Master of Technology - Data Science Engineering

Syllabus - First Semester

ADVANCED DATA STRUCTURE & ALGORITHMS

Course Code: DSE4101 **Credit Units: 03**

Course Objective:
The objective to this course is to equip students with advanced concepts of data structures like Huffman trees, Self-organizing trees, different types of heaps and their time complexity. Advanced topics and graphs and graph algorithms, geometric algorithms and parallel algorithms.

Course Contents:

Module-I
ADVANCED TREES: Definitions Operations on Weight Balanced Trees (Huffman Trees), 2-3 Trees and Red- Black Trees, Splay Tree.Augmenting Red-Black Trees to Dynamic Order Statistics and Interval Tree Applications. Operations on Disjoint sets and itsunion-find problem Implementing Sets. Dictionaries, Priority Queues and Concatenable Queues using 2-3 Trees.

Module-II
MERGEABLE HEAPS: Mergeable Heap Operations, Binomial Trees Implementing Binomial Heaps and itsOperations, 2-3-4. Trees and 2-3-4 Heaps. Amortization analysis and Potential Function of Fibonacci HeapImplementing Fibonacci Heap. SORTING NETWORK: Comparison network, zero-one principle, bitonic sorting andmerging network sorter.

Module-III
GRAPH THEORY DEFINITIONS: Definitions of Isomorphic Components. Circuits, Fundamental Circuits, Cut-sets. Cut-Vertices Planer and Dual graphs, Spanning Trees, Kuratovski's two Graphs.

Module-IV
GRAPH THEORY ALGORITHMS: Algorithms for Connectedness, Finding all Spanning Trees in a Weighted Graph andPlanarity Testing, Breadth First and Depth First Search, Topological Sort, Strongly Connected Components and ArticulationPoint. Single Min-Cut Max-Flow theorem of Network Flows. Ford-Fulkerson Max Flow Algorithms.

Module-V
Geometric algorithms: Point location, convex hulls and Voronoi diagrams, Arrangements. Parallelalgorithms: Basic techniques for sorting, searching, merging

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:
Text:
- RivestCormen, "Introduction to Algorithms";PHI

References:
- Tammasia, "Algorithm Design", Willey

ADVANCED DATABASE MANAGEMENT SYSTEMS

Course Code: DSE4102 **Credit Units: 03**

Course Objective:
The objective of this course is to expose the students to the implementation techniques of database system. This course explains techniques for query processing and optimization with transaction and concurrency control techniques

Course Contents:
Module I: Relational Databases
Integrity Constraints revisited, Extended ER diagram, Relational Algebra & Calculus, Functional, Muiltivalued and Join Dependency, Normal Forms, Rules about functional dependencies.

Module II: Query Processing and Optimization
Valuation of Relational Operations, Transformation of Relational Expressions, Indexing and Query Optimization, Limitations of Relational Data Model, Null Values and Partial Information.
Objected Oriented and Object Relational Databases
Modeling Complex Data Semantics, Specialization, Generalization, Aggregation and Association, Objects, Object Identity, Equality and Object Reference, Architecture of Object Oriented and Object Relational Databases

Module III: Parallel and Distributed Databases
Distributed Data Storage – Fragmentation & Replication, Location and Fragment
Transparency Distributed Query Processing and Optimization, Distributed Transaction Modeling and concurrency Control, Distributed Deadlock, Commit Protocols, Design of Parallel Databases, Parallel Query Evaluation.
Advanced Transaction Processing
Nested and Multilevel Transactions, Compensating Transactions and Saga, Long Duration Transactions, Weak Levels of Consistency, Transaction Work Flows, Transaction Processing Monitors.

Module IV
Multimedia databases,Databases on the Web and Semi–Structured Data. Case Study: Oracle Xi

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:
***Text*:**

- Elmarsi, Navathe, Somayajulu, Gupta, "Fundamentals of Database Systems", 4th Edition, Pearson Education, 2007
- Garcia, Ullman, Widom, "Database Systems, The complete book", Pearson Education, 2007
- R. Ramakrishnan, "Database Management Systems", McGraw Hill International Editions, 1998

***References*:**

- Date, Kannan, Swaminathan, "An Introduction to Database Systems", 8th Edition Pearson Education, 2007
- Singh S.K., "Database System Concepts, design and application", Pearson Education, 2006.
- Silberscatz, Korth, Sudarshan, "Database System Concepts", Mcgraw Hill, 6th Edition, 2006
- W. Kim, "Modern Database Systems", 1995, ACM Press, Addision – Wesley,
- D. Maier, "The Theory of Relational Databases", 1993, Computer Science Press, Rokville, Maryland
- Ullman, J. D., "Principals of database systems", Galgotia publications, 1999
- Oracle Xi Reference Manual
- Dietrich, and Urban, "An Advanced Course in Database Systems", Pearson, 2008.

ADVANCED COMPUTER NETWORKS

Course Code: DSE4104 **Credit Units: 03**

Course Objective:
The objective of the course is to provide thorough understanding & in-depth knowledge of concepts in computer networks Such as Internet protocols and routing, local area networks, wireless communications and networking, performance analysis, congestion control, TCP, network address translation, multimedia over IP, switching and routing, mobile IP, multicasting, IPv6. Peer-to-peer networking, network security, and other current research topics. A focus will be placed on wireless networking, reflecting rapid advances in this area. This course motivates the students to explore current research areas in the same field.

Course Contents:

Module I
Uses computer networks, Reference Models, TCP/IP suite of protocols, MAC protocols for high-speed LANS, MANs, and wireless LANs. (For example, FDDI, DQDB, HIPPI, Gigabit Ethernet, Wireless Ethernet, etc.)Fast access technologies. (For example, ADSL, Cable Modem, etc.)

Module II
Network Layer Design Issues, Routing Algorithms, Congestion Control Algorithms, Quality of Service, Internet Working, Network Layer in Internet.
IPv6 basic protocol, extensions and options, support for QoS, security, etc., Changes to other protocols, Application Programming Interface for IPv6.

Module III
Mobile IP, IP Multicasting. Multicast routing protocols, address assignments, session discovery, etc.

Module IV
The Transport Protocol: The Transport Service, Elements of transport protocol, a simple Transport Protocol, Internet Transport Protocols UDP, Internet Transport Protocols TCP, TCP extensions for high-speed networks, transaction-oriented applications Performance Issues.
The Application Layer: DNS-(Domain Name System), Electronic Mail, World Wide Web Multimedia.

Module V
Overview of network security, Secure-HTTP, SSL, ESP, Key distribution protocols. Digital signatures, digital certificates-mail Security, Web security, Social Issues.

Examination Scheme:

Components	A	CT	H	V/S/O	EE
Weightage (%)	5	10	7	8	70

Text & References:
Text:
- Computer Networks - Andrew S Tanenbaum, 4th Edition. Pearson Education/PHI
- Data Communications and Networking – Behrouz A. Forouzan. Third Edition TMH.

***References*:**
- Computer Communications and Networking Technologies –Michael A.Gallo, William M .Hancock - Thomson Publication.
- W. Stallings. Cryptography and Network Security: Principles and Practice, 2nd Edition, Prentice Hall, 1998.
- W. R. Stevens. TCP/IP Illustrated, Volume 1: The protocols, Addison Wesley, 1994.
- C. E. Perkins, B. Woolf, and S. R. Alpert. Mobile IP: Design Principles and Practices, Addison Wesley, 1997.

DIGITAL COMPUTER ORGANIZATION

Course Code: DSE4103 **Credit Units: 03**

Course Objective:
With increase in availability of system resources, concept of parallel architecture has obtained immense popularity. This course provides a comprehensive study of scalable and parallel computer architectures for achieving a proportional increase in performance with increasing system resources. In this course we have discussed the theory, technology, architecture (hardware) and software aspects of parallel computer and Vector computers.

Course Contents:
Module I: Parallel computer models
The state of computing, Multiprocessors and multicomputers, Multivector and SIMD computers, Architectural development tracks
Program and network properties: Conditions of parallelism, Data and resource dependences, Hardware and software parallelism, Program partitioning and scheduling, Grain size and latency, Program flow mechanisms, Control flow versus data flow, Data flow architecture, Demand driven mechanisms, Comparisons of flow mechanisms

Module II: System Interconnect Architectures
Network properties and routing, Static interconnection networks, Dynamic interconnection Networks, Multiprocessor system interconnects, Hierarchical bus systems, Crossbar switch and multiport memory, Multistage and combining network.

Module III: Processors and Memory Hierarchy
Advanced processor technology, Instruction-set Architectures,CISC Scalar Processors, RISC Scalar Processors, Superscalar Processors,VLIW Architectures, Vector and Symbolic processors
Memory Technology: Hierarchical memory technology, Inclusion, Coherence and Locality, Memory capacity planning, Virtual Memory Technology

Module IV: Backplane Bus System
Backplane bus specification, Addressing and timing protocols, Arbitration transaction and interrupt, Pipelining: Linear pipeline processor, Nonlinear pipeline processor, Instruction pipeline design, Mechanisms for instruction pipelining, Dynamic instruction scheduling, Branch handling techniques, Arithmetic Pipeline Design,Computerairthmetic principles.

