IASLIC Bulletin 45 (2), 2000, p. 73-81

An Artificial Intelligence approach towards Automatic Classification (Part 1)

Dr. P. K. Panigrahi
Reader & Head, Dept. of Library and Information Science
University of Kalyani, Kalyani, Nadia, West Bengal, India
e.mail: pkpanigrahi@hotmail.com (or) pkpanigrahi@library.iitkgp.ernet.in

Emphasizes on the need of Artificial Intelligence application in library and information work and services specially in the area of library classification. Describes the nature of analytico synthetic classification and its compatibility with the computer. Defines the Artificial Intelligence with its short history and the scope of its application. States clearly the objectives of the present study to develop an automatic classification system considering few hypotheses. Also describes the methodology could be adopted for this purpose. Gives an overview of the developed system for automatic classification.

1 Introduction

The advent of computers, ushered a new era in many disciplines. Like arithmetic, application of computers has become an essential prerequisite. Information Technology, a combination of Computer Science and the Communication Science has a significant impact on the library profession, paving the way for Internet and library networks through out the World. In addition, Computers as information processing systems have much relevance and offer many opportunities to the library profession.

Ever since the introduction of computers, it has become a goal to many researchers to automate not only chores of routine work, but also the intellectual work. The question is whether we can incorporate our knowledge to make machines perform jobs, which otherwise requires human intelligence and discretion. After all, the first industrial revolution attempted to build machines that could substitute human physical power and the Artificial Intelligence (AI) revolution is attempting to build machines that can mimic human intellectual power. Whether this goal will be fully realised is a matter of controversy. Though, AI researchers believe that one day, it will be achieved, a few questions regarding its possibility still remain unanswered. Leaving the controversy apart, it can be safely concluded that the intermediate results in the process of these attempts could be of great value, as falling a little short of this kind of Herculean task is no mean achievement. The emergence of the field Artificial Intelligence and the progress of research in this field prompted many researchers in other disciplines to apply the tools and techniques of AI especially that of Expert Systems to develop intelligent systems in their respective fields and Library and Information

Science is no exception. Although such enthusiasm is still in its infancy as is the case with AI field itself, it has posed many challenges in many areas of Library and Information Science, like reference service[1], Information retrievel especially subject indexing[2], classification, natural language front-ends to databases to mention a few. The present work is an attempt in this direction in the area of classification.

Although, Classification is applicable to many areas of library science like indexing, thesaurus construction etc., its major application is in the preparation of Classification Schedules and building the class numbers as well. In this regard, library classification is meant for arrangement of documents in order to bring like documents together, following the APUPA pattern[3] (Alien, Penumbral, Umbral, Penumbral and Alien). The purpose of arranging documents in this pattern is to facilitate effective browsing to users, who may be interested not only the documents of umbral region but also of penumbral region.[4]

Library classification has come a long way from early days of enumerative systems. It is Ranganathan, who bestowed an elaborate scientific theory to library classification and introduced Analytico-Synthetic Classification system and devised Colon Classification. He has done a careful study of the 'Universe of Knowledge', the various ways of the formation of new subjects. His studies have obviously to do a lot with basic ideas and their relations[5,6]. Typically, ideas have a kind of CoSSCo (Co-ordinate, Superordinate, Subordinate, Collateral) relationship among them. The parallels can be easily seen between the study of 'Universe of Knowledge' and that of 'Knowledge based Computer Systems' (KBCS) otherwise known as expert systems in the area of Artificial Intelligence. Although both fields have grown independently, they have a lot to offer to each other. On basis of this conviction, the present study has taken into consideration.

11 Library Classification

Information explosion has posed many challenges to the library and information profession. One such challenge is the classification of documents for effective storage and retrieval. The subjectwise arrangement is the classificatory arrangement. Classification is responsible for developing a classification schedule where the mutual relationship among subjects is established. Among all members (subjects) of the universe of subjects there exists a powerful CoSSCo (Coordinate, Superordinate, Subordinate and Collateral) relationship[7]. This relationship is also called as hierarchical relationship. This proves that the nature of the universe of subject is non-linear or multi-dimensional. But arranging documents in the shelves must be linear in nature as users feel convenient to move in a line while browsing. The library classification forces to change the non-linear nature to linear nature when documents are displayed on the shelves. In other words, the library classification brings related subject groups together and unrelated ones separated from the others[8,9].

However, as far as the classification schemes are concerned, there exists different types like enumerative scheme, almost enumerative scheme, almost faceted scheme, rigidly faceted scheme, freely faceted or analytico-synthetic classification scheme, etc. Of these approaches to classification, the present study deals with the last one (i.e. Analytico–Synthetic Classification scheme.)

12 Analytico-Synthetic Classification

"The term 'Analytico Synthetic Classification' is a generic term to denote any scheme in which a compound subject is first analysed into its facets and synthesized in the verbal and notational plane"[10]. The analytico synthetic classification scheme includes two phases of work in three planes. Firstly, it analyses the subject into its facets in the idea plane. Secondly, it synthesizes in the verbal plane and notational plane. The process of classifying a document is guided by some principles and postulates. *'Freely faceted classification'* does not suggest any rigid facet formula but guides the classifier to use postulates and principles while constructing a classification number. However, there has been attempts to make Colon Classification more and more analytico-synthetic. This tendency is increased gradually in different editions. The Colon Classification 7th edition (CC7) is much more analytico-synthetic in nature than the earlier editions[11].

2 Artificial Intelligence

21 Definition and Scope of AI

Artificial Intelligence means the intelligence that is gathered artificially. But defining artificial intelligence is a hard task because, 'intelligence' is a vague word and is not well-defined. Nevertheless, we can consider artificial intelligence as "getting computers to do things these seem to be intelligent". Artificial Intelligence is the study of mental facilities through the use of computational models.

Artificial Intelligence includes[12]

1. Getting the computer to communicate with us in human languages like English. (Natural Language Processing.)
2. Getting computers to remember complicated interrelated facts, and drawing conclusions from them (Inference).
3. Getting computers to plan sequences of actions to accomplish goals (Planning).
4. Getting computers to offer advice based on complicated rules for various situations (Expert Systems).
5. Getting computers to look through cameras and see what is there (Vision).
6. Getting computers to move themselves and objects around in the real world (Robotics).

This is well summarised in Patterson's definition which states that "AI is a branch of computer science concerned with the study and creation of computer systems that exhibit some form of intelligence : Systems which learn new concepts and tasks, systems that can reason and draw useful conclusions about the world around us, systems that can understand a natural language or perceive and comprehend a visual scene, and systems that perform other types of feats that require human types of intelligence"[13]. In other words, a better understanding of AI largely depends on the understanding of related terms like Intelligence, Knowledge, Cognition, Reasoning thought, learning and a number of computer related terms.

