CHAPTER – 1

INTRODUCTION

1.1 Introduction .............................................................................................. 1
1.2 Open Index Initiative ................................................................................ 3
1.3 Relational Database Management System and Thesaurus ....................... 4
1.4 Statement of Problem ............................................................................... 5
1.5 Need for the Study ................................................................................... 6
1.6 Objectives of the Study ........................................................................... 6
1.7 Scope and Limitation of the Study .......................................................... 7
1.8 Methodology ............................................................................................ 8
1.9 Organization of the Study ....................................................................... 9

References.................................................................................................. 12
1.1 Introduction

The modern age is known as information society and modern economy is known as knowledge economy as information plays a vital role in everyone's life. As a result of scientific and social research, lots of new literature is emanating all over the world. The research, as an information, which is published in various formats is getting scattered into specific knowledge domains. Organization of such information for easy identification and retrieval poses a great challenge and also is highly time consuming. Therefore, for comprehensive search and effective retrieval by the end user, the information sources need to be arranged in some logical sequence or need to be assigned specific notations or tags or keywords. According to Soergel (1974), the process of assigning notations or tags or keywords to the information or information resources is known as indexing.

The indexing of information resources with subject specific keywords can be done in two ways, using controlled vocabulary (Thesaurus) and uncontrolled vocabulary. The retrieval of information becomes meaningful and appropriate, if it is done in a specific time frame and therefore, use of controlled vocabulary (Thesaurus) for indexing information resource becomes pre-eminent and is a paramount prerequisite.

A thesaurus is a list of words pertaining to a specific subject discipline. It is an annotated and an annotated list of descriptive words and phrases. According to Soergel (1974) thesaurus is a conceptual structure of the terminology which contains
structured system of concepts with indication of hierarchical and associative relationships between the concepts. Thesaurus is one of the most important knowledge organization tools for organizing and retrieving information. It is also known as an indexing tool with cross-concordance, which means it is a list of words and phrases used in specific subject domain with cross references.

Research and knowledge are expansive and keep evolving. These characteristics are part of social science research too. Due to continuous expansion and development in the area of social sciences, many new disciplines have come up with the various subjects of social sciences. In view of this, there is a great need for the development of a subject thesaurus. Considering the importance of social sciences, few thesauri such as ERIC, GESIS, HASSET, OECD, UNESCO, etc., have already been developed. However, it is observed that the above mentioned thesauri might not be comprehensive for indexing Indian social science literature as these thesauri lack country specific (indigenous) keywords for effective retrieval. For the above mentioned reason a need has been felt to develop a suitable thesaurus for indexing research materials emanating from India. Therefore, development of such country specific thesaurus will help libraries and information centers to effectively index social science literature published in Indian social science journals, and other primary sources such as working papers, conference papers, theses and dissertations. It is observed that there are two Indian institutions have hosted index to Indian social science literature from their respective websites. ISID (Institute for
Studies in Industrial Development, Delhi) seems to be a pioneer in this effort and has covered literature emanating from India such as journal articles, press clippings, etc., However, the ISID online databases (2016), does not use any standard thesaurus for their online indexing service. Therefore, the search and retrieval is based on the title of the document and subject specific retrieval happens to be a limitation. The next attempt has been made by IGIDR Library, Mumbai (Indira Gandhi Institute of Development Research) and the service has been named as Open Index Initiative (OII). It is also observed that IGIDR Library has developed its own thesaurus, but it’s not as exhaustive as it should be and this thesaurus has not been integrated with the data input worksheet of OII.

1.2 Open Index Initiative (OII)

Open Index Initiative (OII) is an attempt to develop an exhaustive bibliographic database in the public domain for social science literature emanating from India. It was developed as a theoretical model for a doctoral thesis and is now a working model managed by IGIDR library staff at Mumbai, India. It intends to index a) Journal articles from selected Indian social science journals and b) Working papers emanating from some of the Indian social science and management institutes, with abstract and appropriate subject keywords. It has also hosted an online union catalogue of journals available at selected Indian social science and management institutes. The main objective of this initiative was to promote resource sharing and to give wide exposure to Indian social science literature. An important feature of the OII is, when articles are retrieved from its search engine, it gives hyperlink to
the union catalogue of journals so that the end users can identify/locate the availability of a particular journal in their city or nearby cities.

1.3 **Relational Database Management System and Thesaurus**

In simple words, a database management system which is based on relational model is called Relational Database Management System (RDBMS). A relational model represents data in the form of tables. Table in an RDBMS is a two dimensional array containing rows and columns. Each row contains data related to an entity and each column contains the data related to a single attribute of the entity. To delineate RDBMS it is necessary to provide some final details about OII back end. OII has been developed on *MySQL*, an open source relational database management system with PHP as programming language and HTML for the front-end. The openness of OII lies in its open access policy and the use of open source software for organizing information.