Module V: Vector Processing Principles
Vector instruction types, Vector-access memory schemes.
Synchronous Parallel Processing: SIMD Architecture and Programming Principles, SIMD Parallel Algorithms

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:
Text:
- Kai Hwang, "Advanced computer architecture"; TMH, 2000.

References:
- J.P. Hayes, "computer Architecture and organization", MGH, 1998.
- M.J Flynn, "Computer Architecture, Pipelined and Parallel Processor Design", Narosa Publishing, 1998.
- D.A. Patterson, J.L. Hennessy, "Computer Architecture: A quantitative approach", Morgan Kauffmann, 2002.
- Hwang and Briggs, "Computer Architecture and Parallel Processing"; MGH,

ADVANCEDDATABASE MANAGEMENT SYSTEM LAB

Course Code: DSE4106 **Credit Units: 01**

Programs should be based on following topics:

Quick Review of Simple SQL Statements, SQL Built-in Functions,Primary Key, Foreign Key, Normalization,Joins View, Union. **Emphasis** on PL/SQL,Cursors 8. Exception handling,Procedure, Functions, Trigger, concurrency control, transaction processing. Introduction to SQLite.

Recommended Software: PostGreSQL, MySQL, Oracle.

Examination Scheme:

IA				EE	
A	**PR**	**LR**	**V**	**PR**	**V**
5	10	10	5	35	35

Note: IA –Internal Assessment, EE- External Exam, PR- Performance, LR – Lab Record, V – Viva.

ADVANCED COMPUTER NETWORK LAB

Course Code: DSE4107 **Credit Units: 01**

Course Contents:

1. Study of different types of networking cables, and implement cross and straight cable using clamping tool
2. Implementation of Stop and Wait Protocol and Sliding Window Protocol.
3. Study of Socket Programming and Client
4. Write a code simulating ARP /RARP protocols.
5. Write a code simulating PING and TRACEROUTE commands
6. Create a socket for HTTP for web page upload and download.
7. Write a program to implement RPC (Remote Procedure Call)
8. Implementation of Subnetting
9. Applications using TCP Sockets like Echo client and echo server, Chat Server, File Transfer, Applications using TCP and UDP Sockets ,DNS,SNMP
10. Study of Network simulator (NS).and Simulation of Congestion Control Algorithms using NS.
11. Perform a case study about the different routing algorithms to select the network path with its optimum and economical during data transfer Link State routing, Flooding, Distance vector.

Recommend Software:C/C++ on Linux/Unix, NS, Packet Tracer

Examination Scheme:

IA				EE	
A	**PR**	**LR**	**V**	**PR**	**V**
5	10	10	5	35	35

Note: IA –Internal Assessment, EE- External Exam, PR- Performance, LR – Lab Record, V – Viva.

ADVANCED DATA STRUCTURE & ALGORITHMS LAB

Course Code: DSE4105 **Credit Units: 01**

Programs based on Implementation of Graphs using Adjacency Matrix, Linked List , implementation of graph algorithms like BFS,DFS, Minimum Spanning Tree, Binary Search Tree, Knapsack Problem using Greedy Algorithm, Dynamic Programming, Shortest Path Algo (Dijkstra's), Implementing B-Tree,AVL Tree ,Red Black Tree. Implementing Sets, Dictionaries, Priority Queue using Heap.

Recommended Software: Java/C++/C/Python

Examination Scheme:

IA				EE	
A	**PR**	**LR**	**V**	**PR**	**V**
5	10	10	5	35	35

Note: IA –Internal Assessment, EE- External Exam, PR- Performance, LR – Lab Record, V – Viva.

MATLAB

Course Code: DSE4108 **Credit Units: 02**

Understanding The MATLAB Environment, Using the Help System in MATLAB, MATLAB Basics,Linear Algebra; Vectors and Matrices and various operations on them, M files; Scripts and User-defined functions, Plotting, Flow Control and Loops; For and While Loops, If and Case statements, structures, writing basic programs using the above, study of various toolboxes available in matlab and case study of any one tool box.

Recommended Software: MATLAB/Octave

Examination Scheme:

IA				EE	
A	**PR**	**LR**	**V**	**PR**	**V**
5	10	10	5	35	35

Note: IA –Internal Assessment, EE- External Exam, PR- Performance, LR – Lab Record, V – Viva.

PYTHON PROGRAMMING LAB

Course Code: DSE4109 **Credit Units: 01**

Course Contents:

1. Setting up python on Windows/Linux/Mac
2. First program in python
3. Programs related to basic input/ouput.
4. Programs related to variables,strings,numbers
5. Programs related to Lists and Tuples
6. Programs related to.Functions
7. Programs related to If Statements
8. Programs related to While Loops and Input
9. Programs related to Basic Terminal Apps
10. Programs related to Dictionaries
11. Programs related to Classes
12. Programs related to Exceptions
13. Programs related to GUI programming
14. Using Word, Excel, PDF files in python.
15. Web programming in python,
16. Case study of application areas of python.

Examination Scheme:

IA				EE	
A	**PR**	**LR**	**V**	**PR**	**V**
5	10	10	5	35	35

Note: IA –Internal Assessment, EE- External Exam, PR- Performance, LR – Lab Record, V – Viva.

Syllabus - Second Semester

APPLIED STATISTICAL ANALYSIS

Course Code: DSE4201 **Credit Units: 03**

Course Objective: This course is an introduction to statistics, the science of collecting, organizing, and interpreting numerical data. Emphasis will be placed on statistical reasoning and data analysis. The main topics are study design, descriptive statistics, probability theory, and statistical inference. Hands-on data analysis projects will be assigned. Students will be required to use a computer software package to solve various statistical problems.

Course Contents:

Module-I: Introduction to Statistical Analysis
Introduction to Statics, Population Versus Sample, Population and Sample, Basic Terms, Types Of Variables, Quantitative Variables, Qualitative or Categorical Variables, Measurement Scales, Identifying The Scale of Measurement, Normal Variables, Ordinal Variables, Continuous Variables, Cross-Section Data, Time –Series Data, Sources of Data, Types of Statics, Measures of Central Tendency, Mean, Examples, Mode, Mode Scenarios, Symmetry, Skewness, Measures of Spread, Range, Variance and Standard Deviation, Examples, Solution, Population Parameters and Sample Statistics, Measures of Position–Quartiles and Interquartile Range, Quartiles, Example, Solution, Percentiles and Percentile Rank, Example, Solution, Box and Whisker Plot, Solution.

Module-II: Describing Data
Raw Data, Graphical Presentation of Qualitative Data, Graphical Presentation of Qualitative Data, Graphing Quantitative Data, Frequency Distributions, Relative Frequency and Percentage Distributions, Graphing Grouped Data, Cumulative Frequency Distributions, Probability Concepts, Simple and Compound Events, Two Properties of Probability, Classical Probability, Probability Concepts, Complementary Events, Example, Discrete Random Variables, The Binomial Experiment, The Poisson Probability Distribution, Continuous Random Variables, Normal Distribution, Standard Normal Distribution.

Module-III: Testing Hypothesis
Population Distribution, Sampling and Non Sampling Errors, A Point Estimate, Interval Estimation, The t Distribution Hypothesis Testing, The Chi-Square Distribution.

Module-IV: Examining Relationship
Covariance, Pearson Correction Coefficient, Computational Formulas- Covariance, Computing A Correlation, Correlation Analysis, Scatter Plots, Relationships Between Continuous Variables, Correlation, Pearson Correlation Coefficient, Hypothesis Test For A Correlation, Extreme Data Values, Correlation Matrix, Anova Overview, Anova Hypothesis, Anova, Error Sum of Squares, Model Sum of Squares, Anova-Example, Simple Regression, Relationship Between Food Expenditure And Income: (A) Linear Relationship (B) Nonlinear Relationship, Plotting A Linear Equation, Y-Intercept And Slope Of A Line, Simple Linear Regression Analysis, Scatter Diagram, Least Squares Line, Error Sum Of Squares, The Least Squares Line, Example, Solution, Error Of Prediction, Positive And Negative Linear Relationship Between X And Y, Assumptions of Regression Model, Coefficient Of Determination, Regression Analysis.

Module-V: Advanced Techniques
Non Parametric Tests, Chi-Squared Goodness- of-Fit Test, Chi-Square Test of Independence, The Sign Test, Example, Mann-Whitney U Test, The Kruskal-Wallis Test, No Title, Structural Equation Modelling, Cluster Analysis, Factor Analysis, Centroid Method, Principal Components Method.

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Spectrum Editorial Board, Statistical Analysis Graphs and Diagrams, Spectrum Books Pvt Ltd, 2013.
- TorstenHothorn and Brian S. Everitt , "A Handbook of Statistical Analyses Using R", Chapman and Hall/CRC, 2006.
- Pal and Sarkar, "Statistics: Concepts and Applications", Prentice Hall India Learning Private Limited, 2007.