To have a clear idea of AI, we must know what AI is not. AI is not the study and creation of conventional computer systems. Even though one can argue that all programs exhibit some degree of intelligence, an AI program will go beyond this in demonstrating a high level of intelligence. AI is not the study of the mind nor the body, nor of psychology, physiology, cognitive science or linguistics. In fact, there is some

overlap between these fields and AI. All seek a better understanding of human intelligence and sensing processes. But in AI, the goal is to develop working computer systems that are truly capable of performing tasks that require high levels of intelligence. The programs are not necessarily meant to initiate human senses and thought processes. Indeed, in performing some tasks differently, they may actually exceed human abilities. The most important point is that the systems should be capable of performing intelligent tasks effectively and efficiently[14,15].

22 History and Development of AI

AI began to emerge as a separate field of study during the 1940s and 1950s when the computer became a commercial reality. It is generally accepted that Alan Turing is the father of Artificial Intelligence (AI). Turing worked with precursors of modern computers. In 1936, Turing visualized in his paper 'Computing machinery and intelligence' that computers could be programmed to exhibit intelligent behaviour. His most important contribution undoubtedly is the 'Turing Test'. In this test, an operator sits at a computer and formulates queries to an unknown source. If the machine provides answers that do not give any clue to the operator whether he is getting the answer from the machine or a human being then, that machine qualifies as having human-like abilities reasoning[16].

The emergence of a new field 'cybernetics' (coined by Norbert Wiener) brought together many parallels between human beings and machines. Cybernetics, is the study of communication between man and machine and it combines the concepts from information theory, feedback control systems (both biological and machines) and electronic computers.

The developments in the field of linguistics, especially the contributions of Noam Chomsky in formal grammars to a large extent helped the developments in natural language processing.

However, many consider 1956 as a landmark in the history of AI. In this year, the so called *'DARTMOUTH CONFERENCE'* was organized by John Mc Carthy and Morvin Minsky. It is in this conference, Mc Carthy coined the word 'Artificial Intelligence'.

In 1963, A L Samuel wrote a checkers playing program which not only played the game with opponents but also used its experience to improve its performance. In the same year, Newel developed 'Logic Theorist' which attempted to prove mathematical theorems. It solved some of the problems in the first chapter of Russell and Whitehead's Principia. Newel along with J C Shaw and H A Simon developed yet another program—General Problem Solver (GPS), which they applied to several tasks including symbolic manipulation of logical expressions.

In 1965, J A Robinson introduced resolution as an inference method in logic. Also work on DENDRAL began at Stanford University by J Lederberg, Edward Feigenbaum and Carl Djerassi. Dendral is an expert system which interprets molecular structures given the information about the constituents of the compound and mass spectra data. Dendral was the first knowledge based expert system[17].

23 Advantages of AI

The advantages of a system that has human intelligence are many. Such machines can be used wherever, human intelligence is required, like in business, engineering, manufacturing,

mining etc. In addition, such study has intrinsic advantages as it demands understanding of the human thinking process, so as to develop computational models[18]. The following reasons given by Winston (1984), justify the study of AI as means of understanding human intelligence[19].

(a) The use of computers demands a clear statement of problem and clear strategy for solutions and this requires a clear thinking about human thinking process.

(b) Computer models force precision. Implementing a theory uncovers conceptual mistakes and oversights that ordinarily escape even the most meticulous researchers.

(c) Computer implementations qualify task requirements. Once a program performs a task, upper bound statements can be made about how much information processing the task requires.

(d) It is usually simple to deprive a computer program of some piece of knowledge in order to test how important that information really is.

3 Interdisciplinary Approach of the study

Numerous attempts have been made to build automatic classification systems using cluster analysis. However, the present study attempts to build such a system using AI techniques. Basically, the approach is highly interdisciplinary encompassing Natural Language Processing and Expert Systems using Logic programming. It should also be emphasized that the attempt is confined to analytico-synthetic classificaiton scheme viz Colon Classification, 7th edition. The following diagram helps to get the idea about involvment of different subjects.

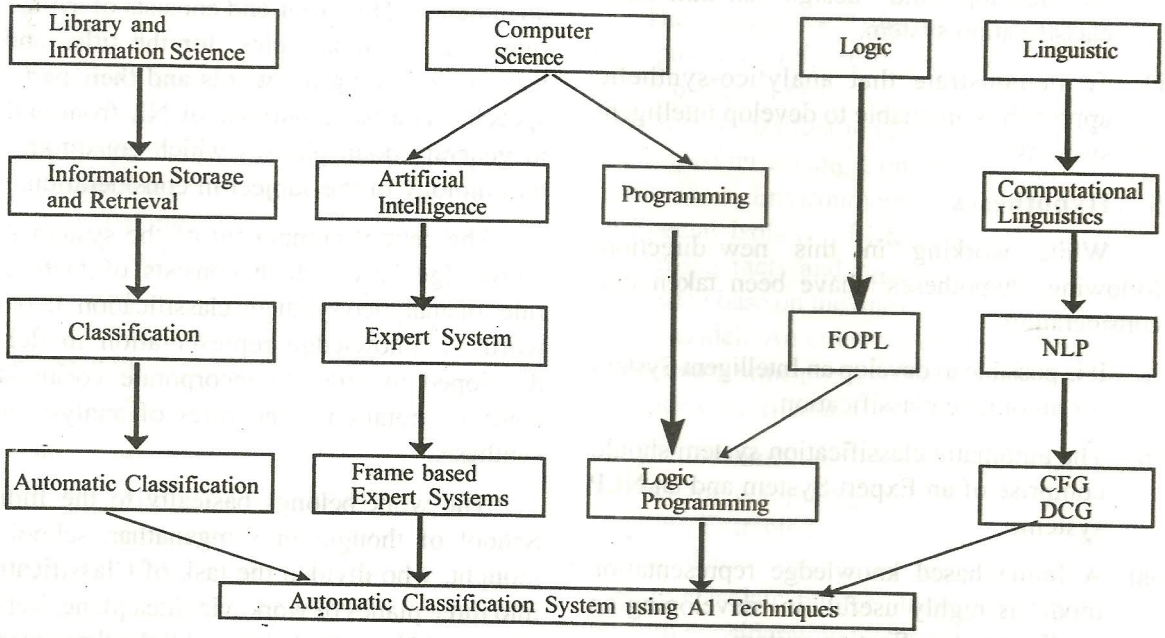

Fig 1 : Scope and Interdisciplinary Approach

As fully automatic text analysis of a document is yet to be achieved, the preparation of expressive titles from the raw titles is to be done manually. This expressive title would be used as input to the system. With the availability of tools and techniques of Artificial Intelligence, the present study makes an attempt to automate steps 2-7 (of total nine steps as suggested by Dr S R Ranganathan). In Fig 1, FOPL implies First Order Predicate Logic, NLP implies Natural Language Processing, CFG implies Context Grammar and DCG implies Definite Clause Grammar[20].

4 Objectives

Following are two major objectives[21] expected to be achieved through this research study

i) To develop and design an automatic classification system.

ii) To demonstrate that analytico-synthetic approach is amenable to develop intelligent systems.

5 Hypotheses

While working in this new direction, following hypotheses[22] have been taken into consideration

i) It is possible to develop an Intelligent System for automatic classification.

ii) The automatic classification system should comprise of an Expert System and an NLP system.

iii) A frame based knowledge representation model is highly useful in developing an intelligent classification system.

iv) Analytico Synthetic classification is amenable for developing such a system.

