

**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.

1.1	Basic Concepts of IR	1-1
1.2	Data Retrieval and Information Retrieval	1-1
1.2.1	Text Mining and IR Relation.....	1-3
1.3	Information Retrieval System : Block Diagram	1-3
1.4	Automatic Text Analysis	1-5
1.5	Luhn's Ideas	1-5
1.6	Conflation Algorithm	1-6
1.7	Indexing and Index Term Weighting	1-8
1.7.1	Indexing.....	1-8
1.7.2	Index Term Weighting.....	1-8
1.8	Probabilistic Indexing.....	1-10
1.9	Automatic Classification.....	1-11
1.10	Measures of Association.....	1-13
1.11	Different Matching Coefficients	1-15
1.11.1	Simple Matching Coefficient	1-15
1.11.2	Dissimilarity Coefficients.....	1-17
1.12	Classification Methods.....	1-20
1.13	Dendrogram.....	1-23
1.14	Inverted Files.....	1-26
1.14.1	Introduction.....	1-26
1.14.2	Searching	1-27
1.14.3	Construction	1-28
1.15	Suffix Trees and Suffix Arrays	1-30
1.15.1	Introduction.....	1-30
1.15.2	Structure of Suffix Tree	1-31

1.15.3	Searching in Suffix Tree	1-33
1.15.4	Construction in Main Memory	1-33
1.15.5	Construction of Suffix Arrays for Large Texts.....	1-34
1.15.6	Difference between Suffix Array and Suffix Tree.....	1-35
1.16	Signature Files	1-36
1.16.1	Structure of Signature Files	1-37
1.16.2	Searching in Signature Files.....	1-38
1.16.3	Construction of Signature File.....	1-38
1.17	Scatter Storage or Hash Addressing.....	1-38
1.18	Clustered Files	1-39
1.19	Hypertext Data Structure	1-39
1.20	XML Data Structures	1-40

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.

2.1	Introduction	2-1
2.2	Information Retrieval Models	2-1
2.2.1	Basic Concepts	2-3
2.2.2	Boolean Model	2-3
2.2.3	Vector Model	2-5
2.2.3(A)	Vector Space.....	2-6
2.2.3(B)	Difference between Boolean and Vector Models	2-10
2.2.4	Probabilistic Model	2-11
2.3	Language Models	2-12
2.3.1	Types of Language Models	2-13
2.4	Inference Networks.....	2-13
2.4.1	Structure of the Inference Network	2-13



2.5	Extended Boolean Retrieval	2-16
2.5.1	Latent Semantic Indexing (LSI).....	2-16
2.5.2	Neural Networks	2-17
2.5.3	Fuzzy Set Retrieval	2-18
2.5.4	Difference between Neural Network-Based Retrieval and Fuzzy Set Retrieval Methods	2-19
2.6	Retrieval Utilities.....	2-20
2.6.1	Relevance Feedback.....	2-20
2.7	Cluster Hypothesis	2-20
2.7.1	Clustering in Information Retrieval	2-22
2.8	Clustering Algorithm	2-23
2.8.1	Definitions	2-23
2.9	Single-Pass Algorithm	2-24
2.10	Single Link Algorithm.....	2-31

UNIT III

Chapter 3 : Retrieval Performance Evaluation and Ontology 3-1 to 3-16

Syllabus :

Performance evaluation : Precision and recall, MRR, F-Score, NDCG, user oriented measures, cross fold evaluation.

Visualisation in Information System : Starting points, document context, User relevance judgement, Interface support for search process.

3.1	Performance Evaluation	3-1
3.1.1	Retrieval Performance Evaluation.....	3-2
3.2	Precision and Recall.....	3-2
3.3	Single Value Summaries	3-8
3.4	Alternative Measures.....	3-10
3.4.1	Harmonic Mean	3-11
3.4.2	E Measure.....	3-11
3.4.3	User - Oriented Measures	3-12
3.4.4	Cross Fold Evaluation.....	3-14
3.5	Visualization in Information System	3-14

UNIT IV

Chapter 4 : Distributed and Multimedia IR 4-1 to 4-40

Syllabus :

Distributed IR : Introduction, Collection Partitioning, Source Selection, Query Processing, web issues.

MULTIMEDIA IR : Introduction, Data Modeling, Query languages, Generic multimedia indexing approach, One dimensional time series, two dimensional color images, Automatic feature extraction.