DATA MINING AND PREDICTIVE ANALYTICS

Course Code: DSE4202 **Credit Units: 03**

Course Objective: This course introduces the topics of Data Mining, and Data Analytics by providing a basic, practical foundation that allows the students to participate in Data Analytics projects. The course incorporates an introduction to the Data Analytics lifecycle, Machine Learning (ML), Data Mining algorithms and computational paradigms that allow computers to find patterns and regularities in databases, perform prediction and forecasting, and generally improve their performance through interaction with data.

Course Contents:

Module-I: Data Preparation
An introduction to data mining and predictive analytics, Data Pre-processing Exploratory Data Analysis, Dimension- Reduction Methods.

Module-II: Classification
k- Nearest Neighbour Algorithm, Decision Trees, Model Evaluation Technique, Cost –Benefit Analysis Using Data Driven Cost.

Module-III: Clustering
Hierarchical and k- Means Clustering, Kohenen Networks, Birch Clustering, Measuring Cluster Goodness.

Module-IV: Association Rules
Affinity Analysis and Market Basket Analysis, Data Representation for Market Basket Analysis, Support, Confidence, Frequent Item sets, and the a Priori Property, How Does a Priori Algorithm Work ? Generating Frequent Item sets, Generating Association Rules, Extension from Flag Data to General Categorical Data, Information-Theoretic Approach: Generalized Rule Induction Method, *J*-Measure, Association Rules are Easy to do Badly, How can we Measure the Usefulness of Association Rules? Do Association Rules Represent Supervised or Unsupervised Learning?

Module-V: Case Study: Predicting Response to Direct Mail Marketing
Business Understanding, Data Preparation, and Eda, Clustering and Principal Components Analysis, Modelling and Evaluation for Performance and Interpretability, Modelling and Evaluation for High Performance.

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Daniel T. Larose, Chantal D. Larose, "Data Mining and Predictive Analytics", John Wiley & Sons, Inc., Hoboken, New Jersey, 2015.
- Thomas W. Miller, "Modelling Techniques in Predictive Analytics", Pearson FT Press, 2013.
- Markus Hofmann, Ralf Klinkenberg, "Rapid-Miner: Data Mining Use Cases and Business Analytics Applications", Chapman and Hall/CRC, 2016

DATA WAREHOUSING AND MULTIDIMENSIONS MODELLING

Course Code: DSE4203 **Credit Units: 03**

Course Objective: This course focuses on the fundamentals of data warehousing and multidimensional Modelling. Data warehouse development life cycle, Data warehouse analysis, CUBE, ROLL UP and STAR queries, Data Warehouse Design - Massive de -normalisation, STAR schema design, Data ware house Architecture, OLAP, ROLAP and MOLAP, concepts of Fact and dimension table are the major areas of coverage of this course. This course also deals with the issues while implementing the multidimensional models

Course Contents:

Module I Introduction
Multidimensional Data Management, Multidimensional History, Related Terminology,

Module II Fundamental Concepts
Cubes ,Dimensions, Facts, Measures, Relational Representations, Star Schemas, Snowflake Schemas, Data Warehouses And Data Marts, Multidimensional Modelling Process, Analysis And Querying ,Roll Up, Drill Down, Drill Out, Slicing And Dicing, Drill Across, Pivot Tables, Ranking, Multi-Dimensional Querying in MDX and SQL, Graphical Querying and Visualizations .

Module III Advance Concepts
Slowly Changing Dimensions, The Problem, Solutions, Other Special Kinds Of Dimensions, Mini dimensions, Outriggers, Degenerate Dimensions, Junk Dimensions, Time Dimensions, Data Quality Dimensions, Advanced Hierarchies, Parent-Child Hierarchies, Unbalanced Hierarchies, Non Covering Hierarchies , Non –Strict Hierarchies, Multiple Hierarchies And Parallel Hierarchies.

Module IV Implementation Issues
Materialized Views, Indexing, Indexing Overview, Bitmap Indices, Join Indices, Query Processing, OLAP Implementations, Extract-Transform-Load.

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Christian S. Jensen, Christian Thomsen, and Professor Torben Pedersen, “Multidimensional Databases and Data Warehousing”, Morgan & Claypool Publisher, 2010.
- Ralph Kimball, Margy Ross, "The Data Warehouse Toolkit: The Definitive Guide”, 3rd Edition, John Wiley & Sons, 2013.
- Len Silverston, Paul Agnew, “The Data Model Resource Book: Volume 3: Universal Patterns for Data Modeling”, John Wiley & Sons., 2009.

DATABASE AND KNOWLEDGE BASE SYSTEMS

Course Code: DSE4204 **Credit Units: 03**

Course Objective: This course discusses design methodology for databases to verify their structural correctness and implements databases. It also provides applications software primarily in the relational model using querying languages, primarily SQL, and other database supporting software applying the theory behind various database models and query languages implementing security and integrity policies relating to databases and preparation for data analytics working in group settings to design and implementing database projects.

Course Contents:

Module-I: Introduction

Database Languages, Object-Base Systems, Knowledge-base Systems, History and Perspective, Data Models for Database Systems: Data Models, The Entity-relationship Model, The Relational Data Model, Operations in the Relational Data Model, The Network Data Model, The Hierarchical Data Model, An Object-Oriented Model, Logic as a Data Model: The Datalog Data Model, Evaluating Non- recursive Rules, Computing the Meaning of Recursive Rules, Incremental Evaluation of Least Fixed Points, Negations in Rule Bodies, Relational Algebra and Logic, Relational Calculus, Tuple Relational Calculus.

Module-II:Relational Query and Object-Oriented Database Language

ISBL: A "Pure" Relational Algebra Language, QUEL: A Tuple Relational Calculus Language, Query-by-Example: A DRC Language, Data Definition in QBE, The Query Language SQL, Data Definition in SQL, The DBTG Data Definition language, The DBTG Query Language, The DBTG Database Modification Commands, Data Definition in IMS, A Hierarchical Data Manipulation Language, Data Definition in OPAL, Data Manipulation in OPAL

Module-III:Physical Data Organization and Design of Relational Databases

The Physical Data Model, The Heap Organization, Hashed Files, Indexed Files, B-trees, Files with a Dense Index, Secondary Indices, Data Structures in DBTG Databases, Data Structures for Hierarchies, Data Structures for Relations, A Search Tree Structure, Functional Dependencies, Lossless-Join Decomposition, Normalization, Generalized Dependencies.

Module-IV:Transaction Management

Basic Concepts, A Simple Transaction Model, The Two-phase Locking Protocol, a Model with Read and write-Locks, Lock Modes, A Read-Only, Write-Only Model, Concurrency for Hierarchically Structured Items, Handling Transaction Failures, Aggressive and Conservative Protocols, Recovery From Crashes, Timestamp-based Concurrency Control.

Module-V: Distributed Database Management

Distributed Databases, Distributed Locking, Distributed Two-phase Locking, Distributed Commitment, A Nonblocking Commit Protocol, Timestamp-based, Distributed Concurrency, Recovery of Nodes, Distributed Deadlocks.

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Jeffrey D. Ullman"Principles of Database and Knowledge-Base Systems", Vol. 1, Computer Science Press, USA, 1988.
- AviSilberschatz, Henry F. Korthand S. Sudarshan, "Database System Concepts", Mcgraw Hill Education, 2000.
- Ngoc Thanh Nguyen, Edward Szczerbicki, "Intelligent Systems for Knowledge Management", Springer-verlagGmbh, 2009.

BIG DATA TECHNOLOGIES

Course Code: DSE4205 **Credit Units: 03**

Course Objective:
This course brings together several key big data technologies used for storage, analysis and manipulation of data. It also introduces the students the key concepts of Hadoop framework, MapReduce, Pig, Hive, and No-SQL. Students will prepare a sample project in Hadoop.

Course Contents:

Module I: Introduction to Big Data
Big Data and its Importance – Four V's of Big Data – Drivers for Big Data – Introduction to Big Data Analytics – Big Data Analytics applications.

Module II: Big Data Technologies
Hadoop's Parallel World – Data discovery – Open source technology for Big Data Analytics – cloud and Big Data –Predictive Analytics – Mobile Business Intelligence and Big Data – Crowd Sourcing Analytics – Inter- and Trans-Firewall Analytics - Information Management.

Module III: Processing Big Data
Integrating disparate data stores - Mapping data to the programming framework - Connecting and extracting data from storage - Transforming data for processing - Subdividing data in preparation for Hadoop Map Reduce.

Module IV: Hadoop Map Reduce
Employing Hadoop Map Reduce - Creating the components of Hadoop Map Reduce jobs - Distributing data processing across server farms –Executing Hadoop Map Reduce jobs - Monitoring the progress of job flows - The Building Blocks of Hadoop Map Reduce - Distinguishing Hadoop daemons - Investigating the Hadoop Distributed File System Selecting appropriate execution modes: local, pseudo-distributed, fully distributed.

