CHAPTER VI

CONCLUSIONS AND AVENUES FOR FUTURE RESEARCH

6.1 NOVELTY OF THE RESEARCH AND MAJOR CONTRIBUTIONS

The research work has clearly demonstrated that any rhetoric concept of structured documents will be reliably extracted with predefined terms/ phrases of the rhetoric. This is qualitatively validated from the responses received from experimental group of the social samples.

It is established by the thesis that Cauchy distributions of structured documents will provide a vivid picture on consistency of the rhetoric concepts. Applying Standard Cauchy distribution for extracting rhetoric concepts of textual documents is the novelty of the research.

Results obtained from unstructured descriptive documents in the web environment have clearly proved that Standard Cauchy distribution will also indicate the predominant rhetoric concept of the documents.

Experimental results have proved clearly that Standard Cauchy distribution values will aid in rank ordering the documents according to the regression (R2) values’ consistency and uniformity of the cognitive dimensions (rhetoric concepts) in the contents. It is also demonstrated that apart from the deviations in slope values of the standard Cauchy curve, the probability value will also help in rank ordering the documents.

The thesis has contributed research and utility values through a proposed framework (algorithm) that will rank order documents which are extracted from the Web for any required rhetoric concept.
6.2 RHETORIC CONCEPT EXTRACTIONS AND CAUCHY DISTRIBUTIONS

6.2.1 On Structuring the Concept Documents

It is concluded that, fully-structured documents of any particular rhetoric concept will show more consistent distribution of Cauchy values, than that of semi or ill-structured documents.

It is concluded that, when rhetoric concept words increase in any document gradually from ill-structured to fully structured forms, non-domain (unknown) word frequencies will also increase.

It is found that more and more the structuring of documents then more and more the consistency of the rhetoric concepts would be. It is concluded that this will lead to better regression coefficient thus better concept extraction. It is concluded that consistency in concept terms will be increased when the documents are structured.

It is found in general that, ‘Reasoning’ rhetoric concept words/phrases are less in numbers, than other rhetoric concepts. However, the word frequency of ‘Reasoning’ when compared with other rhetoric concepts, have found to increase in fully structured ‘Reasoning’ concept documents.

It is clearly demonstrated that rhetoric words are very valuable in representing the concepts in documents for reliable extractions.

Rhetoric concept words are found to be more valuable in structured documents.

Recommendation:

It is recommended to fully structure the domain concept documents with appropriate rhetoric words/phrases for best representation of the concept that will help in reliable extraction.
6.2.2 On Instructional Rhetoric Concepts

General Conclusions:

It is concluded that the highest Cauchy probability value along with the maximum consistent regression ($R^2$) value will be the best to decide about the representation of a rhetoric concept in a document.

It is found that, even when Cauchy distributions are consistent with two or more different rhetoric concept documents, the maximum probability value among these distributions will indicate the predominant rhetoric concept.

It is concluded that extremely lower representation of term frequencies of rhetoric concept will yield erratic regression coefficient.

It is concluded that when documents are fully structured with rhetoric concept terms, Standard Cauchy distribution with maximum regression value, will indicate that rhetoric concept representation.

It is concluded that the least range of values from largest to smallest frequencies of any rhetoric concept will be more representative of that concept, where a few concepts even though have the same regression ($R^2$) value.

Specific Recommendation:

Since ‘Utilization’ rhetoric concepts would be mostly represented by mathematical symbols and numerical values, it is recommended not to consider any weight for such documents.

6.2.3 On Reasoning Rhetoric Concept Documentation

In almost all the case studies, the unknown word frequency has gradually reduced from ill-structured to fully-structured concept documents. Even though rhetoric concept frequencies have increased gradually, the ratio of ‘Reasoning’ to other concepts alone
has increased from about 1/5\textsuperscript{th} to about 1/3\textsuperscript{rd} in the fully structured documents. This shows, ‘Reasoning’ concept (which is generally represented with less number of concept words/terms) has increased in fully-structured documents. It is concluded that ‘Reasoning’ rhetoric concept is more likely will indicate correct extraction in fully structured documents.

6.2.4 On Similarity between Cauchy Distributions of Different Rhetoric Concepts

It is observed that in some Cauchy distributions, the least of the maximum Cauchy probability value of any category, has matched with the opinion of the respondents. It is concluded that even when the Cauchy distributions of one category, if found to be similar to other rhetoric cases, the maximum Cauchy probability value would determine the reliable concept representation.

6.3 EFFECTIVENESS OF STRUCTURING TEXTUAL DOCUMENTS

6.3.1 Structuring Documents and Representation of Concepts

The deviations between the researcher and the ratters on the volumes of rhetoric concepts observed in selected question papers are less than 10%. The distributions of different rhetoric concepts are found to be steeper in the question papers. It is concluded that structured question papers will indicate particular rhetoric concept for the purpose of automation.

6.3.2 Cauchy Distributions of Structured Question Papers

The responses provided by the experimental group samples and the distributions of Cauchy values have matched well. It is concluded that structured question papers will yield correct Cauchy values.

The uniform and consistent Cauchy distributions found on structured question papers have revealed that any structured interrogative documents will truly represent particular
6.3.3 Cauchy Representations of Unstructured Descriptive Documents

The results of unstructured descriptive documents have proved that the Cauchy distribution will still work well even though responses through opinions of users differ a lot.

6.3.4 Rank Ordering Documents According to Rhetoric Concepts

The surveyed results have clearly demonstrated that the Standard Cauchy distribution values will aid in rank ordering the documents according consistency and uniformity of rhetoric concepts of the contents. It is also demonstrated clearly that the proposed method will indicate the deviations in slope values of the Standard Cauchy curve, and can help in rank ordering the documents according to consistency. It is concluded that while simple ‘tf/Idf’ term frequency values are used to extract desired documents; the Standard Cauchy distribution will reliably rank order the documents as per any desired rhetoric concept.

6.3.5 Effectiveness of Cauchy Distributions

Some cases saw sudden increase in deviation of Cauchy values, even the number of documents considered were small in order. This indicates that even with small number of documents, it will be possible to judge the inconsistency through Cauchy distribution.

6.4 SUGGESTIONS FOR FUTURE EXTENSION OF RESEARCH

This research work (thesis) deals with rhetoric concept extraction with the help of fixed number of single words, terms and phrases taken from two educational models’ and the respective taxonomies. The verification of words/phrases in a domain concept document was done with the help of database of rhetoric concept words. The novelty of the
research is the application of Cauchy distribution for reliability and efficiency (probability) of extractions. An intelligent trainer based system may be tried out with varying synonym words and the accumulation of such terms/phrases by automatic searches that may be made on similar rhetoric concepts through an agent driven s/w. Such dynamically updatable databases could be tried out with repeated attempts on concept extractions and such a technique could be determined for optimizing both the parameters namely volumes of databases and the speed of extractions. Such a framework could be proposed and researched upon as a future extension of this work.