These four hypotheses are tested in this research study.

6 Methodology

In order to achieve the specified objectives and to test the hypotheses, about 5000 titles in the field of Medicine are collected. The study is confined to a single subject i.e. Medicine, as the purpose of the study is to develop a Knowledge Representation Model. Secondly, it would be a team work to develop a comprehensive system for all the subjects. However, it is to be noted that as the proposed model works for medicine, the model can be extended to other subjects also.

The system has a natural language front-end, in order to scan the expressive titles of documents. This front-end consists of syntactic rules (i.e. grammar rules) for the titles and a lexicon consisting of words and their parts of speech. The basic purpose of NL front end is to generate noun phrases which constitute the terminology of the subject in consideration.

The central component of the system is a knowledge base, which consists of facts and rules of analytico synthetic classification. In other words, a knowledge representation model is developed in order to incorporate vocabulary control, notation, and rules of analysis and synthesis.

The work belongs basically to the Indian School of thought or Ranganathan school of thought, who divided the task of Classification into three planes of work, viz. Idea plane, Verbal plane and Notational plane. All the three planes are involved while constructing a class number.

He has also suggested the following steps[23] to be followed by a classifier

Step 0 -- Raw title
Step 1 -- Expressive title
Step 2 -- Title in kernel terms
Step 3 -- Analysed title
Step 4 -- Transformed title
Step 5 -- Title in standard terms
Step 6 -- Title in facet number
Step 7 -- Preparation of Class number
Step 8 -- Verification.

While building the computational model, the present study also follows more or less the same approach.

Once such a system is developed, it is proposed to test the system with more titles and compare the results with manually prepared class numbers. The result of comparison will be presented in order to highlight the strengths and weaknesses of the system.

As my work is specially to test the hypotheses and to observe the efficiency of the designed Expert System in building classification numbers, no fixed sampling technique is used in the selection of bibliographical entries from BNB (British National Bibliography). However, it is ensured, that assorted titles are taken into consideration for building, testing and evaluation of the system[24].

7 The present system at a glance

This is an attempt to design and develop a Knowledge Representation Model for Analytico Synthetic Classification specially for colon classification with special reference to medicine and with an intention to develop an Expert System for Automatic Classification System using Natural Language Processing Techniques.

'Medicine' is used as sample for demonstration and testing the hypotheses as well as results. Entries from the British National Bibligraphy are collected for this purpose. A parser is developed with the Context Free Grammar (CFG), to be specific Definite Clause Grammar (DCG) to analyse Expressive Titles. The parser includes a huge lexicon comprising of words and their syntactic categories (i.e. their parts of speech, accepted standard spelling etc). More than two thousand titles are tested. The parser is used to extract noun phrases (technical terms) with the help of Grammar and lexicon. On the other hand, the role of each noun phrases is identified using semantic analysis technique for Frame based Knowledge Representation Model. Considering the advantages of declarative programming, the Prolog language is used to develop the Parser, Knowledge Base, Inference Engine etc.

The Expressive Titles in natural language are input to the Expert System and are analysed by the parser in order to identify noun phrases. The Colon Classification Schedules of Common Isolates, Anteriorising Common Isolates (ACI), Space, Time, Environment Divisions etc and the Special Isolates of Medicine are converted into Prolog facts and rules in order to build a knowledge base on the line of frame based expert system model. An inference engine is developed incorporating principles, canons, laws for Analytico Synthetic Classification.

The inference engine comprises a large set of rules for computing classification numbers properly. Extracted noun phrases are consulted with the knowledge bases for picking up the respective isolate number taking synonyms, if any, into consideration. The indicator digits for respective fundamental categories be picked and

put before the respective isolate numbers. While generating classification number facilities regarding Devices such as Alphabetical Devices, Subject Devices, Chronological Devices etc are incorporated. Also incorporated facet formulas for respective main subjects which could be followed for constructing the classification number accordingly in respect of facet sequence.

It is felt that emphasizes may be given to adopt Context Sensitive Grammar instead of Context Free Grammar to develop a parser which would give more powerful Expert Systems. However, this study directed a new approach towards designing automatic classification system. Lastly it is hoped that this work would be more utilised in developing and designing a total Expert Systems for Analytico Synthetic Classification[25].

8 Conclusion

Developing an automatic classification system stands as one of the major interesting research areas in Library and Information Science. Though the present work may not be considered as a fully automatic classification system, it claims to be a step towards building such a complete system in future. The present work demonstrates fairly well that analytico synthetic classification systems are well suited for computerization to a large extent. It is an interdisciplinary approach and builds an automatic classification system which has been named as *VIŞWAMITRA*[26]. Viswamitra is one of the great ancient sages in Hindu mythology. The etymological meaning of his name (*Viswa* means Universe and mitra means *friend*) is that he is a friend of the Universe. It is believed that he created a sibling object for every object in the World. If two objects are found with the same purpose, it is believed, the latter is Viswamitra's creation. Obviously, the story is a brain child of the earlier classifiers, who were perplexed by the similarities in the World of objects. the automatic classification system *VISWAMITRA* includes a Parser, Inference engine, Lexicon, Grammar and Knowledge base for basic subjects (e.g. Medicine), Space, Time, Anteriorising Common Isolates etc. The inference engine and parsar (which includes Grammar, lexicon) and knowledge base are designed using PROLOG code.

References

1 VICKERY (A) and others. PLEXUS : the expert system for referral. *Information Processing and Management.* 23(2); 1986; 99-117.

2 PRASAD (A R D). Application of computer based natural language processing tools and techniques in developing subject indexing languages. Karnataka University. *Ph D Thesis.* 1993. (*unpublished*)

3 BHATTACHARYYA (G). A general theory of subject indexing languages. Karnataka University. *Ph D Thesis.* 1980. (*unpublished*) 17-19.

4 DEVADASAN (F J). Computer based systems for generating different types of subject indexes and alphabetical classaurus based on deep structure of subject indexing languages. Karnataka University. *Ph D Thesis.* 1984. (*unpublished*)

5 RANGANATHAN (S R). Documentation, computer and classification. *Library Science with a slant to Documentation.* 4(3); 1967; 197-225.

6 RANGANATHAN (S R). Doc finder. *Library Science with a slant to Documentation.* 5,4; 1968; 289-302.

7 BHATTACHARYYA (G). A general theory of subject indexing languages. Karnataka University. *Ph D Thesis.* 1980. (*unpublished*). 25-31.

8. RANGANATHAN (S R). A descriptive account of colon classification. 1965 (Reprinted in 1990). Sarada Ranganathan Endowment for Library Science; Bangalore; 276-277.

9 RANGANATHAN (S R). Prolegomena to library classification. (*Reprinted in 1989*). 3rd edition. Sarada Ranganathan Endowment for Library Science; Bangalore, 1967.

10 RANGANATHAN (S R). Colon classification, edition 7 (1971) : a preview. *Library Science with a slant to Documentation.* 6; 1969, 204-206.

