophisticated IR skills in shaping user's information habits. It concludes that a different model based on the social cognition theories and practical is needed to approach general user's problems with information retrieval on the database and Web.

Ramesh Prajuli
Central Department of Library and
Information Science, T.U. Kirtipur.

CATALOGUE OF THESIS

Main card

D

025.04

P884_b

Prajuli, Ramesh

Boolean operator in logical and efficient information retrieval/ Ramesh Prajuli . – Kirtipur: Central Department of Library and Information Science, 2008.

xx,103p.:ill.;30cm.

Dissertation: Master degree of Library and Information Science from CDLISc.

Shelf list Card

D

025.04

P884 b

Prajuli, Ramesh

Boolean operator in logical and efficient information retrieval/Ramesh Prajuli . – Kirtipur: Central Department of Library and Information Science, 2008.

12 xx,103p.:ill.;30cm.

Dissertation: Master degree of Library and Information Science from CDLISc.

1. Information retrieval 2. Information science 3.
Boolean operator i. Title

Subject added card

D

025.04 INFORMATION RETRIEVAL

P884 b

Prajuli, Ramesh

Boolean operator in logical and efficient information retrieval/ Ramesh Prajuli . – Kirtipur: Central Department of Library and Information Science, 2008.

xx,103p.:ill.;30cm.

Dissertation: Master degree of Library and Information Science from CDLISc.

Subject added card

D

025.04 INFORMATION SCIENCE

P884 b

Prajuli, Ramesh

Boolean operator in logical and efficient information retrieval/ Ramesh Prajuli . – Kirtipur: Central Department of Library and Information Science, 2008.

xx,103p.:ill.;30cm.

Dissertation: Master degree of Library and Information Science from CDLISc.

Subject added card

D

025.04 BOOLEAN OPERATOR

P884 b

Prajuli, Ramesh

Boolean operator in logical and efficient information retrieval/ Ramesh Prajuli . – Kirtipur: Central Department of Library and Information Science, 2008.

xx,103p.:ill.;30cm.

Dissertation: Master degree of Library and Information Science from CDLISc.

Title added card

D

025.04 Boolean operator in logical and efficient information

P884 b retrieval

Prajuli, Ramesh

Boolean operator in logical and efficient information
retrieval/ Ramesh Prajuli . – Kirtipur: Central Department of
Library and Information Science, 2008.

xx,103p.:ill.;30cm.

Dissertation: Master degree of Library and Information
Science from CDLISc.

TABLE OF CONTENT

TITLE PAGE.....	i
LETTER OF RECOMMENDATION	ii
LETTER OF ACCEPTANCE	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	vi
DEDICATION.....	vii
PREFACE.....	viii
CATALOGUE OF THESIS.....	ix
TABLE OF CONTENT	xii
LIST OF TABLES	xvi
LIST OF FIGURES.....	xviii
LIST OF APPENDICES.....	xix
LIST OF ABBREVIATIONS.....	xx

Chapter-I

INTRODUCTION	1
1.1 Background of the study.....	1
1.1.1 Libraries in the past	1
1.1.2 Information retrieval system	2
1.1.3 Boolean logic	5
1.1.4 Database	5

1.1.5 Search engine	7
1.1.6 Library catalogues.....	9
1.1.7 Indexing	9
1.1.8 Pre-coordinate indexing.....	10
1.1.9 Post coordinate indexing.....	10
1.1.10 Subject heading.....	11
1.2 Statement of problem	12
1.3 Objective of the study	14
1.4 Hypothesis	15
1.5 Scope and limitation of the study	15
1.6 Significance of the study	16
1.7 Definition of the terms	16
1.8 Organization of the study	19

Chapter-II

RIVIEW OF LITERATURE	21
----------------------------	----

Chapter-III

FOCUS OF THE STUDY	30
3.1 Information retrieval system	30
3.2 Information retrieval models	31
3.2.1 Boolean retrieval model	31
3.2.2 Vector space model	39
3.2.3 Statistical and probabilistic retrieval model.....	41

3.4 Natural language processing and information retrieval	43
3.5 Boolean retrieval system facilitated libraries.....	44
3.5.1 Tribhuvan University Central Library (TUCL)....	44
3.5.1.1 TUCL Database with Boolean search facilitation...	44
3.5.2 International Centre for Integrated Mountain Development Library (ICIMODL).....	45
3.5.2.1 ICIMODL Database management.....	45
3.5.3 Social Science Baha Library (SSBL).....	46
3.5.3.1 Social Science Baha Library (SSBL) Database.....	46
3.5.4 Kathmandu University School of Management Library (KUSOML).....	46
3.5.4.1 Kathmandu University School of Management Library (KUSOML) Database.....	47

