K.S.RANGASAMY COLLEGE OF ARTS AND SCIENCE (AUTONOMOUS)

TIRUCHENGODE – 637 215.

DEPARTMENT OF COMPUTER SCIENCE – DATA SCIENCE

ELECTIVE COURSES:

SEMESTER - V

1. **Social Media Mining**
2. **Natural Language Processing**

SEMESTER - VI

1. **Artificial Intelligence and Expert Systems**
2. **Information Retrieval Techniques**

Enclosures:

i. **Copy of Scheme of Examination**
ii. **Syllabus copy of the elective course**
iii. **Mapping of the courses.**

HoD – B.Sc [CS-DS]

HEAD,
Department of Computer Science,
K.S. Rangasamy College of Arts and
Science (Autonomous),
TIRUCHENGODE-637 215.

CONTROLLER OF EXAMINATIONS

CoE

Mr. M. PRASAD, M.Sc., M.B.A., M.Phil.
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu. India.

NON MAJOR ELECTIVE COURSE

The department offers the following two subjects as Non Major Elective Course for other than the computer science - data science students for third and fourth semesters.

S.No.	Semester	Subject Code	Subject
1	III	20UDSNM301	Internet Technology
2	IV	20UDSNM401	Principles of Web Design

ELECTIVE I

(Student shall select any one of the following subject as Elective in fifth semester)

S.No	Subject Code	Subject
1.	20UDSEL501	Social Media Mining
2.	20UDSEL502	Natural Language Processing

ELECTIVE II

(Student shall select any one of the following subject as Elective in sixth semester)

S.No	Subject Code	Subject
1.	20UDSEL601	Artificial Intelligence and Expert Systems
2.	20UDSEL602	Information Retrieval Techniques

K.S. RANGASAMY COLLEGE OF ARTS & SCIENCE (AUTONOMOUS) CONTROLLER OF EXAMINATIONS TIRUCHENGODE-637 215

Mr. M PRASAD, M.Sc., M.B.A., M.Phil.
Controller of Examinations
K.S.Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu. India.

Fifth Semester							
Part III							
20UDSM501	Core IX: Data Mining and Warehousing	6	3	25	75	100	5
20UDSM502	Core X: Software Engineering	5	3	25	75	100	5
20UDSM503	Core XI: Operating Systems	5	3	25	75	100	4
	Elective : I	5	3	25	75	100	4
20UDSMP501	Core Practical VI: R Programming	3	3	40	60	100	2
20UDSMP502	Core Practical VII: Computer Hardware	3	3	40	60	100	2
Part IV							
20UDSSBP501	SBC Practical III: MySQL(Internal Evaluation)	2	3	100	-	100	2
Part V							
20UDSE501	Extension Activity	-	-	-	-	-	2
Non Credit							
18ULS501	Career Competency Skills III	1	-	-	-	-	-
		30				**700**	**26**
Sixth Semester							
Part III							
20UDSM601	Core XII: Python Programming	6	3	25	75	100	5
20UDSM602	Core XIII: Computer Networks [Fifth Unit as Self- study]	6	3	25	75	100	4
	Elective II	6	3	25	75	100	4
20UDSMP601	Core Practical VIII: Python Programming	3	3	40	60	100	2
20UDSMP602	Core Practical IX: Computer Networking	3	3	40	60	100	2
20UDSPR601	Project Work	3	3	40	60	100	5

KSR KANGASAMY COLLEGE OF ARTS & SCIENCE (AUTONOMOUS) CONTROLLER OF EXAMINATIONS

Mr. M. PRASAD, M.Sc., M.B.A.,
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu. India.

20UDSEL501	ELECTIVE I: SOCIAL MEDIA MINING	SEMESTER - V

COURSE OBJECTIVES:
The Course aims to

- Understand how accurately analyze voluminous complex data set in social media and other sources.
- Understand the models and algorithms to process large data sets.
- Understand social behavior and recommendation challenges and methodologies.