Use of a standard thesaurus is vital for effective and appropriate retrieval of information from a database. In its model, OII attempted to integrate a thesaurus for social sciences developed by GESIS (German Social Science Infrastructure Services) with the data input worksheet of OII. According to Manjunath & Sangam (2004) this thesaurus developed on RDBMS was provided to the IGIDR library in MS-Excel files, which were subsequently converted to hash (#) separated text files and were uploaded to corresponding MySQL tables. Even though thesaurus was successfully integrated with OII data input worksheet on an experimental basis, according to the developer of the OII model, a thesaurus was not used for indexing purpose as it did not completely suit for indexing Indian literature.
1.4 Statement of Problem

Open Index Initiative (OII) is an upcoming gateway and an attempt to provide a single search platform for most of the social science research literature emanating from India. OII experimented integration of the GESIS thesaurus with its data input worksheet to derive subject keywords. However, it is observed that many keywords required for indexing Indian social science literature were not available in GESIS thesaurus. Therefore, development of an exhaustive thesaurus for indexing social science literature is very much essential. It has also been observed that, keywords which represent various concepts in the Indian social science literature are not considered and covered in the foreign thesauri.

Following are the some of the important keywords, to exemplify few, are not found in some of the existing thesauri:

Indo-Pak Partition, Scheduled Caste and Scheduled Tribes, Bhoodan Movement, Dalits, Emergency, Devadasis, Monsoon, Madrasas, Panchayats, Ayurveda, Unani, Finance Commission, Five Year Plans, prominent national political parties, such as Congress, BJP, Tribals such as Jarva Tribes, Toda Tribes, Soliga Tribes, etc., etc.,

Certain keywords which are irrelevant to Indian social science literature available in some of the existing thesauri are:

Federal Administrative Courts, Federal Assembly, Federal Border Guard (used as a Border Security Force (BSF) in Indian context)

Some of the thesauri use proper nouns i.e. personal and geographical names which make thesaurus very bulky. However, provision to use names of cities, districts, states, countries etc., may be used as standard subdivision to reduce bulkiness of the thesaurus. Therefore, use of standard subdivisions, with instruction in the
introductory notes to the thesaurus would give extra freedom and flexibility to the indexer and increases mnemonic quality. Inclusion of professionals such as teachers, librarians, physicians, economists, etc., in the standard subdivision would also give lots of flexibility to the indexers.

1.5 Need for the Study

The most important objective of developing thesaurus is to provide an exhaustive list of terms (descriptors and non-descriptors) enabling indexers to use appropriate terms and thereby enabling end users to retrieve appropriate and specific information from the database. Such a development need to study vast literature, keywords used in the existing thesaurus and to identify most preferred term by the end users. Even though the thesaurus developed by the OII coordinator became a de-facto base material, the present research work has considered appropriate keywords from various thesauri such as ERIC, HASSET, OECD, UNESCO, etc.

1.6 Objectives of the Study

The objectives of the present study are to:

- Develop indigenous/country specific thesaurus for social science literature emanating from India;
- Develop an online thesaurus and host it on the web for the benefit of the larger user community; and
- Integrate the thesaurus with OII data input worksheet with proper relations between descriptors and non-descriptors through a search interface.
1.7 Scope and Limitation of the Study

Social science itself is a very vast area. Even though the scope is to develop a thesaurus for social sciences, only certain subject disciplines such as sociology, social anthropology, economics, public administration, education, trade & commerce, social problems & services, political sciences, demography, statistics, customs etc., have been considered under the scope of the study. The subject discipline law is even though a part of social sciences, itself is a very vast subject area and therefore, keywords for the subject discipline law has not been considered in the present research. However, certain important keywords such as Civil Law, Criminal Law, Corporate Law, etc., have been incorporated in the thesaurus.

According to the directory developed by Tyagi & Johry (2001), majority of the social science library and information centers are using DDC system (Dewey Decimal Classification) for classification of the documents. Therefore, various subject disciplines under social sciences have been identified as enunciated in DDC.

To derive and develop this thesaurus, various primary and secondary sources such as selected Indian social science journals, Sociological Abstract, Social Science Encyclopaedia, UNESCO thesaurus, GESIS thesaurus, OECD Macrothesaurus, etc., have been used as base materials for the current research work.

In this study, 70 regular and well known Indian social science journals (Annexure - I) published between 1996 and 2015 were considered for identifying appropriate keywords. In case of such journals, an attempt has been made to cover journals starting from time of their inception, from first volume to the latest volume.
published till 2015. However, as an exception, one of the oldest and most widely read journal *Economic and Political Weekly*, which covers almost all the areas of social sciences, has been considered for the study purpose since 1966.

Any thesaurus may have two sequences - 1. Alphabetic sequence and 2. Classified sequence. However, the thesaurus constructed under this research is available only in alphabetic sequence and thus, this happens to be a limitation. However, it is also observed that there is a great scope for further research and to develop a classified thesaurus based on the present research work.