Chapter-IV

RESEARCH METHODOLOGY	49
4.1 Research design	49
4.2 Population	50
4.3 Sampling method	51
4.4 Data collection methods	51
4.5 Data analysis procedure	52
4.6 Use of Spears man rank correlation	52
4.7 Use of Chi-Square (t^2) test	53

Chapter- V

ANALYSIS AND PRESENTATION OF FINDINGS	56
---	----

5.1 Respondents frequent requirement.....	57
5.2 Factors affecting for successful query formulation.....	58
5.3 Respondents searching habits.....	59
5.4 Preference for search engine.....	60
5.5 Preference for library database.....	60
5.6 Kinds of approach for navigation.....	60
5.7 Awareness of checklist.....	61
5.8 Respondents Follow-up of checklist in navigating process.....	62
5.9 Necessity of retrieving technique.....	63
5.10 Respondent using the Boolean operator.....	63
5.11 Respondent's skills for coordinating the key words	64
5.12 Evaluating the skills for successful Boolean query formulation.	65
5.13 Ability of distinguishing relevant and non relevant one.....	71
5.14 User friendly of Boolean operator.....	71
5.15 Efficiency of Boolean operator.....	73
5.16 Respondent's problems for retrieving information.....	74
5.17 Respondent's response to overcome their retrieving problems....	74

Chapter- VI

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS.....	75
6.1 Summary of the findings	75
6.2 Conclusions	78
6.3 Recommendations	79

REFERENCES.....	82
-----------------	----

LIST OF TABLES

<i>Table: 3.1</i> Example for AND logic	33
<i>Table: 3.2</i> Example for more than two keywords AND logic	34
<i>Table 3.3</i> Example for OR logic	35
<i>Table 3.4</i> Example for more than two keywords OR logic	36
<i>Table 3.5</i> Example for NOT logic	37
<i>Table 3.6</i> Format of truth table	38
<i>Table 3.7</i> Results of truth table for AND, OR, & NOT operators...	38
<i>Table 4</i> Sample respondents profile	52
<i>Table 5.1</i> Respondents profile with their responses	56
<i>Table 5.2</i> Respondents ranked to their sources of information	58
<i>Table 5.3</i> Respondents ranked to factors affecting	59
<i>Table 5.4</i> Respondents profile for approach of information	60
<i>Table 5.5</i> Respondents awareness for checklist.....	61
<i>Table 5.6</i> Respondents obligation for checklist.....	62
<i>Table 5.7</i> Respondents preference for retrieving model/technique.....	63
<i>Table 5.8</i> Respondents profile for using the Boolean operators	64
<i>Table 5.9</i> Respondents profile for use of Boolean operators	66
<i>Table 5.10</i> Retrieved results of problem no 15 by specialist	67
<i>Table 5.11</i> Retrieved results for problem no. (a)	67
<i>Table 5.12</i> Retrieved results for problem no. (b)	68

<i>Table 5.13</i> Retrieved results for problem no. (c)	68
<i>Table 5.14</i> Retrieved results for problem no. (d)	69
<i>Table 5.15</i> Retrieved results for problem no. (e)	69
<i>Table 5.16</i> Decision table for problem no. 15	70
<i>Table 5.17</i> User friendliness of Boolean operator.....	72

LIST OF FIGURES

<i>Fig: 3.1</i> Venn diagram for AND operator.....	33
<i>Fig: 3.2</i> Venn diagram for AND operator for three.....	34
<i>Fig: 3.3</i> Venn diagram for OR operator.....	35
<i>Fig: 3.4</i> Venn diagram for OR operator for.....	36
<i>Fig: 3.5</i> Venn diagram for NOT operator.....	37

LIST OF APPENDICES

Appendix: 1 Rank correlation used for question no 2 and 15.....	85
Appendix: 2 t^2 - Test for use of Boolean operators.....	89
Appendix: 3 t^2 - Test of 50 Professional staffs for question no 13.	90
Appendix: 4 t^2 - Test of 65 General users for question no13.....	92
Appendix: 5 ICIMOD Library Database Search.....	97
Appendix: 6 Questionnaire.....	99

LIST OF ABBREVIATION

ICIMODL: International Centre for Integrated Mountain Development Library

IRS: Information Retrieval System

KUSOML: Katmandu University School of Management Library

MARC: Machine Readable Catalogue

MEDLARS: Medical Literature Analysis and Retrieval System

MeSH: Medical Subject Heading

POPSI: Postulate-based Permuted Subject Indexing

PRECIS: Preserved Context Index System

SSBL: Social Science Baha Library

TUCL: Tribhuvan University Central Library

VRM: Vector Retrieval Model