Credit Points: 4 **Total Hours: 50**

UNIT	CONTENTS	Hrs	CO
I	Social Media Mining - Introduction – Atoms – Molecules – Interactions – Social Media mining Challenges - Graphs - Basics – Nodes – Edges – Degree of Distribution - Types - Directed – Undirected – Weighted - Graph Connectivity - Tress and Forests – Bipartite graphs - Complete Graphs – Sub graphs - Planar Graphs - Graph Representation - Graph Traversal Algorithms - Shortest path algorithms Dijkstra"s - Spanning tree algorithms – Prims - Bipartite matching - Ford-Fulkerson algorithm.	10	CO1
II	Network Models – Measures – Node : Eigen Centrality – Page Rank – Group Measures – Betweenness centrality - group degree centrality, centrality, and group - Closeness centrality - Node Linking Behavior - Transitivity and reciprocity - Linking Analysis - Cluster coefficient – Jaccard - Case Study: Modeling small networks with real world model.	10	CO2
III	Social media Communities – Social Communities – Member based Detection – Node degree – Node Similarity – Node reachability - Group Based detection methods - balanced –	10	CO3

K.S. RANGASAMY COLLEGE OF ARTS & SCIENCE (AUTONOMOUS) CONTROLLER OF EXAMINATIONS TIRUCHENGODE-637 215

Mr. M. PRASAD, M.Sc., M.B.A.,
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu. India.

	robust - modular – dense - hierarchical - Spectral Clustering: Balanced Community algorithm Community Evolution - Evaluation.		
IV	Social Network – Information Diffusion – Types - herd behavior - information cascades diffusion of innovation – epidemics – Diffusion Models Case Study – Herd Behavior – Information Cascades Methods – Social Similarity – assortativity – Social Forces - Influence homophily – Confounding - Assortativity measures – Influence measures – Predictive Models.	10	CO4
V	Recommendation Vs Search – Recommendation Challenges – Recommender algorithms - Content-Based Methods - Collaborative Filtering – Memory Based – Model Based – Social Media Recommendation – User friendship – Recommendation Evaluation – Precision – Recall –Behavioral – User Behavior – User – Community behavior – User Entity behavior – Behavioral Analytics – Methodology.	10	CO5

TEXT BOOKS:

1.	*Reza Zafarani, Mohhammad AbiElasi. 2014.* **Social Media Mining: An Introduction.** Cambridge press.

REFERENCE BOOKS:

1.	*Memon, N., Xu, J.J., Hicks, D.L., Chen, H. (Eds.).* **Data Mining for Social Network Data.**

WEB REFERENCES:

1.	http://dmml.asu.edu/smm/chapter
2.	http://learn.g2.com

Mr. M. PRASAD, M.Sc., M.B.A.,
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu. India

COURSE OUTCOMES (CO):

After completion of the course, the students will be able to:

CO1	Understand the concepts of Graph Models, social communities.
CO2	Understand the network models and measures to evaluate information.
CO3	Understand and apply algorithms to model data using graph and network structures and recommendations.
CO4	Apply algorithms on social data diffusion for various domains.
CO5	Distinguish and Suggest the appropriate algorithms for domain specific applications for data modeling.

MAPPING:

PSO/CO	PSO1	PSO2	PSO3	PSO4	PSO5
CO1	H	M	M	H	H
CO2	H	M	M	H	H
CO3	H	H	M	H	H
CO4	H	H	M	H	H
CO5	H	H	M	H	H

H-High; M-Medium; L-Low

K.S. RANGASAMY COLLEGE OF ARTS & SCIENCE (AUTONOMOUS) CONTROLLER EXAMINATIONS

Mr. M. PRASAD, M.Sc., M.B.A.,
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215, Tamilnadu, India.

20UDSEL502	ELECTIVE I: NATURAL LANGUAGE PROCESSING	SEMESTER - V

COURSE OBJECTIVES:
The Course aims to

- Make students understand syntactic and semantic elements of natural language processing.
- Conceive basics of knowledge representation and inference.

Credit Points: 4 **Total Hours: 50**

UNIT	CONTENTS	Hrs	CO
I	**INTRODUCTION:** Applications of NLP techniques and key issues - MT – grammar checkers – dictation - document generation - NL interfaces - Natural Language Processing key issues - The different analysis levels used for NLP: morpho-lexical - syntactic – semantic - pragmatic - markup (TEI, UNICODE) - finite state automata - Recursive and augmented transition networks – open problems.	10	CO1
II	**LEXICAL LEVEL:** Error-tolerant lexical processing (spelling error correction) - Transducers for the design of morphologic analyzers Features - Towards syntax: Part-of-speech tagging (Brill, HMM) – Efficient representations for linguistic resources (lexica, grammars,..) tries and finite – state automata.	10	CO2
III	**SYNTACTIC LEVEL:** Grammars (e.g. Formal/Chomsky hierarchy, DCGs, systemic, case, unification, stochastic) - Parsing (top-down, bottom-up, chart (Earley algorithm), CYK algorithm) - Automated estimation of probabilistic model parameters (inside-outside algorithm) - Data Oriented Parsing - Grammar formalisms and treebanks -	10	CO3

K.S. RANGASAMY COLLEGE OF ARTS & SCIENCE (AUTONOMOUS)

Mr. M. PRASAD, M.Sc., M.B.A.,
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu.