Module V: Big Data Tools and Techniques
Installing and Running Pig – Comparison with Databases – Pig Latin – User- Define Functions – Data Processing Operators – Installing and Running Hive – Hive QL – Tables – Querying Data – User-Defined Functions – Oracle Big Data.

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Michael Minelli, Michehe Chambers, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business", 1st Edition, AmbigaDhiraj, Wiely CIO Series, 2013.
- Tom White, "Hadoop: The Definitive Guide", 3rd Edition, O'reilly, 2012.
- Arvind Sathi, "Big Data Analytics: Disruptive Technologies for Changing the Game", 1st Edition, IBM Corporation, 2012.
- Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", 1st Edition, Wiley and SAS Business Series, 2012.

RESEARCH METHODOLOGY AND TECHNICAL REPORT WRITING

Course Code: DSE4206 **Credit Units: 02**

Course Objectives: The course will enhance scientific , technical and research writing skills and impartknowledge about various stages of research process, statistical analysis, statistical tests and their applications in statistical decision making.

Course Contents:

Module I: Introduction to research: Definition, motivation, need, objectives, significance and characteristics of research; types of research; steps in research process; planning a research proposal; literature review, web searching.

Module II: Population and sample, parameter and statistic, sampling and data collection, sampling design: steps, types, sample size, sampling methods, large and small samples, primary and secondary data, data processing and analysis. Sample surveys and questionnaire designing, scaling techniques.

Module III: Dependent and independent variables, univariate, bivariate and multivariate analysis, means-arithmetic, geometric and harmonic; measure of dispersion of data, standard deviation, variance, coefficient of variation and degree of freedom. Hypothesis testing: null hypothesis and alternate hypothesis, errors in hypothesis testing, significance and confidence levels, parametric tests and non-parametric tests, one-tailed and two-tailed tests, analysis of variance. Regression analysis and curve fitting, method of least-squares, explained and unexplained variations, coefficient of correlation, coefficient of determination**.**

Module IV: Technical/scientific/research report writing: structure and components of scientific reports, formats of dissertations, research report, report writing skills, report preparation, referencing , bibliography and footnotes. Making presentation-use of visual aids and PPTs. Publication of research papers, citations,. Intellectual property rights and copy rights, plagiarism, patents and patent laws, commercialization and ethical issues.

Examination Scheme:

Attendance	Assignment/Library consultation / Thesis writing	Class test	Final Exam	Total
5	15	10	70	100

Text Books:

- Blake, G. and Bly, R.W. 1993, The Elements of Technical Writing. MacMillan, New York
- Booth, V. 1981. Writing a Scientific Paper and Speaking at Scientific Meetings. The Biochemical Society, London
- Chawla,D and Sondhi, N. 2016, Research Methodology- Concepts and Cases. Vikas Publishing House Pvt Ltd. New Delhi
- Kothari, C.R.2008. Research Methodology- Methods and Techniques, 2nd.ed. New Age International Publishers, New Delhi.

Reference Books:

- Geode, Millian J.& Paul K. Hatl, Methods in Research, McGraw Hills, New Delhi.
- Montomery, Douglas C.(2007), 5th Ed. Design and Analysis of Experiments, Wiley India.
- Panneerselvam, R.2009. Research Methodology, PHI Learning Pvt.Ltd., New Delhi-110001
- Ranjit Kumar 2009. Research Methodology- A step –by- step Guide for beginners; 2nd ed. Dorling Kindersley (India) Pvt. Ltd. Patpargang, Delhi- 110092

APPLIED STATISTICAL ANALYSIS LAB

Course Code: DSE4207 **Credit Units: 01**

Programs should be based on following topics:

Quick Reviewof interpretation of mathematical models such as formulas, graphs, tables, and schematics, and draw inferences from them;
Representation of mathematical information symbolically, visually, numerically, and verbally;
Use of arithmetic, algebraic, geometric, and statistical methods to solve problems;
Estimation and check answers to mathematical problems in order to determine reasonableness,
Identification of alternatives, and select optimal results; and recognize the limitations of mathematical and statistical models.

Recommended Software: SPSS (IBM).

Examination Scheme:

IA				EE	
A	**PR**	**LR**	**V**	**PR**	**V**
5	10	10	5	35	35

Note: IA –Internal Assessment, EE- External Exam, PR- Performance, LR – Lab Record, V – Viva.

DATA MINING AND PREDICTIVE ANALYTICS LAB

Course Code: DSE4208 **Credit Units: 01**

Course Objective:

- To analyse the data using statistical methods.
- To understand and demonstrate data mining using any open source data mining tool.

Recommended Software: ORANGE, Rapid Miner

List of Experiments

1. Data Analysis- Getting to know the Data (Using ORANGE, Rapid Miner)
    - Parametric - Means, T-Test, Correlation
    - Prediction for numerical outcomes - Linear regression
    - Correlation analysis
    - Preparing data for analysis
    - Pre-processing techniques

2. Data Mining (Using ORANGE, Rapid Miner or any open source data mining tool)
    - Implement clustering algorithm
    - Implement classification using
    - Decision tree
    - Back propagation
    - Visualization methods.

Examination Scheme:

IA				EE	
A	**PR**	**LR**	**V**	**PR**	**V**
5	10	10	5	35	35

Note: IA –Internal Assessment, EE- External Exam, PR- Performance, LR – Lab Record, V – Viva.

DATA WAREHOUSING AND MULTIDIMENSIONAL MODELLING LAB

Course Code: DSE4209 **Credit Units: 01**

Programs should be based on following topics:
Quick Review SQL Statements, SQL Built-in Functions,**Emphasis** on PL/SQL,Cursors 8. Exception handling, Procedure, Functions,Trigger, concurrency control, transaction processing. Introduction to ETL Tools:Talend Open Source Data Integrator, Scriptella, KETL
Pentaho Data Integrator - Kettle, Jaspersoft ETL, GeoKettle, CloverETL, HPCC Systems

Recommended Software: SQL Server, ETL Tools (Open Source)

Examination Scheme:

IA				EE	
A	**PR**	**LR**	**V**	**PR**	**V**
5	10	10	5	35	35

Note: IA –Internal Assessment, EE- External Exam, PR- Performance, LR – Lab Record, V – Viva.

INFORMATION MANAGEMENT SYSTEM

Course Code: DSE4210 **Credit Units: 03**

Course Objective:
This course exposes students with the basics of managing the information and explores the various aspects of database design and modelling. It examines the basic issues in information governance and information integration and also helps to understand the information architecture

Course Contents:

Module-I:Database Modelling, Management and Development
Database design and modelling – Business Rules and Relationship; Java database Connectivity (JDBC), Database connection Manager, Stored Procedures. Trends in Big Data systems including NoSQL – Hadoop HDFS, MapReduce, Hive, and enhancements.

Module-II: Data Security and Privacy
Program Security, Malicious code and controls against threats; OS level protection; Security – Firewalls, Network Security Intrusion detection systems. Data Privacy principles. Data Privacy Laws and compliance.

Module-III: Information Governance
Master Data Management (MDM) – Overview, Need for MDM, Privacy, regulatory requirements and compliance. Data Governance – Synchronization and data quality management.

Module-IV:Information Architecture
Principles of Information architecture and framework, Organizing information, Navigation systems and Labelling systems, Conceptual design, Granularity of Content.

Module-V: Information Lifecycle Management
Data retention policies; Confidential and Sensitive data handling, lifecycle management costs. Archive data using Hadoop; Testing and delivering big data applications for performance and functionality; Challenges with data administration;

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Jeffrey A. Hoffer, HeikkiTopi, V Ramesh – MODERN DATABASE MANAGEMENT, 10 Edition, PEARSON, 2012.
- Alex Berson, Larry Dubov MASTER DATA MANAGEMENT AND DATA GOVERNANCE, 2/E, Tata McGraw Hill, 2011.
- Security in Computing, 4/E, Charles P. Pfleeger, Shari Lawrence Pfleeger, Prentice Hall; 2006
- Information Architecture for the World Wide Web; Peter Morville, Louis Rosenfeld ; O'Reilly Media; 1998.
- http://ibm.com/big-data – Four dimensions of big data and other ebooks on Big Data Analytics

INFORMATION SYSTEM SECURITY

Course Code: DSE4211 **Credit Units: 03**

Course Objective:
This course provides a deep and comprehensive study of the security principles and practices of information systems. Topics include basic information security concepts, common attacking techniques, common security policies, basic cryptographic tools, authentication, access control, software security, operating system security, and legal and ethical issues in information systems security

Course Contents:

Module-I: Introduction
Computer Security Concepts, Threats, Attacks, and Assets Security Functional Requirements A Security Architecture for Open Systems Computer Security Trends.

Module-II: COMPUTER SECURITY TECHNOLOGY AND PRINCIPLES
Cryptographic Tools, User Authentication, Access Control, Database Security, Malicious Software, Denial-of-Service Attacks.

