UNIT I

Chapter 1 : Introduction 1-1 to 1-41

Syllabus : Basic Concepts of IR, Data Retrieval and Information Retrieval, text mining and IR relation, IR system block diagram.

Automatic Text Analysis : Luhn's ideas, Conflation Algorithm, Indexing and Index Term Weighing, Probabilistic Indexing Inverted file, Suffix trees and suffix arrays, Signature Files, Scatter storage or hash addressing, Clustered files, Hypertext and XML data structures.

UNIT II

Chapter 2 : Storage and Searching Techniques

2-1 to 2-36

Syllabus :

Retrieval strategies : Vector Space model, Probabilistic retrieval strategies, Language models, Inference networks, Extended Boolean retrieval, Latent semantic indexing, neural networks, Fuzzy set retrieval.

Retrieval utilities : Relevance feedback, Cluster Hypothesis. **Clustering algorithms :** Single Pass Algorithm, Single Link Algorithm.

Syllabus :

Searching the Web : Challenges, Characterizing the Web, Search Engines, Browsing, Mata-searchers, Web crawlers, Meta-crawler, Web data mining, Finding needle in the Haystack, Searching using Hyperlinks, Page ranking algorithms: Pagerank, Rank SVM.

Syllabus :

Semantic Search systems : G Semantic Web Google knowledge graphs, Ontology, Searching across ontologies, semantic web search.

Recommendation system : Collaborative Filtering and Content Based Recommendation of Documents and Products.

Information Extraction and Integration : Extracting Data from Text. Collecting and Integrating Specialized Information on the web.

❑❑❑