	Efficient parsing for context-free grammars (CFGs) - Statistical parsing and probabilistic CFGs (PCFGs) - Lexicalized PCFGs.		
IV	**SEMANTIC LEVEL:** Logical forms – Ambiguity resolution - Semantic networks and parsers - Procedural semantics - Montague semantics - Vector Space approaches - Distributional Semantics - Lexical semantics and Word Sense Disambiguation - Compositional semantics. Semantic Role Labeling and Semantic parsing. **PRAGMATIC LEVEL:** Knowledge representation – Reasoning - Plan/goal recognition - speech acts/intentions - belief models - discourse – reference.	**10**	**CO4**
V	**NATURAL LANGUAGE GENERATION:** content determination - sentence planning – surface realization. **SUBJECTIVITY AND SENTIMENT ANALYSIS:** Information extraction – Automatic summarization Information retrieval and Question answering - Named entity recognition and relation extraction - IE using sequence labeling - Machine translation: Basic issues in MT - Statistical translation - word alignment - phrase-based translation and synchronous grammars.	**10**	**CO5**
TEXT BOOKS:			
1.	*Daniel Jurafsky and James H. Martin.*2009. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition**. Prentice Hall.		
2.	*Ian H. Witten and Eibe Frank, Mark A. Hall.*2013. **Data Mining: Practical Machine Learning Tools and Techniques**, Morgan Kaufmann.		

K.S. RANGASAMY COLLEGE OF ARTS & SCIENCE (Autonomous) CONTROLLER OF EXAMINATIONS

Mr. M. PRASAD, M.Sc., M.....
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu. India.

REFERENCE BOOKS:	
1.	*Christopher Manning and Hinrich Schütze.* 2008. **Foundations of Statistical Natural Language Processing.** MIT Press.
2.	*James Allen.*1995.**Natural Language Understanding.** Addison Wesley, 1995.
3.	*Steven Bird, Ewan Klein, and Edward Loper.2009.* **Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit.** O'Reilly Media, Sebastopol.

COURSE OUTCOMES (CO):

After completion of the course, the students will be able to:

CO1	Understand the basics of natural language.
CO2	Acquire knowledge about lexical level.
CO3	Understand the concepts of syntactic level.
CO4	Know the concepts of semantic and pragmatic level.
CO5	Gain knowledge of sentiment analysis.

MAPPING:

PSO/CO	PSO1	PSO2	PSO3	PSO4	PSO5
CO1	H	M	M	H	H
CO2	H	M	M	H	H
CO3	H	H	M	H	H
CO4	H	H	M	H	H
CO5	H	H	M	H	H

H-High; M-Medium; L-Low

K.S. RANGASAMY COLLEGE OF ARTS & SCIENCE (Autonomous) CONTROLLER OF EXAMINATIONS 105

Mr. M. PRASAD, M.Sc., M.B.A., M.
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu. India

20UDSEL601	ELECTIVE II: ARTIFICIAL INTELLIGENCE AND EXPERT SYSTEMS	SEMESTER–VI

COURSE OBJECTIVES:

The Course aims to

- Demonstrate the knowledge of the building blocks of AI.
- Analyze and formalize the problems.

Credit Points: 4 **Total Hours: 50**

UNIT	CONTENTS	Hrs	CO
I	**Introduction to Artificial Intelligence:** Introduction - Brief History - Intelligent Systems: ELIZA - Categorization of Intelligent Systems - Components of AI Program - Foundations of AI - Sub-areas of AI - Applications. Tic-Tac-Toe Game Playing: Approach1 - Approach2 - Approach3 - Development of AI Languages - Current Trends in AI. **Problem Solving State-space Search and Control Strategies:** Introduction - General Problem Solving: Production System - State-Space Search - Control Strategies.	10	CO1
II	**Exhaustive Searches:** Breadth-First Search - Depth-First Search - Depth-First Iterative Deepening - Bidirectional Search - Analysis of Search methods. **Heuristic Search Techniques:** General-Purpose Heuristic-Branch and Bound Search - Hill Climbing - Beam Search - A* Algorithm - Optimal Solution by A* Algorithm - Monotonic Function.	10	CO2
III	**Expert System and Applications:** Introduction - Phases in Building Expert Systems: Knowledge Engineering - Knowledge Representation. Expert System Architecture: Knowledge Base - Inference Engine - Knowledge Acquisition - Case History -	10	CO3