11 RANGANATHAN (S R). Colon classification. *7th ed, ed by* M A Gopinath. Sarada Ranganathan Endowment for Library Science; Bangalore, 1987, 8-9.

12 CHARNIAK (E) and Mc DERMOTT (D). Introduction to artificial intelligence. Addision Wesley; Massachusetts, 1985, 6-7.

13 PATTERSON (D W). Introduction to artificial intelligence and expert systems, Prentice Hall of India; New Delhi, 1992, 2.

14 ROWE (Neil C). Artificial intelligence through PROLOG, Prentice Hall; New Jersey, 1988.

15 CHARNIAK (E) and Mc DERMOTT (D). Introduction to artificial intelligence, Addission Wesley; Massachusetts, 1985, 6-7.

16 PRASAD (A R D). Introduction to artificial intelligence. DRTC *Refresher Workshop on Artificial Intelligence and its application to Library and Information work.* Documentation Research and Training Centre, Bangalore, May 26-28, 1993. Documentation Research and Training Centre; Bangalore. 1993. Paper A.

17 WINSTON (P H). Artificial intelligence. *2nd ed.* Addision Wesley; London, 1984, 10-23.

18 PATTERSON (D W). Introduction to artificial intelligence and expert systems, Prentice Hall of India; New Delhi, 1992, 3.

19 WINSTON (P H). Artificial intelligence. *2nd ed.* Addision Wesley; London, 1984, 39-45.

20 PANIGRAHI (P K). Design and development of a knowledge representation model for analytico synthetic classification. *Ph D Thesis*. Vidyasagar University, Medinipur, 1986, (*unpublished*), 46-64.

21 PANIGRAHI (P K). Design and development of a knowledge representation model for analytico synthetic classification. *Ph D Thesis*. Vidyasagar University, Medinipur, 1996, (*unpublished*), 6.

22 PANIGRAHI (P K). Design and development of a knowledge representation model for analytico synthetic classification. *Ph D Thesis*. Vidyasagar University, Medinipur, 1996, (*unpublished*), 6.

23 RANGANATHAN (S R). Colon classification, *7th ed, ed by* M A Gopinath. Sarada Ranganathan Endowment for Library Science; Bangalore, 1978, 14-15.

24 PANIGRAHI (P K). Design and development of a knowledge representation model for analytico synthetic classification. *Ph D Thesis*. Vidyasagar University, Medinipur, 1996, (*unpublished*), 6-8.

25 PANIGRAHI (P K). Design and development of a knowledge representation model for analytico synthetic classification. *Ph D Thesis*. Vidyasagar University, Medinipur, 1996, (*unpublished*), 338.

26 PANIGRAHI (P K). Design and development of a knowledge representation model for analytico synthetic classification. *Ph D Thesis.* Vidyasagar University, Medinipur, 1996, (*unpublished*), 8.

Acknowlegements

The author expresses deep gratitute to Prof. Asok Basu, University Librarian, Calcutta University and Prof. A R D Prasad, Documentation Research and Training Centre, Indian Statistical Institute, Bangalore for their valuable guidance in relation to this research study.

IASLIC Bulletin 45 (3), 2000, p. 105–118

An Artificial Intelligence approach towards Automatic Classification (Part 2)

Dr. P. K. Panigrahi
Reader & Head, Dept. of Library and Information Science,
University of Kalyani, Kalyani, Nadia, West Bengal, India
(e.mail : pkpanigrahi@hotmail.com or pkpanigrahi@library.iitkgp.ernet.in)

Gives an overview on expert systems (ES) and natural language processing (NLP), the two major areas of artificial intelligence (A1), with their short histories and landmarks. Indicates previous researches in connection with automatic text processing emphasising to specify the impact of linguistics/computational linguistics. Brings out the possible ways of analysing expressive titles to identify noun phrases (i.e. facets, isolates etc.). Discusses the phrase structure tree for analyzing expressive titles those are in natural language. Describes the scope of applying NLP techniques in library and information works. Mentions different areas in the field of library and information sciences where artificial intelligence techniques can be applied. Shows the successful result of applying AI towards developing an automatic classification system, the main objective of the present study.

1 Introduction

Numerous attempts have been made by researchers to design a powerful automatic classification system, but those could not bring any well accepted result for a long time. The main problem is successfully automatic analysis of titles of documents and identifying subject propositions. Because being a purely mental process, classification demands human intelligence for analysing the title to find out its basic subject and other facets, if any, along with its category and also synthesising those facets according to principles and postulates to construct classification number. In the other words, the document title, which is in the natural language, is analysed carefully to pick up relevant words (i.e. subject propositions) and those are synthesised using classifier's expertise to build the classification number. Application of Artificial Intelligence (AI) is the solution to this problem. Use of Natural Language Processing (NLP) techniques will help in automatic analysis of title and an Expert System can be developed to work exactly in the same way as a classifier does to build classification numbers being guided by canons, principles and postulates.[1]

2 Expert Systems (ES)

21 *Definition*

Expert systems are sophisticated computer programs that manipulate knowledge to solve problems efficiently and effectively in a narrow problem area. The term 'expert systems' can also be referred to computer programs in which

substantial knowledge of an expert on some specific area can be applied to the problem solving process. Also, a system can be thought of as an expert system if it is in a position to handle real world complex problems considering all possibilities suggested by a human expert, and solve these problems using the concept of human reasoning, reactivity etc. Again an expert system can be compared with a computer-based system that achieves the high level of performance in the task areas for which human being requires years of special education and training.[2] However, Edward Fiegenbaum defined an expert system as ".....an intelligent computer program that uses knowledge and inference procedure to solve the problems that are difficult enough and also requires significant expertise for their solution. Knowledge necessary to perform at such a level plus the inference procedure used, can be thought of as a model of the expertise of the best..... practitioners of the field. The knowledge of an expert systems consists of facts and heuristics. The 'facts' constitute a body of information that is widely shared, publicly available and generally agreed upon by experts in a field."[3]

The process of building an expert system is often called Knowledge Engineering[4] that relies heavily on the study of human experts in order to develop intelligent, skilled programs. The central notion of intelligent problem solving is that a system must construct its solution selectively and efficiently from a space of alternatives. When resources are limited, the expert needs to search this space selectively, with as little unfruitful activity as possible. An expert's knowledge helps to spot useful data early, suggests ways to exploit them and helps to avoid low-payoff efforts by pruning blind alleys as early as possible.

