



DEPARTMENT OF COMPUTER SCIENCE

COURSE OUTLINE

Unit Code & Name	CCS416: Information Retrieval
Prerequisite	Data Mining; Artificial Intelligence; Database Administration
Cohort	CCS Y4 S2, January 2024
Lecturer	Dr. Obuhuma James
Contact	jobuhuma@maseno.ac.ke / +254710 463 258

Purpose

The course is aimed at enabling the student to study of information retrieval and data mining techniques. It is about how to find relevant information and subsequently extract meaningful patterns out of it.

Learning Outcomes

By the end of the course, the student should be able to:

1. Explain the common algorithms and techniques for information retrieval.
2. Describe the techniques and algorithms existing in practical retrieval and data mining systems.
3. Discuss the challenges and existing techniques for the emerging topics of MapReduce, portfolio retrieval and online advertising.
4. Apply the quantitative evaluation methods for the IR systems and data mining techniques.

Course Content

Information retrieval systems. Representation and storage of information: Boolean Representation. Vector space model. Knowledge-based schemes. Genetic algorithms and association rules for document representation. Document Processing: Classification of document. Ranking and indexing schemes using probabilistic, fuzzy-set theoretic and knowledge-based techniques. Query-document similarity computation: Conventional matching schemes. Conceptual cohesiveness. Relevance feedback. Interactive query modification techniques. Performance evaluation of information retrieval systems. Application to Internet Search: Web mining and information filtering.

Delivery Methodology

Lectures, laboratory exercises, assignments and projects

Learning Resources

Books, Computers, Internet

Course Contents

Period	Topic	Outline
Week 1 – 2	1. Introduction	<ul style="list-style-type: none">▪ History of IR▪ Components of IR▪ IR Issues▪ Open-Source Search Engine Framework▪ The Impact of the Web on IR▪ The Role of AI in IR▪ IR Vs Web Search▪ Components of a Search Engine
Week 3 – 5	2. Boolean and Vector Space	<ul style="list-style-type: none">▪ Term Weighting

	Retrieval Models	<ul style="list-style-type: none"> ▪ Preprocessing ▪ Language Model Based IR ▪ Probabilistic IR
Week 6 – 8	3. Web Search Engine	<ul style="list-style-type: none"> ▪ Web Search Overview ▪ Web Search Architectures ▪ Crawling
Week 9 – 11	4. Web Search	<ul style="list-style-type: none"> ▪ Link Analysis ▪ Specialised Search
Week 10 – 13	5. Document Text Mining	<ul style="list-style-type: none"> ▪ Information Filtering, organization and relevant feedback ▪ Text Mining ▪ Text Classification ▪ Categorisation Algorithms: naive Bayes; decision trees; and nearest neighbor ▪ Clustering Algorithms: agglomerative clustering; k-means; expectation maximization (EM)

Course Assessment

Continuous Assessment Tests	30%
CATs	- 20%
Lab work	- 10%
End of Semester Examination	<u>70%</u>
	100%

Course Textbook

1. “Introduction to Information Retrieval,” Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Cambridge University Press, ISBN 0521865719
2. “Modern Information Retrieval,” Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison Wesley, ISBN 9780201398298
3. “Introduction to Data Mining,” Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Addison-Wesley, ISBN 978-0321321367
4. “Pattern Recognition and Machine Learning ,” Christopher M. Bishop, Springer, ISBN 978-0-387-31073-2

Reference Textbooks

Course Journals

1. IEE Transactions on Computers, IEEE Computer Society, ISSN: 00189340.
2. The SIAM Journal on Computing, Society of Industrial and Applied Mathematics, ISSN: 0097-5397.
3. Journal of Computer Science and Technology, Springer, ISSN: 1000-9000.

Online Resources