Module-III: SOFTWARE SECURITY
Buffer Overflow: Stack Overflows, Defending Against Buffer Overflows, Other Forms of Overflow Attacks, Software Security: Software Security Issues, Handling Program Input, Writing Safe Program Code, Interacting with the Operating System and Other Programs, Handling Program Output

Module-IV: TRUSTED SYSTEMS SECURITY
Operating System Security: Introduction to Operating System Security, System Security Planning, Operating Systems Hardening, Application Security, Security Maintenance, Linux/Unix Security, Windows Security, Virtualization Security. Trusted Computing and Multilevel Security: Bell-LaPadula Model for Computer Security, Other Formal Models for Computer Security, Concept of Trusted Systems, Application of Multilevel Security, Trusted Computing and the Trusted Platform Module, Common Criteria for Information Technology Security Evaluation, Assurance and Evaluation

Module-V: MANAGEMENT ISSUES
IT Security Management and Risk Assessment, IT Security Controls, Plans, and Procedures, Physical and Infrastructure Security, Human Resources Security, Security Auditing, Legal and Ethical Aspects

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- W. Stallings, "Computer Security: Principles and Practice," 2st Edition, Prentice Hall, ISBN: 0132775069, 2011.
- M. Stamp, "Information Security: Principles and Practice," 2st Edition, Wiley, ISBN: 0470626399, 2011.
- M. E. Whitman and H. J. Mattord, "Principles of Information Security," 4st Edition, Course Technology, ISBN: 1111138214, 2011.
- M. Bishop, "Computer Security: Art and Science," Addison Wesley, ISBN: 0-201-44099-7, 2002.
- Elmsari and Navathe, "Fundamentals of Database Systems", 4th Ed., A. Wesley, 2004
- Ullman J. D., "Principles of Database Systems", 2nd Ed., Galgotia Publications, 1999.

Syllabus - Third Semester

DESCRIPTIVE ANALYSIS

Course Code: DSE4301 **Credit Units: 03**

Course Objective:
This course introduces some elementary statistical methods of analysis of data and compute various measurements of central tendency, dispersion, skewness and kurtosis. Also discusses computation of the correlation coefficient from ungrouped bivariate data and interpret them and analyse data pertaining to attributes and to interpret results.

Course Contents:

Module-I: Introduction to Statistics
Definitions: Webster's and Secrist's definition of Statistics, Importance of Statistics, Scope of Statistics: In the field of Industry, Biological Sciences, Medical Sciences, Economics Sciences, Social, Sciences, Management Sciences, Agriculture, Insurance, Actuarial Science, Education and Psychology.

Module-II: Population and Sample
Types of characteristics: Attributes: Nominal scale, ordinal scale. Variables: Interval scale, ratio scale, discrete and continuous variables, Types of data: Primary data, Secondary data, Notion of a statistical population: Finite population, infinite population, homogeneous population and heterogeneous population. Notion of sample, random sample and non-random sample, Methods of sampling: Simple random sampling with and without replacement (SRSWR and SRWOR) stratified random sampling, systematic sampling, cluster sampling and two-stage sampling.

Module-III: Presentation of Data
Classification: Raw data and its classification, Discrete frequency distribution, Sturge's rule, continuous frequency distribution, inclusive and exclusive methods of classification, Open end classes, cumulative frequency distribution and relative frequency distribution, Graphical Presentation of Data: Histogram, frequency curve, frequency polygon, ogive curves, stem and leaf chart, Check sheet, Parato diagram, Examples and Problems.

Module-IV: Measures of Central Tendency
Concept of central tendency of statistical data, Arithmetic Mean (A.M.), combined mean of a number of groups, merits and demerits, Geometric Mean (G.M.), Harmonic Mean (H.M.), Weighted Mean, Weighted A.M., G.M. and H.M. , Mode, Median, Empirical relation between mean, median and mode, Order relation between arithmetic mean, geometric mean, harmonic mean.

Module-V: Measures of Dispersion
Concept of dispersion, characteristics of good measure of dispersion, Range, Mean deviation, Mean square definition, Variance and standard deviation, Combined variance, Combined standard deviation, generalization for n groups, Measures of dispersion for comparison: coefficient of range, coefficient of quartile deviation and coefficient of mean deviation, coefficient of variation.

Module-VI: Skewness and Kurtosis
Concept of skewness of frequency distribution, positive skewness, negative skewness, symmetric frequency distribution, Bowley's coefficient of skewness, interpretation using Box plot, Karl Pearson's coefficient of skewness, Measures of skewness basked on moments ($\beta 1$, $\gamma 1$), Concepts of kurtosis, leptokurtic, mesokurtic and platykurtic frequency distributions, Measures of kurtosis based on moments, ($\beta 2$, $\gamma 2$).

Module-VII: Correlation & Regression

Bivariate data, bivariate frequency distribution, Concept of correlation between two variables, positive correlation, negative correlation, Scatter diagram, conclusion about the type of correlation from scatter diagram, Covariance between two variables, Karl Pearson's coefficient of correlation, Spearman's rank correlation coefficient, In case of ties, compute Karl Pearson's correlation coefficient between ranks. Regression: lines of regression, fitting of lines of regression by the least squares method, interpretation of slope and intercept. 9.2 Regression coefficient (byx, bxy), Effect of change of origin and scale, Angle between the two lines of regression, Mean residual sum of squares, Residual plot and its interpretation.

Examination Scheme:

Components	**A**	**CT**	**S/V/Q**	**HA**	**EE**
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text and References :

- Goon A.M.,Gupta M. K., Dasgupta B.(1998): Fundamentals of Statistics (V-1),World Press.
- Miller and Fruend: Modern Elementary Statistics.
- Snedecor and Cochran: Statistical Methods, Oxford and IBH Publishers.
- Mukhopadhyay, P: Mathematical Statistics (1996), New Central Book Agency, Calcutta.
- Introduction to Mathematical Statistics, Ed. 4 (1989), MacMillan Publishing Co. New York.
- Gupta and Kapoor: Fundamentals of Mathematical Statistics, Sultan Chand and Sons, New Delhi.
- Neil Weiss: Introductory Statistics: Pearson Publishers.

LEARNING AND REASONING WITH BAYESIAN NETWORKS

Course Code: DSE4302 **Credit Units: 04**

Course Objective:
This course provides an in-depth exposition of knowledge representation, reasoning, and machine learning under uncertainty using the framework of Bayesian networks. Both theoretical underpinnings and practical considerations will be covered, with a special emphasis on constructing and learning graphical models, and on various exact and approximate inference algorithms.

Course Contents:

Module-I: Introduction
Reasoning about beliefs using Logic and Probability Propositional Logic, Probability Calculus and Bayesian Reasoning, Bayesian Networks,: Syntax and Semantics, Building Bayesian Networks.

Module-II: Bayesian Networks Inference
Inference by variable elimination, Inference by Factor Elimination (Jointree), Compiling Bayesian Networks, Complexity of probabilistic inference, compiling bayesian networks.

Module-III: Approximate Inference
Inference by Belief Propagation: Algorithm, Iterative belief propagation, semantics of IBP, Join graphs, edge-detection semantics, Approximate Inference by Stochastic Sampling: Simulating a Bayesian network, direct sampling, expectations, estimating a conditional probability, Markov chain simulation.

Module-IV: Learning: The Maximum Likelihood Approach
Introduction, estimating parameters from complete data, estimating parameters from incomplete data, learning network structure, searching for network structure.

Module-V: Learning: The Bayesian Approach
Introduction, Meta Networks, Learning with Discrete Parameter Sets, Learning with Continuous Parameter Sets, Learning Network Structure

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Adnan Darwiche, Modelling and Reasoning with Bayesian Networks. Cambridge University Press 2009
- Richard E. Neapolitan, "Learning Bayesian networks", Prentice Hall Series in Artificial Intelligence, 2004.
- TimoKoski, John Noble, "Bayesian Networks: An Introduction", Wiley series in Probability and Statistics, 2009.
- Finn V. Jensen and Thomas Nielsen. Bayesian Networks and Decision Graphs. Springer 2007.

SOCIAL NETWORK DATA ANALYTICS

Course Code: DSE4303 **Credit Units: 03**

Course Objective:
This course gives an introduction to social network analysis, with a focus on modelling. It provides an overview of research questions connected to social networks, and of descriptive measures, models, and methods of analysis that can be used to analyse empirical social network data. It helps to understand the online interactive demonstrations and hands-on analysis of real-world data sets

Course Contents:

Module-I: Introduction
Overview: Social network data-Formal methods- Paths and Connectivity-Graphs to represent social relations-Working with network data- Network Datasets-Strong and weak ties - Closure, Structural Holes, and Social Capital.