22 History of Expert Systems

The ultimate goal of the AI scientists had always been to develop computer programs that could have some sense of thinking as human being. In fact physicians attempted to analyze the brain functions of those whose brains are not working properly. They have tried remedies, ranging from opium to electric shock as they have attempted to have better understanding in the inner workings of brain. The psychologists concern themselves with measuring human intelligence and study factors that influence its development. At the same time, the linguists attempt to investigate the development of language and the procedure of communication. Most of these directly relevant works were sponsored by intelligent agencies. Those were mostly interested in automated translation of scientific journals and automatic interpretation of machine readable data.[5]

Pre 1956

Alan Turing foresaw thinking machines in 1950. He suggested computers can be programmed to exhibit its intelligence and proposed his famous "Turing Test".[6] Claude Shanon suggested the computer chess model and also shared Alan Turing's conviction about the possibilities of machine's thinking process.[7]

1956

For the first time the term 'artificial intelligence' was coined at the Dartmouth Conference organized by John McCarthy and Marvin Minsky during the summer of 1958.[8] Alan Newel and Herbert Simon developed the General Problem Solver systems.[9]

1960's

Scientists tried to simulate complicated process of thinking by finding general methods

for solving broad classes of problems. Developing General Problem Solving programs was too difficult and ultimately fruitless. A single program could handle more classes of problems, the more poorly it seemed to do on any individual problem.[10]

1970's

Scientists realized that there must be a way to make a computer program intelligent. Though it was too difficult to make the entire program as a general purpose one, they concentrated instead on developing general methods or techniques to be used in more specialized programs. They concentrated on representation techniques i.e. how to formulate the problems so that it would be easy to solve, and search techniques i.e. how to cleverly control the search for a solution so that it would not take too long or use too much of the computers memory capacity. This way of thinking and its implementation got some success but not breakthroughs. However, the scientists slowly realized by late 1970's that the problem solving power of a program comes from the knowledge it possesses, not just from the formalism and inference schemes it employs. And it was felt that to make a program intelligent, provide it with lots of high-quality, specific knowledge about some problem area. This realization led to the development of special purpose computer programs, systems that were experts in some narrow problem area. The programs were called expert systems.[11]

Mid 1970's

Several expert systems had begun to emerge. A few investigators who recognized the central role of knowledge initiated efforts to develop comprehensive knowledge representation theories and associated general purpose systems. It became apparent that these efforts had limited success for reasons similar to those that doomed the first general problem solvers. 'Knowledge' as a target of study is too broad and diverse; efforts to solve knowledge based problems in general are premature. On the other hand, several different approaches to knowledge representation proved sufficient for the expert systems that employed them. The lesson learned from these experiences was that the expert's knowledge provides the key to expert performances while knowledge representation and inference schemes provide the mechanism for its use.[12]

23 Types of Expert System

Expert sysem can be grouped broadly into the following, depending on their objectives.[13]

	Systems	Objectives
1.	Interpretation	Inferring situation description from sensor data.
2.	Prediction	Inferring likely consequences of a given situation.
3.	Diagnosis	Inferring systems malfunctions from observable facts.
4.	Design	Configuring objects under constraints
5.	Planning	Designing actions.
6.	Monitoring	Comparing observations to plan vulnerabilities.
7.	Debugging	Prescribing remedies for malfunction.
8.	Repair	Executing a plan to administer a prescribed remedy.
9.	Instructions	Diagnosis, debugging, and repairing ~~student~~ System behaviour.
10.	Control	Interpreting, predicting, repairing and monitoring system behaviour.

However, some expert systems are found to be as a combination of any two or more systems listed above.

3 Natural Language Processing (NLP)

31 Definition

The attempt to make computers understand natural language was called Automatic Language Processing. However, the term, 'Natural Language Processing,' (NLP) is at present widely used. The first generation of NLP work in the 1950's, in automatic language processing ended in disappointment as they had the most ambitious goal i.e., Machine Translation (MT). Now it is well realized that natural language is one of the most complex artifacts of human mind. The present day research is mostly with limited goals, and tries to find solutions at different levels viz., morphological, lexical, syntactic, semantic and pragmatic levels.[14,15]

32 Scope of NLP

Theoretical linguists are primarily interested in producing a structural discipline of natural language. They are not interested in parsing or language generation from structural descriptions. A major objective of theoretical linguists is to develop theories that hold good across languages. In other words, they attempt to characterize the general organizing principles that underlie all human languages and do not usually examine any particular language. A linguistic theory is only useful to the extent that it explains actual behaviour. Consequently, Psycho-linguists are interested in both the representations of linguist's structure from actual sentences. The primary tool that is used in experimentation in which actual measurements are made from people as they produce and understand language, including how much time a person needs to decide whether a given item is a legal word or not ; what types of errors people make as they perform various linguistic tasks and so on. Experimental data is used to validate or reject a specific hypothesis about language, which are often taken from the theories that linguists and computational linguists propose.[15]

Natural language understanding requires a knowledge of how the words are formed, how the words in turn form clauses and sentences. In addition, to successfully understand a set of sentences in a given context, it should have a higher level knowledge. In general, the knowledge that is to be used in natural language understanding is divided into the following[16]:

Morphological Knowledge: This deals with the morphological structure of word root, prefix, suffix and infixes. The basic unit in a written word is a morpheme. Thus, this level gives knowledge of word formation.

Lexical Knowledge: This level deals with thesaurus look up, spell corrections, acronyms and abbreviations and the parts of speech.

Syntactic Knowledge: Syntax deals with the structure and validity of sentences and now a right combination of words in a particular sequence constitute a valid sentence.

Semantics Knowledge: Semantics deals with the meaning of words and that of sentences.

Pragmatic Knowledge: Pragmatic level deals with sentences in a particular context. This requires a higher level knowledge that relates to the uses of sentences in different contexts. It does not only deal with the meaning but also with the intentions of the statements.

World Knowledge: In order to carry out effective communication, both the communicator and the communicatee should have a background

knowledge either to send or to receive a message without any noise. This background knowledge is considered as the world knowledge of a particular domain.

4 Automatic Processing of Text in Expressive Titles

The primary objective of this research study is to develop an automated classification system using artificial intelligence. Success of such a system is much dependent on its capability of automatic analysis of expressive titles which is a text in natural language. There have been a remarkable amount of studies reflecting recent developments in automated language processing area which emphasizes on computer understanding, analysis and assessment of various technical and non-technical terms. Automatic language processing in the context of library classification means automated text processing especially automatic processing of text appeared as expressive titles. In fact, automatic language processing is the more general term, encompassing all studies, theoretical or applied, of the use of computers or the computational techniques in the processing of lanugage, primarily natural language. On the other hand, Linguistics has important role to play in information science particularly in relation to the use of computers for linguistics operations involving document analysis, description and retrieval or as a whole in documentation.

41 Role of Linguistics and Computational Linguistics

In relation to automated language processing and information science, linguistics is the primary reference point, although the perspective on computational linguistics implicates psychology as well as descirptive lingustics and computation.[17] Computational linguistics is considered to be a proper subset of automatic language processing, distinctive in its relation to linguistcs. However, it has been cleared that computation linguistics is the primary focus for works in the whole area. Again by nature computational linguistic itself has become a genuine interdisciplinary field with a focus on "linguistics algorithm", their structure and application, and with close historical, conceptual and practical ties to linguistics, although with a concern legitimately far broader than natural language.[18]

Actually, the greater availability of computers and ever-increasing sophisticated technology of interactive programming have provided much hope of allowing computational linguistics who keep current with the rapid pace of developments in linguistics of feed back the result of their computational research to linguist in time to be of use to him. Consequently, an accurate assessment of the potential value of computational linguistics in support of the linguist's research remains for the future. However, it is contingent on the willingness and ability of the linguist to formulate problems in computationally relevant ways as well as on the proximity and responsiveness of the computational linguist. For a long time, scientists have tried to design a fully developed system of linguistic description that is equally relevant for computation linguistics. As a result of such researches it is found that "the linguistic description of a language

(i) must characterize, for each lexical item in the language.

