AMCN-20 WEB SEARCH AND INFORMATION RETRIEVAL

UNIT-1 INTRODUCTION TO INFORMATION RETRIEVAL

1.1 Information retrieval problem, an inverted index, Processing Boolean queries ,
1.2 The extended Boolean model versus ranked retrieval , an inverted index,
1.3 Bi-word indexes, Positional indexes, Combination schemes

UNIT-2 INDEX CONSTRUCTION HARDWARE

2.1 Blocked sort-based indexing, Single-passin-memory indexing ,Distributed indexing,
2.2 Dynamic indexing, Other types of indexes Index compression:
2.3 Statistical properties of terms in information retrieval, Heap‘s law:
2.4 Estimating the number of terms, Zipf‘slaw: Modeling the distribution of terms, Dictionary compression, Dictionarys a string, Blocked storage, Postings file compression.

UNIT-3 SCORING

3.1 term weighting and the vector space model Parametric and zone indexes ,Weighted zone scoring, Learning weights ,The optimal weight,
3.2 Term frequency and weighting, Inverse document frequency, Tf-idf weighting, The vector space model for scoring, Variant tf-idffunctions.

UNIT-4 COMPUTING SCORES IN A COMPLETE SEARCH SYSTEM

4.1 Efficient scoring and ranking, In exact top K document retrieval, Index elimination, Champion lists, Static quality scores and ordering,
4.2 Impact ordering, Cluster pruning, Component so fan information retrieval system, Tiered indexes

UNIT-5 WEB SEARCH BASICS BACKGROUND AND HISTORY,

5.1 Web characteristics, The web graph, Spam, Advertising as the economic model,
5.2 The search user experience, User query needs Crawling, Crawler architecture,
5.3 DNS resolution, The URL frontier, Link analysis, The Web as a graph, Anchor text and the web graph ,Page Rank, Markov chains,
5.4 The Page Rank computation, Topic-specific Page Rank

UNIT-6 LANGUAGE MODELS FOR INFORMATION RETRIEVAL

6.1 Language models, Finite automata and language models, Types of language models, Multinomial distributions over words ,
6.2 The query likelihood model, Using query like lihood language models in IR, Estimating the query generation probability ,Language modeling versus other approaches in IR

References Book:

1. C. D. Manning, P. Raghavan and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008