Date: 25-07-2012

1. Introductory class.
   Overall Idea of IR system

Date: 26-07-2012

2. Definition of IR
3. Structured, Unstructured & semi-Structured search
4. Boolean Retrieval
5. Term-Document incidence matrix
6. [web 2.0, Trending Topic Search-Twitter, Hash Tag-Manual Indexing]

Date: 27-07-2012

7. Inverted Index – Dictionary + Posting
8. Inverted Index construction
9. Query Processing, Hit List – Intersecting two posting list.
10. Query Optimization – Query Processing order

Date: 01-08-2012

11. Term, Token
12. Parsing a documents-
    a. Format
    b. Language
    c. Character set etc.
13. Normalization
14. Asymmetric expansion
15. Tokenization Problem- Hyphen, numbers, language specific characteristics
16. Equivalence Class
    a. Case Folding
    b. Stop Words
    c. Soundex

Date: 02-08-2012

    d. Thesauri
    e. Lemmatization
    f. Stemming
    g. [Spell correction]
17. Phrase queries
18. Rare Query

Date: 03-08-2012

19. Biword Index, Extended Biword.
20. Positional Index, Proximity Search
21. Combinational Scheme-Positional index+ popular phrases

22. Ziph's Law
23. Skip Pointers

Date: 08-08-2012

24. Intersecting posting lists using skip pointer, # of comparison
25. Dictionary
26. Data Structure- Hash & Tree
27. Merit & Demerit of these Data Structure
28. Wildcard Query, Tolerant Search

Date: 09-08-2012

29. K-gram index, Permuterm Index
30. Spell Correction – Isolated Word, Context Sensitive
31. Edit Distance, Levenshtein Distance
32. [Soundex]

Date: 10-08-2012

Tutorial Class on Posting, intersection, query processing order, some true/false question regarding precision recall

Date: 16-08 & 17-08

33. Problem with Boolean Retrieval system
34. Ranking
35. Jaccard Coefficient – Query-Doc matching score
36. Query: Bag of word model
37. tf-idf score
38. log frequency weight
39. Collection frequency, Document frequency
40. Vector Space model---Document vector, Query Vector
41. cosine similarity
42. Euclidean distance

Date: 22-08-2012

43. Selecting documents for considering ranking.
44. Champion List

Date: 23-08-2012

45. HUB score
46. Authoritative Score
47. Centrality

Date: 24-08-2012
48. Evaluation of Search Engine
49. Precision, Recall, Accuracy

Date: 29-08-2012
50. PR Curve
51. F-measure
52. MAP, MAP@K
53. R-precision
54. Interpolation of PR curve
55. NDCG@K

Date: 30-08-2012
56. Test Collection for IR evaluation-Benchmark
57. Information Need, Doc Collection, Human judged relevance.
58. Inter annotator agreement/ Disagreement.
59. Kappa Measure
60. Critique of pure relevance.

Date: 31-08-2012
61. Evaluation of Large Search Engine
62. A/B Testing
63. Relevance Measure:-
    a. Click Through First Result
    b. Mouse Position
64. Result Presentation
65. Result Summarization
    a. Static
    b. Dynamic
66. Quick Links
67. Relevance Feedback
68. Best Query for a set of Doc.
69. Rocchio Algorithm

Date: 05-09-2012
70. Relevance feedback-improves Recall
71. Query Drifting Problem
72. Pseudo Relevance Feedback
73. Query Expansion
    a. Retrieved result analysis
    b. Query Log analysis

Date: 06-09-2012
74. Thesaurus Generation
    a. Manual
    b. Automatic-Co-citation
75. [click stream mining]
76. Probabilistic IR model

77. Basic probability theory

Date: 07-09-2012
78. Probability Ranking Principle
79. Binary Independence Model [Detailed Derivation]
80. Ranking Query terms

Date: 12-09-2012
81. XML Retrieval
82. RDBMS, unstructured IR, XML
83. Document Object Model, DOM Tree,
84. xpath
85. Challenges-granularity, schema heterogeneity, query formulation etc
86. Context, Structural Term, Non-structural Term
87. Lexicalised Subtree.
88. Context Resemblance
89. Doc similarity Measure, Evaluation

Date: 13-09-2013
90. Tutorial Class#2 [Computing tf-idf score, Document vector computation, Document Ranking, MAP, R-Precision, Rocchio relevance feedback]

Date: 14-09-2012
91. Class Test#1

Date: 19/20/21
92. Language Model
93. Smoothing
94. Mixture Model
95. Vector Space Model

MID SEM

Date: 04-10-2012
96. Query Log Analysis [Class By Monojit Choudhury]

Date: 10-10-2012
97. Ranking--Classification
98. Rocchio Algorithm for text classification

Date: 11-10-2012
99. Naïve Bayes Classifier
100. Naïve Bayes Independence Assumption

Date: 12-10-2012

101. Naïve Bayes Model-Multivariate Bernoulli
102. K-NN Algorithm
103. Advantage & Disadvantage of K-NN

Date: 17-10-2012

104. Linear Classifier
105. Support Vector Machine
106. Functional Distance
107. Clustering vs Classification
108. Clustering Algorithm- Flat, Hierarchical
109. Dendrogram- Cut point
110. Distance- Euclidian, Manhattan, Cosine Similarity

Date: 31-10-2012

111. Clustering Concept- hard/Soft Clustering
112. K-Means Algorithm
113. Convergence of K-Means
114. Time Complexity of K-Means

Date: 01-11-2012

115. Enhancing speed of K-Means
116. Problem of K-Means – Outlier Detection
117. Evaluation of Clustering Algorithm
    a. Purity Measure
    b. NMI- Normalized Mutual Information
    c. Rand Index and F-measure

118.