(a) the grammatical construction in which it can occur,

(b) the grammatical processes to which it is subject in each relevant context,

(c) the grammatical processes which its presence in a construction determines, and

(d) information about speech and conditions, conversation rules, and semantic interpretation which must be associated in an idiosyncratic way with the lexical item in question;

(ii) it must provide the apparatus which characteristics

(a) the grammatical structure of sentences on the 'deep' or abstract level, and

(b) the grammatical processes by which abstract linguistic structures are processed and become surface sentence;

(iii) it must contain a component for calculating the complete semantic and pragmatic description of sentence given its grammatical structure and information associated with each lexical item.

(iv) it must able to draw on the theory of illocutionary acts, in terms of which the calcutations of (iii) are empowered to provide a ful account of potential illocutionary force of each sentence.

(v) it must be able to draw on a theory of discousre which relates the use of sentences in social and conversational situations; and

(vi) it must be able to draw on a theory of 'natural logic' by means of which such judgements as the success of an argument or the appropriateness of elements in conversation can be deduced."[19]

The description provided a framework that began to develop systems for automated language processing to make it capable to analyse language, though for a long time none of the problems could be considered solved, but many promising direction could be explored with appreciation.

A remarkable number of the efforts have made changes in the systems reflecting deep and surface structures of a language. Syntactic and semantic features lead to analysis of sentence. Procedures for deduction and inference have become more sophisticated and developments in artificial intelligence have reflected in new programming techniques and new heuristics. Slowly a transformation technique for a class of transformational grammar is developed emphasizing the role of lexicon as an important component of generative and recognition grammar. Syntactic and semantic analysis modules are customized to provide an English-language interface between a user and an arbitrary computer system to query library catalog. However, it is considered that a lot of more work needs to be done in lignuistics, in information science and particularly in attempts to relate these two fields in order to establish exactly what the most productive relation between the two fields can be.

4 2 Title Analysis and Retrieval of Noun Phrases

Sometime, it is felt that the boundaries between linguistics / computational linguistics and rest of the automated language processing are not clear and distinct. Actually it is reasonable both to predict and to argue for an increasing comprehensiveness in the models of language underlying computational linguistics. As a result, all manipulation of language data in the information science is able to reflect the resulting increase in sophistication. A good number of

attempts are made to develop a simple system for personal information storage and retrieval for facilitating personal information processing. These systems include heuristic-based parser for establishing syntactic dependencies in topic descriptions, a network structure for organizing topics, improved retrieval mechanisms, and various kinds of topics network-processing techniques. The further development of personal and intellctually supportive information system is likely to highten the awareness of lack of our understanding of the significance of differences among users in a system[20].

Development of a large number of systems for analysing the structure and content of documents are in progress. Already developed systems attempt either to present data representation code for use in text-processing system, or to handle computer-based recognition of semantically related words in any document/text or to describe programming language for use in text-manipulation. But none of these systems could present major innovations except a few of them could provide similar capabilities on different computer facilities. Some of them adopted statistical approach in text analysis to identify noun phrases. Through experiences, scientists could realise that NLP should be considered as a primary concern to develop any content analysis and retrieval system due to the success of experiments with SMART retrieval system dealt with language elements specifically[21]. Since that period a good number of research results are coming up in this area. The general question of natural language as a query language came into picture again. The variety of intermediate languages for automatic language processing between natural language and machine language are provided. Also a procedure for automatically constructing word discriminator lists for use in the SMART system is developed which in addition could provide an automatic way for constructing a term dictionary.

Bhattacharyya[22] in his study on "A General Theory of Subject Indexing Language (GT/SIL)" demonstrates the normal method of transforming deep structure into surface structure index entries. Afterwards, the questions of NLP in natural language interfaces to databases are raised. In this regard, Vickery[23] observed that most parsers described in NLP requires a verb to be identified during elucidation of the sentence. They have shown that rarely does a query put to online system contain a verb. However, the basic approach of identifying the associated categories of a verb could be taken as model, by replacing verb with noun phrases.

Recently, Prasad[24] has succeeded to landmark in automated language processing area with the approaches of Bhattacharyya and Vickery. In building PROMETHEUS, an automatic indexing system using NLP techniques, he has developed an NLP system using a knowledge representation model to represent the meaning of given expressive titles. This system has two basic components — syntactic and semantic. The syntactic parser is basically developed to build syntactic structures for a given sentence and the syntactic parsing aimed at producing noun phrases. In other words the captured noun-phrases be identified along with associated categories. The dictionary contains words indicating their grammatical classes, their category(ies), pointers to their position(s) in hierarchy, pointers to entries which are to be regarded as synonyms and a tag to signal how the entry is to be truncated etc. Where as the second components, semantic processing aims at

converting the syntactic representation into a kind of internal representation to the machine. The present research study on developing an automatic classification system using AI techniques is influenced by Prasad's work.

5 Phrase Structure Tree to Analyse Titles

The present study is mainly carried out with an intention to develop an automatic classification system for analytico synthetic schemes with special reference to Ranganathan's Colon Classification 7th edition. This model suggests to have two major components— semantic and syntactic representation. The semantic representation (i.e. semantic knowledge) includes the knowledge base where the facts (i.e. schedule) are stated to make the system sufficiently knowledgeable about the domain, the inference engine, expert system. The syntactic representation (i.e. syntax of the language) includes lexicon (lexical knowledge), the parser, the grammar formalism (i.e. structural knowledge of the natural language sentences). For the present work, the expressive titles are prepared in natural language using abstruct, titles and keywords and are input to the system for processing. The system accepts and parses these natural language expressive titles to check their syntax with the help of lexicon and grammar formalism. These natural language expressive titles are analyzed in two phases. The first one is syntactic analysis which would check the syntactical correctness of structure of the sentences (here expressive titles) with the help of grammar formalism where the correct structure is defined in the light of phrase structure tree depending on the lexical status of the words or word compounds[25]. A noun phrase (NP) is a self contained unit and can fully interpreted by the time it has been read. That means it can always be determined what objects a noun phrase refers to. However, the important part is that the noun phrase is the valuable aid to achieve the objective of the present work. Because once these noun phrases (i.e. isolate ideas) are identified, then it is straight forward for the expert system to extract all information about it from the knowledge base.