	User Interfaces - Explanation Module - Special Interfaces. Expert Systems versus Traditional Systems: Characteristics of Expert Systems - Evaluation of Expert Systems - Advantages and disadvantages of Expert Systems - Languages for ES Development. **Rule Based Expert Systems:** Expert System Shell in Prolog - Problem-Independent Forward chaining - ES Shells and tools - MYCIN Expert System and various Shells - Applications of Expert Systems.		
IV	**Machine-Learning Paradigms:** Introduction - Machine-Learning Systems: Components of a Learning System - Rote Learning - Learning by Taking Advice - Learning by parameter Adjustment - Learning by Macro-Operators - Learning by Analogy. Supervised and Unsupervised Learnings: Neural Network Based Learning - Supervised Concept Learning - Probability Approximating Correct Learning - Unsupervised Learning - Reinforcement Learning.	**10**	**CO4**
V	**Artificial Neural Networks: Introduction** - Artificial Neural Networks: The Neuron Networks - The Neuron Model - Activation Functions - Neural Network Architectures. Single-Layer Feed - Forward Networks: Perceptron: Neuron Model - Learning Algorithm for Perceptron - Perceptron for OR Function: Example - Limitations of Perceptron. Multi-Layer Feed – Forward Networks: Back-Propagation Training Algorithm for FFNN - Weight Update Rule - Delta Rule (Least Mean Square) for Error Minimization.	**10**	**CO5**

TEXTBOOKS:	
1	*SarojKaushik.* 2014. **Artificial Intelligence.** [Sixth Edition]. Cengage Learning India Pvt. Ltd.
REFERENCE BOOKS:	
1	*Dan W.patterson.* 1992. **Introduction to Artificial Intelligence and Expert Systems.** Prentice Hall of India, New Delhi.
2	*SturatJ.Rusell and Peter Norvig.*2010. **Artificial Intelligence.** Prentice.
3	*Elaine Rich, Kevin Knight, B,Nair.*2010. **Artificial Intelligence: A Modern Approach.** [Third Edition]. Prentice Hall of India, New Delhi.
WEB REFERENCES:	
1	https://www.tutorialspoint.com
2	http://www.epub.uni-regensburg.de.pdf
3	http://www.investopedia.com
4	https://www.sas.com

COURSE OUTCOMES (CO):

After completion of the course, the students will be able to:

CO1	Acquire the basics of AI.
CO2	Analyze and formalize the problem as a state space, design and heuristics.
CO3	Attain the capability to various expert system methods.
CO4	Characterize machine learning algorithms as supervised, semi-supervised, and unsupervised.
CO5	Understand the concepts of Artificial Neural Networks.

KSR KANGASAMY COLLEGE OF ARTS & SCIENCE (AUTONOMOUS) CONTROLLER OF EXAMINATIONS TIRUCHENGODE

Mr. M. PRASAD, M.Sc., M.B.A., M.Phil.
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Namakkal Dt.

MAPPING:

PSO/CO	PSO1	PSO2	PSO3	PSO4	PSO5
CO1	M	H	H	M	H
CO2	H	H	H	H	H
CO3	M	H	M	H	H
CO4	H	H	H	H	H
CO5	H	H	H	H	H

H-High; M-Medium; L-Low

K.S. RANGASAMY COLLEGE OF ARTS & SCIENCE (AUTONOMOUS) CONTROLLER OF EXAMINATIONS

Mr. M. PRASAD, M.Sc., M.C.A., M.Phil.
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu.

20UDSEL602	ELECTIVE II: INFORMATION RETRIEVAL TECHNIQUES	SEMESTER - VI

COURSE OBJECTIVES:

The Course aims to

- Learn the information retrieval models.
- Familiarize with Web Search Engine.

