This thesis discusses the extraction of the hidden information available in documents and its presentation in a visual form to the user for effective corpus analysis, understanding, document search and for finding out relationships. Information extraction and document visualization are combined to show only the relevant contents of the document as the main dimensions in a visual form. In order to achieve the above objectives in this work, hidden features of the text documents such as topical terms, category information, plagiarized information, author identity information, ontology generation terms and summary sentences are extracted using various information extraction techniques. This extracted information is displayed to the user in visual forms applying visualization techniques to increase cognition.

Visualizations have to consider perception for designing effective visuals, since the visuals are to be perceived by the eye. Therefore, the visuals created here are based on a novel Gestalt Visual Perception (GVP) model, which defines four spaces and the mapping functions between the spaces. The spaces are considered for applying a transformation sequence to convert extracted document information into visual scenes and to perform cognitive tasks out of the generated visuals. The mapping functions between the spaces are based on Gestalt perception principles such as proximity, similarity, symmetry, continuity, closure, figure/ground, smallness/area, surroundedness.
and pragnanz. The visuals so generated help in easy analysis of documents and perform various visual tasks that illustrate summaries, patterns, relationships, comparisons, clusters and paths.

First, the key words or the content bearing terms are extracted from the documents. These terms, also called topical terms, are extracted along with the temporal terms using the Booksteins model. Later, the GVP model is applied over the extracted terms to perform mapping and model the topical and temporal term visualizations using wavelets. Next, the document category information, namely the multi-category and multiple level of categorization information are extracted from the documents using a novel set, rank and priority based algorithm and the extracted information is visualized applying thematic, temporal and animated visualization techniques. Later, plagiarism and author style information are extracted from the documents using detector cell and readability complexity measures. This is then modelled applying the GVP principles on terrain surfaces, bobbies and tree structures. The visualization techniques used in this case are animated three dimensional fractals. Next, the ontology related information is extracted from documents to visualize the user profile, user browsing pattern, personalized search results and ontological terms. The ontology so created is used for developing a personalized search and question answering system. This extracted information is then visualized as nested circles and tree maps.

Finally, single document and multi-document summarization is performed over the documents and the summary information is visualized as fractal patterns. An integrated visualization is also performed to visualize all
the above said features using a single visual. In addition, an evaluation strategy to judge the effectiveness and expressiveness of visualization is also designed. This thesis discusses extraction and visualization techniques by adapting methodologies from the Gestalt perception principles to build the framework of information extraction based document visualization.