4.1	Distributed Information Retrieval - Introduction	4-1
4.2	Collection Partitioning.....	4-2
4.3	Source Selection	4-3
4.4	Query Processing.....	4-4
4.5	Web Issues	4-6
4.6	Multimedia Information Retrieval - Introduction	4-6
4.6.1	Multimedia Information System vs. Traditional System	4-7
4.6.2	Data Modelling	4-7
4.6.3	Data Retrieval.....	4-7
4.7	Data Modelling in Multimedia Information Retrieval	4-9
4.7.1	Multimedia Data Support in Commercial DBMS.....	4-10
4.7.2	MULTOS Data Model	4-12
4.8	Techniques to Represent Audio and Visual Document	4-15
4.8.1	Format for Images	4-15
4.8.2	Format for Audio.....	4-16
4.8.3	Format for Video.....	4-16
4.9	Query Languages	4-16
4.9.1	Request Specification.....	4-17
4.9.2	Conditions on Multimedia Data	4-18
4.9.3	Uncertainty, Proximity and Weights in Query Expressions	4-20
4.9.4	Query Languages to Support Retrieval of Multimedia Objects	4-21
4.9.4(A)	SQL 3 Query Language.....	4-21
4.9.4(B)	MULTOS Query Language	4-22



4.10	Multimedia Information Retrieval : Indexing and Searching	4-25
4.10.1	Introduction.....	4-25
4.10.2	Spatial Access Methods	4-26
4.11	Generic Multimedia Indexing Approach	4-27
4.12	One-Dimensional Time Series	4-31
4.12.1	Distance Function.....	4-31
4.12.2	Feature Extraction and Lower - Bounding.....	4-31
4.12.3	Experiments	4-33
4.13	Two-Dimensional Colour Images.....	4-35
4.13.1	Image Features and Distance Function	4-36
4.13.2	Lower Bounding	4-36
4.13.3	Experiments using Bounding Theorem	4-38
4.14	Automatic Feature Extraction.....	4-39

UNIT V

Chapter 5 : Web Searching 5-1 to 5-36

Syllabus :

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

5.1	Introduction.....	5-1
5.2	Searching the Web	5-1
5.3	Challenges	5-3
5.4	Characterizing the Web	5-5
5.4.1	Measuring the Web	5-5
5.4.2	Modelling the Web.....	5-6
5.5	Search Engines	5-7
5.5.1	Challenges in Web Search Engine	5-8
5.5.2	Centralized Architecture	5-9
5.5.3	Distributed Architecture	5-11
5.5.4	User Interfaces	5-13
5.5.5	Ranking.....	5-16

5.5.6	Crawling the Web	5-17
5.5.7	Indices.....	5-17
5.6	Browsing	5-18
5.6.1	Web Directories	5-18
5.6.2	Combining Searching with Browsing and Some Helpful Tools	5-20
5.7	Meta Searchers	5-20
5.8	Web Crawlers.....	5-22
5.8.1	How it works?.....	5-23
5.8.2	Architecture of Web Crawler	5-23
5.8.3	Common Uses of Web Crawler	5-24
5.9	Meta-crawler.....	5-25
5.10	Web Data Mining.....	5-26
5.10.1	Three Perspectives on Data Mining.....	5-27
5.10.2	Uses of Web Data Mining	5-28
5.10.3	Web Data Mining Tools	5-28
5.11	Finding Needle in the Haystack	5-29
5.12	Searching using Hyperlinks	5-29
5.13	Page Ranking Algorithms	5-30
5.13.1	Page Rank Algorithm.....	5-31
5.13.1(A)	Importance of Page Ranking	5-33
5.13.2	Rank SVM.....	5-33
5.13.3	Ranking SVM Algorithm.....	5-34
5.13.4	Application of Ranking SVM	5-36

UNIT VI

Chapter 6 : Advanced Information Retrieval 6-1 to 6-30

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.



6.1	Semantic Search Systems.....6-1	6.7.2	Steps to Follow.....6-16
6.2	Google Semantic Web.....6-2	6.8	Recommendation System : Introduction6-18
6.3	Google Knowledge Graph6-3	6.9	Collaborative Filtering.....6-18
6.4	Taxonomy and Ontology.....6-3	6.9.1	Methodology.....6-19
6.4.1	Ontology.....6-5	6.9.2	Types of Collaborative Filtering6-19
6.4.2	Taxonomy6-5	6.9.3	Advantages of Collaborative Filtering6-21
6.4.3	Ontology Development from Taxonomy.....6-6	6.9.4	Disadvantages of Collaborative Filtering.....6-22
6.4.4	Ontology Extraction from Text6-8	6.10	Content Based Recommendation of Documents and Products6-22
6.5	Searching across Ontologies6-10	6.11	Information Extraction and Integration6-24
6.5.1	Content Explication.....6-11	6.11.1	Extracting Data from Text.....6-24
6.5.2	Query Model.....6-12	6.11.2	Semantic Web6-26
6.5.3	Verification6-12	6.11.3	Collecting and Integrating Specialized Information from Web6-28
6.6	Semantic Web Search.....6-12		
6.7	Ontology Creation.....6-15		
6.7.1	Methodology : General Ideas.....6-15		