Credit Points: 4 **Total Hours: 50**

UNIT	CONTENTS	Hrs	CO
I	**INTRODUCTION** Introduction - History of IR - Components of IR - Issues - Open source Search engine Frameworks - The impact of the web on IR - The role of artificial intelligence (AI) in IR – IR Versus Web Search - Components of a Search engine - Characterizing the web.	10	CO1
II	**INFORMATION RETRIEVAL** Boolean and vector - space retrieval models - Term weighting - TF-IDF weighting - cosine similarity - Preprocessing - Inverted indices - efficient processing with sparse vectors – Language Model based IR - Probabilistic IR – Latent Semantic Indexing - Relevance feedback and query expansion.	10	CO2
III	**WEB SEARCH ENGINE - INTRODUCTION AND CRAWLING** Web search overview, web structure, the user, paid placement, search engine optimization/ spam. Web size measurement - search engine optimization/spam – Web Search Architectures - crawling - meta-crawlers - Focused Crawling - web indexes – Near-duplicate detection - Index Compression - XML retrieval.	10	CO3
IV	**WEB SEARCH - LINK ANALYSIS AND SPECIALIZED SEARCH**		

	Link Analysis – hubs and authorities – Page Rank and HITS algorithms - Searching and Ranking – Relevance Scoring and ranking for Web – Similarity – Hadoop & Map Reduce - Evaluation - Personalized search - Collaborative filtering and content - based recommendation of documents and products – handling "invisible" Web - Snippet generation, Summarization, Question Answering, Cross - Lingual Retrieval.	10	CO4
V	**DOCUMENT TEXT MINING** Information filtering; organization and relevance feedback – Text Mining - Text classification and clustering - Categorization algorithms: naive Bayes; decision trees; and nearest neighbor - Clustering algorithms: agglomerative clustering; k-means; expectation maximization (EM).	10	CO5

TEXT BOOKS:

1.	*C. Manning, P. Raghavan, and H. Schütze,* **Introduction to Information** Retrieval, Cambridge University Press, 2008.
2.	*Ricardo Baeza -Yates and BerthierRibeiro* - Neto, Modern Information Retrieval: **The Concepts and Technology behind Search** 2 nd Edition, ACM Press Books 2011.
3.	*Bruce Croft, Donald Metzler and Trevor Strohman,* **Search Engines: Information Retrieval in Practice,** 1 st Edition Addison Wesley, 2009.
4.	*Mark Levene,* **An Introduction to Search Engines and Web Navigation**, 2nd Edition Wiley, 2010.

K.S. RANGASAMY COLLEGE OF ARTS & SCIENCE (AUTONOMOUS) CONTROLLER OF EXAMINATIONS TIRUCHENGODE-637 215.

Mr. M. PRASAD, M.Sc.,
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu, India

REFERENCE BOOKS:	
1.	*Stefan Buettcher, Charles L. A. Clarke, Gordon V. Cormack,* **Information Retrieval: Implementing and Evaluating Search Engines,** The MIT Press, 2010.
2.	*OphirFrieder* "**Information Retrieval: Algorithms and Heuristics: The Information Retrieval Series** ", 2 nd Edition, Springer, 2004.
3.	*Manu Konchady,* **"Building Search Applications: Lucene, Ling Pipe"**, and First Edition, Gate Mustru Publishing, 2008.
WEB REFERENCES:	
1.	https://www.springer.com
2.	https://www.tutorialspoint.com
3.	http://www.scholarpedia.org

COURSE OUTCOMES (CO):

After completion of the course, the students will be able to:

CO1	Differentiate between supervised, unsupervised, semi-supervised machine learning approaches
CO2	Discuss the decision tree algorithm and identity and overcome the problem of over fitting.
CO3	Discuss and apply the back propagation algorithm and genetic algorithms to various problems
CO4	Apply the Bayesian concepts to machine learning
CO5	Analyze and suggest appropriate machine learning approaches for various types of problems

K.S. RANGASAMY COLLEGE OF ARTS & SCIENCE (AUTONOMOUS) CONTROLLER OF EXAMINATIONS TIRUCHENGODE-637 215

Mr. M. PRASAD, M.Sc., M.B.A.,
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu. India.

MAPPING:

PSO/CO	PSO1	PSO2	PSO3	PSO4	PSO5
CO1	H	M	M	H	H
CO2	H	M	M	H	H
CO3	H	H	M	H	H
CO4	H	H	M	H	H
CO5	H	H	M	H	H

H-High; M-Medium; L-Low

K.S. RANGASAMY COLLEGE OF ARTS & SCIENCE (AUTONOMOUS) TIRUCHENGODE
CONTROLLER OF EXAMINATIONS

Mr. M. PRASAD, M.Sc., M.B.A., M.....
Controller of Examinations
K.S. Rangasamy College of Arts & Science (Autonomous)
Tiruchengode - 637 215. Tamilnadu. India.