Module-II: Community Discovery in Social Networks: Applications, Methods and Engineering Trends
Introduction, Communities In Context, Core Methods, Quality Functions, The Kernighan-Lin (Kl) Algorithm, Agglomerative/Divisive Algorithms, Spectral Algorithms, Multi-Level Graph Portioning, Markov Clustering. Other Approaches, Emerging Fields and Problems, Community Discovery in Dynamic Networks, Community Discovery in Heterogeneous Networks, Community Discovery in Directed Networks, Coupling Content and Relationship Information for Community Discovery,

Module-III: Information Networks and the World Wide Web
The Structure of the Web- World Wide Web- Information Networks, Hypertext, and Associative Memory- Web as a Directed Graph, Bow-Tie Structure of the Web- Link Analysis and Web Search, Searching the Web: Ranking, Link Analysis using Hubs and Authorities- Page Rank- Link Analysis in Modern Web Search, Applications, Spectral Analysis, Random Walks, and Web Search. Module IV

Module-IV: Node Classifications in Social Networks
Introduction, Problem Formulation, Representing Data As A Graph, The Node Classification Problem, Methods Using Local Classifiers, Iterative Classification Method, Random Walk Based Methods, Label Propagation, Graph Regularization, Adsorption, Applying Node Classification To Large Social Networks, Basic Approaches, Second-Order Methods, Implementation Within Map-Reduce, Inference Using Graphical Models, Metric Labelling, Spectral Partitioning, Graph Clustering, Variations on Node Classification.

Module-V: Data and Text Mining In Social Media
Data Mining In Nutshell, Social Media, Motivations For Data Mining In Social Media, Data Mining Methods For Social Media, Data Representation, Data Mining- A Process, Social Networking Sites: Illustrative Examples, Related Efforts, Ethnography And Netnography, Event Maps, Text Mining: Keyword Search, Query Semantics And Answer Ranking, Keyword Search over Xmland Relational Data, Keyword Search Over Graph Data, Classification Algorithms, Clustering Algorithms, Transfer Learning in Heterogeneous Networks.

Module-VI: Overview of Social Tagging
Introduction, Problems With Metadata Generation and Fixed Taxonomies, Tags: Why And What?, Different User Tagging Motivations, Kinds Of Tags, Linguistic Classifications Of Tags, Game-Based Tagging, Tag Generation Models, Tagging System Design, Tag Analysis, Tagging Distributions ,Identifying Tag Semantics, Tags Versus Keywords, Visualization Of Tags, Tag Clouds For Browsing/Search, Tag Selection For Tag Clouds, Tag Hierarchy Generation, Tag Cloud Display Formats, Tag Evolution Visualization.

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Easley and Kleinberg, "Networks, Crowds, and Markets: Reasoning about a highly connected world", Cambridge Univ. Press, 2010.
- Charu C. Aggarwal, "Social Network Data Analytics", Springer, 2011.
- Robert A. Hanneman and Mark Riddle, "Introduction to social network methods", University of California, 2005.
- Jure Leskovec, AnandRajaraman, and Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2 edition, 2014.
- Wasserman, S., & Faust, K, "Social Network Analysis: Methods and Applications", Cambridge University Press; 1 edition, 1994.

R PROGRAMMING LAB

Course Code: DSE4304 **Credit Units: 01**

Course Objective:
This lab will provide a basic introduction to the R programming Language and the use of R to perform basic statistics and programming tasks. The main objectives of this lab is to impart the students with the knowledge of R Programming, Machine Learning using R Mining from streaming Data, Mining from Distributed Data.

R Programming Objective

- Master the use of the R interactive environment
- Expand R by installing R packages
- Explore and understand how to use the R documentation
- Read Structured Data into R from various sources
- Understand the different data types in R
- Understand the different data structures in R
- Understand how to use dates in R
- Use R for mathematical operations
- Use of vectorised calculations
- Write user-defined R functions
- Use control statements
- Write Loop constructs in R
- Use Apply to iterate functions across data
- Reshape data to support different analyses
- Understand split-apply-combine (group-wise operations) in R
- Deal with missing data
- Manipulate strings in R
- Understand basic regular expressions in R
- Understand base R graphics
- Focus on GGplot2 graphics for R
- Be familiar with trellis (lattice) graphics
- Use R for descriptive statistics
- Use R for inferential statistics
- Write multivariate models in R
- Understand confounding and adjustment in multivariate models
- Understand interaction in multivariate models
- Predict/Score new data using models
- Understand basic non-linear functions in models
- Understand how to link data, statistical methods, and actionable questions

Examination Scheme:

IA				EE	
A	**PR**	**LR**	**V**	**PR**	**V**
5	10	10	5	35	35

Note: IA –Internal Assessment, EE- External Exam, PR- Performance, LR – Lab Record, V – Viva.

HADOOP LAB

Course Code:DSE4305 **Credit Units: 02**

Course Objective:

- To provide an overview of several key technologies used in manipulating, storing, and analysing big data.
- To understand the fundamentals of Hadoop.
- To apply the learning specific problems in various domains.

Recommended Tools

Big Data Tools and Technology [Learning and Demonstration of Big Data Ecosystem]

- Hadoop
- HBase
- NoSQL
- Hive
- Pig

Examination Scheme:

IA				EE	
A	**PR**	**LR**	**V**	**PR**	**V**
5	10	10	5	35	35

Note: IA –Internal Assessment, EE- External Exam, PR- Performance, LR – Lab Record, V – Viva.

SUMMER INTERNSHIP EVALUATION

Course Code: DSE4335 **Credit Units: 06**

Guidelines:

There are certain phases of every Intern's professional development that cannot be effectively taught in the academic environment. These facets can only be learned through direct, on-the-job experience working with successful professionals and experts in the field. The internship program can best be described as an attempt to institutionalize efforts to bridge the gap between the professional world and the academic institutions. Entire effort in internship is in terms of extending the program of education and evaluation beyond the classroom of a university or institution. The educational process in the internship course seeks out and focuses attention on many latent attributes, which do not surface in the normal classroom situations. These attributes are intellectual ability, professional judgment and decision-making ability, inter-disciplinary approach, skills for data handling, ability in written and oral presentation, sense of responsibility etc.

In order to achieve these objectives:

- **Each student will be allotted a supervisor** for proper guidance.
- **Student will first submit synopsis in the format given by coordinator/supervisor.**
- Student will maintain a file (**Internship File/Project Report**). **Further, coordinator will provide NTCC project guidelines and sample to help in preparation of file.** The Internship File aims to encourage students to keep a personal record of their learning and achievement throughout the Programme. It can be used as the basis for lifelong learning and for job applications. Items can be drawn from activities completed in the course modules and from the workplace to demonstrate learning and personal development. The File will assess the student's analytical skills and ability to present supportive evidence, whilst demonstrating understanding of their organization, its needs and their own personal contribution to the organization.

The **layout guidelines** for the Project Report

1. File should be in the following specification

- A4 size paper
- **Font**

For normal text Font Type and Size must be- Times New Roman, 12 pt. The minimum font size of materials within a table or a figure can be 10 point.

- **Margins**

A margin of 3.75 cm (1½ inch) is to be given on the binding edge while on the other sides it is to be 2.5 cm (1 inch). The text of the report, including headings, figures, tables, and notes, but excluding page numbers, must be accommodated within the page area.

- **Line Spacing**

The line spacing in the main text must be between one-and-a-half (1.5). Single line spacing should be given for figure captions, table titles, figure legends, and footnotes. Equations, tables, figures, and quotations should be set off from the main text with adequate space (not less than the normal line spacing adopted for the main text). Two consecutive paragraphs should be separated by a spacing which must be larger than the line spacing adopted for the text.

- **Tables and Figures**

Each sketch, drawing, graph and photograph should have a figure number and title below the figure etc. Numbering should be sequential, chapter wise. For instance, if there are 24 figures chapter 3 spread over all of its sections the figure numbers run from Figure 3.1 through Figure 3.24. In figures experimental data should typically be represented by centered symbols, and theoretical data by continuous curves.

Each table should have a table number and caption above the table. Numbering should be sequential, chapter wise, as in the case of Figure numbers. For instance, if there are 18 tables in chapter 3 the table numbers run from Figure 3.1 through Figure 3.18.

Make sure that figures and tables are complete in other respects such as legends, references (if any) and coordinate labels with units. Each figure and table must be explicitly referred to in the text and located where its first reference occurs, preferably after the reference.

- **Drawings**

All engineering drawings must conform to relevant Standards and should include a title block. If drawings are large they should be included at the back of the report in a separate pocket. In case drawings are made using CAD packages, a CD ROM should be included which contains all the files and details of the packages used.

- **Equations**

The numbering of equations should be sequential, chapter wise. Numbered equations must be explicitly referred to in the text.

2. Report Size: The maximum number of pages of the Report should be preferably between 50-80 pages.

3. Report Layout: The report should contain the following components

Front Page
Table of Content
Acknowledgement
Student Certificate
Company Profile (optional)
Introduction
Main Body
References / Bibliography

The File will include ***five sections*** in the order described below. The content and comprehensiveness of the main body and appendices of the report should include the following:

1. **The Title Page**--Title - An Internship Experience Report For (Your Name), name of internship organization, name of the Supervisor/Guide and his/her designation, date started and completed, and number of credits for which the report is submitted.

2. **Declaration by the Students-**This is page number (i), the beginning of the small case Roman numeral page numbers. The student has to give a declaration to the effect that the data used for the work, the work depicted in the report, and the written material contained in the report are not copied from others and that due permission has been taken from, and due credit has been given to, the sources whenever they are used.

3. Certificate-This is page number (ii). The certificate will be signed by the Faculty Supervisor(s) before the viva-voce after verifying the format and by the Head of the Department after review with the Supervisor(s).

4. Acknowledgements-This is page number (iii). Keep this brief and avoid using informal language. This page must be signed by the candidate.

5. Abstract and **Keywords-**This is page number (iv). The abstract (preferably one page) should contain the context/relevance of the problem at hand, a description of what was done and a gist of the significant observations/results.

The keywords (maximum 6) are a hint that what is contained in the report.

7. Contents-This is page number (v). The table of Contents should be titled just *Contents* (not Table of Contents). Try to fit it into one or two pages.

8. **Introduction**--short, but should include how and why you obtained the internship experience position and the relationship it has to your professional and career goals.

9. **Main Body**--should include but not be limited to daily tasks performed. Major projects contributed to, dates, hours on task, observations and feelings, meetings attended and their purposes, listing of tools and materials and their suppliers, and photographs if possible of projects, buildings and co-workers.

10. **References / Bibliography** --This should include papers and books referred to in the body of the report. These should be ordered alphabetically on the author's surname. The titles of journals preferably should not be abbreviated; if they are, abbreviations must comply with an internationally recognised system.

ASSESSMENT OF THE INTERNSHIP FILE

Continuous Internal Assessment	Final Assessment
40 Marks	60 Marks

Continuous Internal Assessment consists of topic relevance, progress report and synopsis marks. Final Assessment includes viva, presentation and report marks.

Examination Scheme:

Components	**V**	**S**	**R**	**PR**	**FP**
Weightage (%)	20	20	20	20	20

V – Viva, S – Synopsis, FP – Final Presentation, R – Report, PR-Progress Report

PROJECT-DISSERTATION-I

Course Code: DSE4337 **CreditUnits: 05**

GUIDELINES FOR DISSERTATION

Research experience is as close to a professional problem-solving activity as anything in the curriculum. It provides exposure to research methodology and an opportunity to work closely with a faculty guide. It usually requires the use of advanced concepts, a variety of experimental techniques, and state-of-the-art instrumentation.

Research is genuine exploration of the unknown that leads to new knowledge which often warrants publication. But whether or not the results of a research project are publishable, the project should be communicated in the form of a research report written by the student.

Sufficient time should be allowed for satisfactory completion of reports, taking into account that initial drafts should be critiqued by the faculty guide and corrected by the student at each stage.

The File is the principal means by which the work carried out will be assessed and therefore great care should be taken in its preparation.

In general, the File should be comprehensive and include

A short account of the activities that were undertaken as part of the project;

A statement about the extent to which the project has achieved its stated goals.

A statement about the outcomes of the evaluation and dissemination processes engaged in as part of the project;

Any activities planned but not yet completed as part of the DISSERTION, or as a future initiative directly resulting from the project;

Any problems that have arisen that may be useful to document for future reference.

- **Report Layout**

The report should contain the following components:

- **Title or Cover Page**

The title page should contain the following information: Project Title; Student's Name; Course; Year; Supervisor's Name.

- **Acknowledgements** (optional)

Acknowledgment to any advisory or financial assistance received in the course of work may be given.

- **Abstract**

A good"Abstract" should be straight to the point; not too descriptive but fully informative. First paragraph should state what was accomplished with regard to the objectives. The abstract does not have to be an entire summary of the project, but rather a concise summary of the scope and results of the project

- **Table of Contents**

Titles and subtitles are to correspond exactly with those in the text.

- **Introduction**

Here a brief introduction to the problem that is central to the project and an outline of the structure of the rest of the report should be provided. The introduction should aim to catch the imagination of the reader, so excessive details should be avoided.

- **Materials and Methods**

This section should aim at experimental designs, materials used. Methodology should be mentioned in details including modifications if any.

- **Results and Discussion**

Present results, discuss and compare these with those from other workers, etc. In writing these section, emphasis should be given on what has been performed and achieved in the course of the work, rather than discuss in detail what is readily available in text books. Avoid abrupt changes in contents from section to section and maintain a lucid flow throughout the thesis. An opening and closing paragraph in every chapter could be included to aid in smooth flow.

Note that in writing the various secions, all figures and tables should as far as possible be next to the associated text, in the same orientation as the main text, numbered, and given appropriate titles or captions. All major equations should also be numbered and unless it is really necessary never write in "point" form.

- **Conclusion**

A conclusion should be the final section in which the outcome of the work is mentioned briefly.

- **Future prospects**
- **Appendices**

The Appendix contains material which is of interest to the reader but not an integral part of the thesis and any problem that have arisen that may be useful to document for future reference.

- **References / Bibliography**

This should include papers and books referred to in the body of the report. These should be ordered alphabetically on the author's surname. The titles of journals preferably should not be abbreviated; if they are, abbreviations must comply with an internationally recognised system.

Examples
For research article
Voravuthikunchai SP, Lortheeranuwat A, Ninrprom T, Popaya W, Pongpaichit S, Supawita T. (2002) Antibacterial activity of Thai medicinal plants against enterohaemorrhagic*Escherichia coli* O157: H7. *ClinMicrobiol Infect*, **8** (suppl 1): 116–117.

For book
Kowalski,M.(1976) Transduction of effectiveness in *Rhizobium meliloti*. SYMBIOTIC NITROGEN FIXATION PLANTS (editor P.S. Nutman IBP), **7**: 63-67

ASSESSMENT OF THE DISSERTATION FILE

Essentially, marking will be based on the following criteria: the quality of the report, the technical merit of the project and the project execution.
Technical merit attempts to assess the quality and depth of the intellectual efforts put into the project.
Project execution is concerned with assessing how much work has been put in.
The File should fulfill the following ***assessment objectives:***

Range of Research Methods used to obtain information

Execution of Research
Data Analysis
Analyse Quantitative/ Qualitative information
Control Quality

Draw Conclusions
Examination Scheme:

Dissertation	50
Viva Voce	50
Total	**100**

ata, leading to production of a structured report.

Selecting the Dissertation Topic
It is usual to give you some discretion in the choice of topic for the dissertation and the approach to be adopted. You will need to ensure that your dissertation is related to your field of specialization.

Deciding this is often the most difficult part of the dissertation process, and perhaps, you have been thinking of a topic for some time.

It is important to distinguish here between 'dissertation topic' and 'dissertation title'. The topic is the specific area that you wish to investigate. The title may not be decided until the dissertation has been written so as to reflect its content properly.

Few restrictions are placed on the choice of the topic. Normally we would expect it to be:

- relevant to business, defined broadly;
- related to one or more of the subjects or areas of study within the core program and specialisation stream;
- clearly focused so as to facilitate an in-depth approach, subject to the availability of adequate sources of information and to your own knowledge;
- of value and interest to you and your personal and professional development.

Planning the Dissertation

This will entail following:

- Selecting a topic for investigation.
- Establishing the precise focus of your study by deciding on the aims and objectives of the dissertation, or formulating questions to be investigated. Consider very carefully what is worth investigating and its feasibility.
- Drawing up initial dissertation outlines considering the aims and objectives of the dissertation. Workout various stages of dissertation
- Devising a timetable to ensure that all stages of dissertation are completed in time. The timetable should include writing of the dissertation and regular meetings with your dissertation guide.

The Dissertation plan or outline

It is recommended that you should have a dissertation plan to guide you right from the outset. Essentially, the dissertation plan is an outline of what you intend to do, chapter wise and therefore should reflect the aims and objectives of your dissertation.

There are several reasons for having a dissertation plan

- It provides a focus to your thoughts.
- It provides your faculty-guide with an opportunity, at an early stage of your work, to make constructive comments and help guide the direction of your research.
- The writing of a plan is the first formal stage of the writing process, and therefore helps build up your confidence.
- In many ways, the plan encourages you to come to terms with the reading, thinking and writing in a systematic and integrated way, with plenty of time left for changes.
- Finally, the dissertation plan generally provides a revision point in the development of your dissertation report in order to allow appropriate changes in the scope and even direction of your work as it progresses.

Keeping records

This includes the following:

- Making a note of everything you read; including those discarded.
- Ensuring that when recording sources, author's name and initials, date of publication, title, place of publication and publisher are included. (You may consider starting a card index or database from the outset). Making an accurate note of all quotations at the time you read them.
- Make clear what is a direct a direct quotation and what is your paraphrase.

Dissertation format

All students must follow the following rules in submitting their dissertation.

- Front page should provide title, author, Name of degree/diploma and the date of submission.
- Second page should be the table of contents giving page references for each chapter and section.
- The next page should be the table of appendices, graphs and tables giving titles and page references.
- Next to follow should be a synopsis or abstract of the dissertation (approximately 500 words)
- Next is the 'acknowledgements'.

- Chapter I should be a general introduction, giving the background to the dissertation, the objectives of the dissertation, the rationale for the dissertation, the plan, methodological issues and problems. The limitations of the dissertation should also be hinted in this chapter.
- Other chapters will constitute the body of the dissertation. The number of chapters and their sequence will usually vary depending on, among others, on a critical review of the previous relevant work relating to your major findings, a discussion of their implications, and conclusions, possibly with a suggestion of the direction of future research on the area.
- After this concluding chapter, you should give a list of all the references you have used. These should be cross - references with your text. For articles from journals, the following details are required e.g.

  Draper P and Pandyal K. 1991, The Investment Trust Discount Revisited, Journal of Business Finance and Accounting, Vol18, No6, Nov, pp 791-832.

  For books, the following details are required:
  Levi, M. 1996, International Financial Management, Prentice Hall, New York, 3rd Ed, 1996

- Finally, you should give any appendices. These should only include relevant statistical data or material that cannot be fitted into the above categories.

The Layout Guidelines for the Dissertation

- A4 size Paper
- Font: Arial (10 points) or Times New Roman (12 points)
- Line spacing: 1.5
- Top and bottom margins: 1 inch/ 2.5 cm; left and right margins: 1.25 inches/ 3 cm

Guidelines for the assessment of the Dissertation

While evaluating the dissertation, faculty guide will consider the following aspects:

1. Has the student made a clear statement of the objective or objective(s).
2. If there is more than one objective, do these constitute parts of a whole?
3. Has the student developed an appropriate analytical framework for addressing the problem at hand.
4. Is this based on up-to-date developments in the topic area?
5. Has the student collected information / data suitable to the frameworks?
6. Are the techniques employed by the student to analyse the data / information appropriate and relevant?
7. Has the student succeeded in drawing conclusion form the analysis?
8. Do the conclusions relate well to the objectives of the project?
9. Has the student been regular in his work?
10. Layout of the written report.

Assessment Scheme:

Continuous Evaluation:		40%
(Based on Abstract, Regularity, Adherence to initial plan, Records etc.)		
Final Evaluation: Based on,		60%
Contents & Layout of the Report,	20	
Conceptual Framework,	05	
Objectives & Methodology and	05	
Implications & Conclusions	10	
Viva & Presentation	20	

WEB TECHNOLOGY

Course Code: DSE4306 **Credit Units: 03**

Course Objectives:
This course provides knowledge on Core technologies that are needed for the web likeHTML and XML and facilitate how to build XML applications with DTD and style sheets thatspan multiple domains ranging from finance to vector graphics to genealogy for use with legacy browsers.

Course Contents:
MODULE-I: Introduction
HTML Common tags, Cascading Style sheets - Introduction to Java Scripts - Objects in Java Script - Dynamic HTML with Java Script.

MODULE-II: Vbscript Language Elements
Constants - Variables and Data Types - Mathematical Operations – LogicalOperators - Looping and Decision Structures - VBScript Functions and Objects:Data Conversion Functions - Mathematical Functions - Data FormattingFunctions -Text Manipulation Functions - Data and Time Functions - Built-in
Objects.

MODULE-III: ASP Fundamentals
Using Server – Side Includes- Learning the SSI Directives – Creating ModularASP Code -Using the Request Object: Using Form Information - Using QueryString Information – Using Server Variables - Using the Response Object:Create Output – Managing Output – Managing the Connection.

MODULE-IV: Using Cookies
Introduction to Cookies: Cookies and Your Browser – Creating a Cookie –Modifying and removing Cookies – Tracking Preferences with Cookies Usingthe Application, Session and Server Objects: The application Object – TheSession Object – The Server Object –Using the global .asa file - Active DataObjects Essentials: Microsoft's Universal Data Access Strategy – TheConnection Object – The Record set and Field Objects – The Command andParameter Objects – Using the Errors Collection.

MODULE-V: Introducing XML
XML: The Life of an XML documents - Related technologies- First XMLDocument: Hello XML – Exploring the Simple XML Document – AssigningMeaning to XML Tags – Writing a Style Sheet for an XML Document – Attachinga Style Sheet to an XML Document – Style Languages: CSS Style Sheets,CSS Layouts, CSS Text Styles.

MODULE-VI: Attributes, Empty Tags & XSL
Attributes – Attributes versus Elements – Empty Elements and Empty ElementTags – XSL-DTDs and Validity: Document Type Definitions – ElementDeclarations – DTD Files – Document Type Declarations – Validating Againsta DTD-Element Declaration - Entity Declarations: What Is an Entity – InternalGeneral Entities – External General Entities – Internal Parameter Entities –External Parameter Entities – Building a Document from Places-AttributeDeclaration: What is an Attribute – Declaring Attributes in DTDs – DeclaringMultiple Attributes – Specifying Default Values for Attributes – Attribute Types –Predefined Attributes – A DTD for Attribute- Based Baseball Statistics.

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Dave Mercer, "ASP 3.0 Beginners Guide", Tata McGraw-Hill Edition, Sixthreprint, 2004.
- Rajkamal, "Web Technology", 1st Edition, Tata McGraw - Hill, 2001.

SERVICE ORIENTED ARCHITECTURE

Course Code: DSE4307 **Credit Units: 03**

Course Objective:
The subject gives an introduction to the fundamentals and issues relating to Service Oriented Architecture and bring out the importance of service orientation and web services. It also teaches appropriate tools as technique on how to build the Service Oriented Architecture with web services.

Course Contents:

Module I: Introduction
Basic definition - Fundamentals of SOA - Characteristics and misperceptions about SOA-Benefits and pitfalls of SOA.

Module II: Evolution of SOA
The evolution of SOA - Web service and primitive SOA - The extension of SOA - Web service extension.

Module III: Web Service and Contemporary SOA
Message Exchange Pattern- Service Activity- Coordination- Atomic Transaction- Business Activity- Orchestration – Choreography- Addressing- Reliable Messaging- Correlation and Policies- Meta data Exchange- Security- Notification and Eventing.

Module IV: Principles of Service Orientation
Principles of service orientation -Building SOA-Planning and Analysis- SOA delivery strategies - Service Oriented Analysis Introduction -Service Modelling of Service Oriented Analysis.

Module V: Service Oriented Design
Introduction to service oriented design - WSDL related XML Schema language Basics - WSDL Language Basics - SOAP Language Basics - Service interface design tools - Steps to composing SOA - Consideration for choosing service layers, positioning core SOA standards and choosing SOA extension – Service design and business process design.

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Thomas Erl, "Service Oriented Architecture, Concepts, Technology and Design", Pearson Education, 2009.
- Shankar Kambhampaty, "Service Oriented Architecture for Enterprise Architecture for Enterprise Application", 1st Edition, Wiley Publication, 2008.

NATURAL LANGUAGE PROCESSING

Course Code: DSE4308 **Credit Units: 03**

Course Objective:
This course provides a general introduction including the use of state automata for language processing and syntax including a basic parse. It explains advanced feature like feature structures and realistic parsing methodologies. It also gives concepts of remotes processing and detail information about a typical natural language processing applications.

Course Contents:

Module-I: Introduction
Introduction: Knowledge in speech and language processing - Ambiguity - Models and Algorithms - Language, Thought and Understanding- Regular Expressions and automata: Regular expressions - Finite-State automata. Morphology and Finite-State Transducers: Survey of English morphology - Finite-State Morphological parsing - Combining FST lexicon and rules - Lexicon- Free FSTs: The porter stammer - Human morphological processing.

Module-II: Syntax Analysis
Word classes and part-of-speech tagging: English word classes - Tagsets for English - Part-of-speech tagging - Rule-based part-of-speech tagging - Stochastic part-of-speech tagging - Transformation-based tagging – Other issues - Context-Free Grammars for English: Constituency - Context-Free rules and trees - Sentence-level constructions - The noun phrase -
Coordination - Agreement - The verb phase and sub categorization – Auxiliaries - Spoken language syntax - Grammars equivalence and normal form - Finite- State and Context-Free grammars - Grammars and human processing. Parsing with Context-Free Grammars: Parsing as search - A Basic Top-Down parser - Problems with the basic Top-Down parser - The early algorithm - Finite-State parsing methods.

MODULE-III: Advanced Features and Syntax
Features and Unification: Feature structures - Unification of feature structures
- Features structures in the grammar - Implementing unification - Parsing with unification constraints - Types and Inheritance. Lexicalized and Probabilistic
Parsing: Probabilistic context-free grammar - Problems with PCFGs -
Probabilistic lexicalized CFGs - Dependency Grammars - Human parsing.

MODULE-IV: Semantic
Representing Meaning: Computational desiderata for representations – Meaning structure of language - First order predicate calculus - Some linguistically relevant concepts - Related representational approaches – Alternative approaches to meaning. Semantic Analysis: Syntax-Driven semantic analysis - Attachments for a fragment of English - Integrating semantic analysis into the early parser - Idioms and compositionality - Robust semantic analysis. Lexical semantics: relational among lexemes and their senses - WordNet: A database of lexical relations - The Internal structure of words - Creativity and the lexicon.

MODULE-V: Natural Language Generation
Introduction to language generation - Architecture for generation – Surface realization - Discourse planning - