**PROJECT SURVEY**

**BACKGROUND AND MOTIVATION**

Document clustering has been investigated for use in a number of different areas of text mining and information retrieval. Initially, document clustering was investigated for improving the precision or recall in information retrieval systems, and as an efficient way of finding the nearest neighbours of a document. More recently, clustering has been proposed for use in browsing a collection of documents or in organizing the results returned by a search engine in response to a user’s query.

Earlier keywords were only the criteria through which documents closeness were measured which was again a very inefficient method for clustering.

Hierarchical based clustering technique is easy to implement technique with the nominal complexity which further reduced to quadratic complexity by also considering the most frequent words by the means of Tf( Term – frequency ) and Idf ( Inverse – Document-Frequency ).

**Table of Contents**

* Introduction
* Project Survey
* Technology Overview
* Proposed Methodology
* Flowchart
* Applied Algorithm Details
* Experimental Results
* Code
* Conclusion
* Bibliography