1.8 Methodology

To construct the thesaurus, the present research work has used ‘Tematres’ an open source software. The above mentioned open source software uses relational database management system for managing all semantic relations. The present research has adopted exploratory and descriptive methods for constructing the thesaurus. All the subject keywords collected from the above mentioned sources have been compiled using ‘Tematres’. This open source software has been developed using PHP, a web scripting language and ‘MySQL’ backend database. The above mentioned open source software is installed and configured on OII server. All the keywords collected from the various information sources are entered into the software with all semantic relations. The software has created various tables in MySQL database which includes user administration for operating software, keywords
management, export-import and backup of database. The software, on executing the web based installation script, would generate ten tables in the database. Out of these ten tables, there are five tables, namely *m_tema*, *m_tabla_rel*, *m_values*, *m_indice*, *m_notas*, on which entire thesaurus resides. The table *m_tema* stores all the descriptors and non-descriptors. The table *m_tabla_rel* provides relations to all the descriptors and non-descriptors. The third table *m_values* contains the information on the relation type such as USE, UF (Used For), BT (Broad Term), NT (Narrow Term), RT (Related Term) etc., The fourth table *m_indice* correlates all the descriptors and non-descriptors and provides relation between keywords and also provides link to the descriptors and non-descriptors. The fifth and the last table, *m_notas* contains scope notes which are associated with various indigenous keywords. A PHP programme has been developed to integrate developed thesaurus with data input worksheet of OII. The entire programme has been provided in one of the chapters of this thesis.

1.9 Organization of the Study

The entire research work, including the thesaurus, the value additions to OII and all the recommendations have been presented in eight chapters.

Chapter - 1 Introduction

The first chapter starts with introduction, which consists of sub-chapters such as introduction to OII, relational database management system, statement of problem,
need for the study, scope and limitations of the study, objectives of the study and methodology. The process of entire research work has been described in this chapter.

Chapter – 2 Thesaurus Constructions and its Role in Indexing

The second chapter elaborates the steps which need to be followed in constructing a thesaurus. This chapter also highlights the importance of thesaurus and its role in indexing as well as identifying the relevant and appropriate documents.

Chapter - 3 Literature Review

Any research will not become unique until and unless past research done on a specific research topic is reviewed. This chapter conducts elucidated literature review on the research topic. The review of literature has emphasized on various aspects of thesaurus construction, automated thesaurus, use of SKOS in generation of automated thesaurus as well as the use of relational database with special reference to social sciences. The critical analysis of the literature unveils that, UNESCO, GESIS and OECD have developed a thesaurus on social sciences. Out of these three thesauri, only GESIS thesaurus has been constructed on relational database management system. Moreover, GESIS has used proper nouns as descriptors due to which it has become bulky. It has also been observed that the coverage in GESIS is more country specific where it has been developed. Once again it is observed that all these three thesauri are not suitable to index Indian social science literature. Therefore, an exhaustive thesaurus to organize knowledge emanating from Indian social sciences is very much necessary.
Chapter – 4 Social Science Research in India: Some Indicators

The thesaurus will become more exhaustive, exclusive and everlasting in the light of background information about social science research in India. The social science is a field of erudition that contemplates society. Generally, a social science is referred as science other than natural science. The chapter describes analysis of social science research in India to get abreast understanding of the major developments in social sciences from various social science institutions across the country.

The present research work for its bibliometric analysis of the research work in India, between 1996 and 2014, used time series data released by SCImago Journal & Country Rank portal. The SCImago Journal & Country Rank is a portal that provides visual representation of indicators derived from the Scopus database. The indicators facilitate analysis of specific subject domain or country with a statistical data captured from 1996 to 2014.

Chapter -5 Construction of Social Science Thesaurus with a Slant towards Indian Social Science Literature

The fifth chapter has dealt with thesaurus construction for Indian social science literature on relational database management system. Total 4684 terms have been derived from the various sources of information which are mentioned in the scope and limitation of the study. The entire thesaurus developed as a core research work is available in this chapter with all semantic relations in an alphabetic sequence.
For the end users’ convenience this chapter also gives user guide with appropriate instructions.

**Chapter – 6 Integration of Social Science Thesaurus with Open Index Initiative (OII)**

One of the important aspects of the current research is to develop an interface which can communicate with thesaurus database and fetch results in specific order with all semantic relations. This chapter lists entire PHP programme used for integrating thesaurus with OII data input worksheet, with ease. This chapter also demonstrates process of assigning keywords to a specific document at the time of indexing and adding such records to OII database.

**Chapter – 7 Contributions to Open Index Initiative (OII)**

The seventh chapter lists three specialized PHP programmes written for OII to generate output in standard bibliographic exchange format i.e. MARC21. The first programme enable OII administrator to convert entire backend database structure from horizontal to vertical. Another two programmes allow OII administrator to export and import OII data into MARC21 format for data exchange.

**Chapter – 8 Findings, Suggestions and Conclusion**

The eighth and final chapter presents summary of entire research work, a few suggestions for further research based on current research and finally the conclusion.

**References**