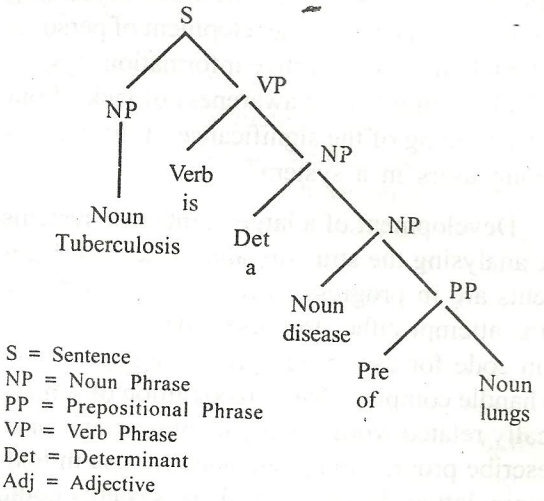

Fig. 1 :Illustration of Phrase Structure Tree

The second phase of this analysis part of NLP system is semantic analysis to represent the meaning of the words with the help of knowledge of the specific subject domain represented or modeled following the techniques of analytico-synthetic classification. The present system analyses the text and picks up the noun phrases. The knowledge base is developed on the lines of subject classification based on Ranganathan's ideas. That means all facets related to this domain are represented in terms of their fundamental categories, numbers etc. For example, if 'Treatment of Shoulder Pain by Injection' is input as expressive title to the system, the title will be analyzed on the basis of

grammar and the lexical status of the terms. The noun phrases are formed and picked up analyzing sentences, prepositional phrases, verb phrases and ignoring prepositions, articles, conjunctions, gerund etc. So from the above example, 'Treatment', 'Shoulder Pain' and 'Injection' would be collected.

6 NLP in Library and Information Work

Natural language is one of the most complex artifacts of human mind. NLP offers much promises to library professionals, although the present state of research is far from the actual goal, some of the advantages of NLP based library and information work are as following[26].

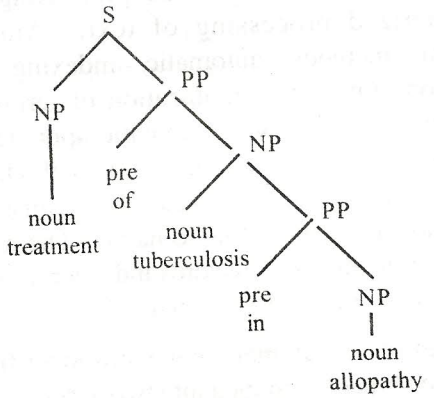

Fig. 2 : Tree structure for 'Treatment of tuberculosis in allopathy'

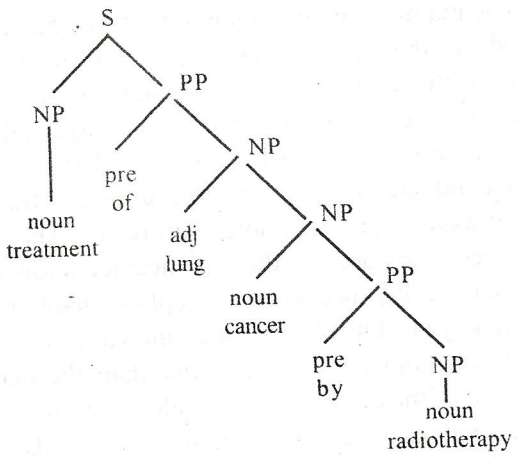

Fig. 3 : Tree structure for 'Treatment of lung cancer by radiotherapy'

i) Natural language front ends to the existing bibliographic database management systems could be developed so that the user could avoid learning the search languages in order to retrieve the required information.

ii) Consequently, such interfaces with a proper knowledge of the search strategies could substitute search intermediaries.

iii) Natural language processing could well be used in the automatic identification of noun phrases from the expressive title and assigning their respective isolate numbers (i.e. the notion of the present work) to build classification numbers.

The present study is based on the NLP application in classification and a step towards developing an automatic classification system for Ranganathan's Colon Classification.

61 Information Retrieval

Information retrieval has become an important research area for a long period to enrich the performance efficiency of the system. The relationship between Linguistics and Information retrieval has been a controversial one. A number of researchers among them Lancaster[27], Salton and McGill[28], noted in the 1970's and early 1980's that there had been a move towards simplification in information retrieval and that there might in fact be no need for the level of detail sophistication provided by linguistics analysis. However, the review of literature gives the impression that there is a large common ground between linguistic theory

and information science and this is further supported by the renewed interest in linguistics as part of natural language processing application in information retrieval[29].

The application of NLP assumes that the documents and user queries under consideration in information retrieval are often available as natural language formulations and it is important, therefore, to be aware of the automatic methods currently used to process natural language text.

In Information Retrieval, NLP offers the potential for friendly, flexible interface through which the user can state the information requirements in natural language instead of expressing in a query language. These systems are also called as Natural Language frontends and they allow users to state relatively complex queries without being familiar with complicated formal retrieval languages. It is also expected that the use of queries in natural language could raise the efficiency and effectiveness by making possible formulation of precise requests that correctly reflect user needs and by simplifying the user-system interaction. In information storage, NLP systems allow us to structure large quantity of textual information so that retrieval of individual facts, gathering of statistics and preparation of summaries are possible[30].

Another possible application of NLP in Information Retrieval Systems is with regard to the construction of thesauri. A thesaurus can be constructed using word co-occurrence, information. However, if linguistic descriptors are available to characterize the individual text units, the thesaurus class might be defined on the set of words occurring in similar contexts in the documents of a collection[31]. Although, it is a distant possibility, the success in NLP systems could give rise to automatic abstracting and machine translation using titles and abstracts of documents, if not the full text of a document.

62 Automatic Indexing

An ideal automatic indexing will be one that has the ability to create and modify new subject terms automatically, by minimizing or without the help of human intellect. During the last few decades, investigators have sought ways to replace human intellectual processing by computerized processing of text. Among different methods automatic indexing by extraction, automatic organization of terms or documents into classes, automatic approaches to thesaurus construction, and method for matching natural language requests against the text of document are a few to name. The main objective is to identify keywords/index terms with the help of controlled vocabulary[32].

Attempts in automatic key word identification can be broadly grouped into two approaches. One is using statistical methods, initiated by H P Luhn and the other is using NLP techniques. The major problem with purely statistical methods is that they do not make any attempt at semantic analysis of the thought content of a given document. In natural language, all the pronouns and words like 'former', 'latter', 'first', 'second' etc. are used to indicate to a particular keyword appearing more than once. They are called anaphora. As the statistical techniques do not have any mechanism to replace anaphora by the keyword it refers to, they increase the count of anaphora occurrence rather than the occurrence of the keywords. Secondly, the statistical techniques deal with compound words in a rather clumsy way, by studying the frequency of co-occurrence of words. Identification of

three or four worded compound terms becomes even more difficult as they depend on frequency of co-occurrence of words. The very nature of recursive structures in natural language can be tackled by rising recursive data structures in NLP to deal with compound terms like *deep blue sea* or *reinforced cement concrete bridges*. Natural Language system can also attempt the problems of compound words when the words are presented in different forms, for examples, 'College libraries' and 'Libraries of colleges'. Again, noun phrases like 'college and research libraries', pose a different kind of problems. There are, in fact, two compound words viz. 'college libraries' and 'research libraries'. However, heuristics can be developed to identify compound words by developing a rule like — *when two adjectives are separated by a conjunction, followed by a noun, each adjective and the noun are to be combined to form a separate compound term*. NL parsers could be easily supplemented by a lexicon or thesaurus to deal with the problems of plurals, synonyms etc. Problems with regard to plurals and irregular plurals (like focus-foci, stadium – stadia etc.) can be dealt with by Natural Language systems more effectively[33].

63 Automatic Classification

Numerous researches have been carried out for a long time in this area. One of the main problems in automatic classification is the analysis of thought content of the document as appeared in the expressive title. Being a purely mental process, library classification demands human intelligence for analyzing the title to find out its basic subjects and other facets, if any, along with its categories. Researches in automatic classification can be broadly grouped into two: (i) Cluster analysis and (ii) using AI techniques for finding the facets. Survey of literature shows that most of the works are related to cluster analysis. They are based on statistical calculations.

Another approach to automatic classification is the use of AI techniques in building class numbers. Scientists feel that if human intelligence is evaluated by the computer artificially, it may then be possible to solve the problem of identification of keywords/noun phrases for automatic classification. Expressive titles are generally in natural language i.e. a language which is learnt by human easily. Human brain is capable of analyzing the natural language and understanding its semantics inspite of its ambiguities. The crux of natural language processing is to develop semantics of any piece of information by machine[33]. NLP is knowledge dependent. Every sentence of natural language contains information. An NLP system should able to contain the syntactic knowledge, knowledge of the vocabulary of a particular language and most importantly it should contain knowledge of particular domain in order to understand the meaning of sentence. Different domains contain specific knowledge of a subject area. An NLP technique can be applied to analyze the expressive title depending on the relation on the domain knowledge and natural language used to express it. These systems basically use an NL front-end and an expert system to achieve the goal.

My attempt is to design and develop a Knowledge Representation Model for Analytico Synthetic Classification specially for colon classification with an intention to develop an Expert System for Automatic Classification using Natural Language Processing Techniques.

The two major components of the present system are syntactic component and semantic component. The syntactic component includes a parser containing the grammar formalism (i.e. structural knowledge of the natural language sentences) and a lexicon (i.e. lexical knowledge). On the other hand, the semantic component includes a knowledge base which consists of the classification scheme represented in terms of a frame based knowledge representation model and a rule base consisting of rules to construct class number for the expressive title (which represents the thought content) of the document. The syntactic component generates the noun phrases and passes the syntactic structure to be processed by the expert system, inference engine and rule base. Here, it has been believed that a noun phrase is a self-content unit and can fully be interpreted by the time it is read. That means, it can always be determined as to what objects a noun phrase refers to. However, the important part is that the noun phrase is a valuable aid to achieve the goal of the present study[34].

A parser is developed with the Context Free Grammar (CFG), to be specific Definite Clause Grammar (DCG) to analyze expressive titles. The parser includes a huge lexicon comprising of words and their syntactic categories. The parser is used to extract noun phrases(technical terms) with the help of grammar and lexicon. On the other hand, the role of each noun phrases is identified through semantic analysis using Frame-based Knowledge Representation Model. An inference engine is developed incorporating scientific theory of library classification comprising a large set of rules for computing classification numbers properly. Noun phrases are extracted and consulted with the knowledge bases for picking up the respective isolate number taking synonyms, if any, into consideration. The indicator digits for respective fundamental categories be picked and put before the isolate number. Also facet formulas for respective main subjects be incorporated which could be followed for constructing the classification number accordingly in respect of facet sequence.

7 Conclusion

This new approach will be useful in developing and designing a total expert system for automating library classification procedure. The expressive titles in natural language are input to the Expert System and are analyzed by the parser in order to identify noun phrases. It has been proved through this study that the analytico-synthetic classification especially the Colon Classification (7th edition) is more computer compatible. The Colon Classification Schedules of Common Isolates, Anteriorising Common Isolates (ACI), Space, Time, Environment Divisions etc. and the Special Isolates of different basic subjects may be translated into Prolog facts in order to build a knowledge base on the line of frame based expert system model.

It seems appropriate to conclude, on basis of above discussion, that an automatic classification system for Colon Classification as a whole can be developed and that it will have impact in information organization and retrieval in the automated environment. However, such a big system may have little complexity than most people anticipated, but once such a system is developed, it will be easier in bring out future editions of the Colon Classification Schedule.

References

1. PANIGRAHI (PK). Design and development of a knowledge representation model for analytico

22. BHATTACHARYYA(G). A general theory of subject indexing language. *Ph. D. Thesis*. Karnataka University. Dharwad. 1980. (unpublished).

23. VICKERY (B) and VICKERY (A). Application of language processing for search interface. *Journal of Documentation*. 48(3); 1992. p 255-275.

24. PRASAD (ARD). Application of computer based natural language processing tools and techniques in developing subject indexing language. Karnataka University. *Ph. D. Thesis*. 1993. *(unpublished)*. p 155-231.

25. CHOMSKY (N). Aspect of the theory of syntax. 1965. The MIT Press; Massachusetts. p 31-119.

26. PRASAD (ARD). Application of computer based natural language processing tools and techniques in developing subject indexing language. Karnataka University. *Ph. D. Thesis*. 1993. 292p.(unpublished).

27. LANCASTER (FW). Vocabulary control for information retrieval. 1972. Information Press; Washington. 65-80.

28. SALTON (G) and McGILL (MJ). Introduction to modern information retrieval. 1983. McGraw Hill; New York. p 257-96.

29. PANIGRAHI (PK). Design and development of a knowledge representation model for analytico synthetic classification. *Ph. D. Thesis*. Vidyasagar University, Medinipur. 1996. 57-58 *(unpublished)*.

30. GRISHMAN (R). Natural language processing. *Journal of the American Society for Information Sciences*. 35(5); 1986; p 291-296.

31. SALTON (G) and McGILL (MJ). Introduction to modern information retrieval. 1983. McGraw Hill; New York. p 257-96.

32. PANIGRAHI (PK). Computerised indexing and clustering techniques. In GHOSH (S.B) and SATPATHI (JN), *Ed.* Subject Indexing Systems : comcepts, methods and techniques. 1998. 2nd ed. IASLIC; Calcutta. p 277-307.

33. PRASAD (ARD). Natural language processing. *In* GHOSH (SB) and SATPATHI (JN), *Ed.* Subject Indexing Systems : concepts, methods and techniques. 1998. 2nd ed. IASLIC; Calcutta. p 308-325.

34. PANIGRAHI (PK). Design and development of a knowledge representation model for analytico synthetic classification. *Ph. D. Thesis*. Vidyasagar University, Medinipur. 1996. 338 p *(unpublished)*.

35. WINOGRAD (Terry). Understanding natural language. Academic Press; New York. 1972. 200p.

Acknowledgements

I express deep gratitude to Prof. Asok Basu, University Librarian, Calcutta University, Calcutta and Prof A. R. D. Prasad, Documentation Research and Training Centre, Indian Statistical Institute, Bangalore for their valuable guidance and suggestion in relation to this research study.

We deeply acknowledge the help and guidance we have received from the following persons to bring out the present issue of the Bulletin :

Sri S B Banerjee
Sri Subrata Dutta
Sri Bhabhani Ghosh
Sri Mani Mohon Ghosh
Sm. Sangita Pal
Dr. J. N. Satpathi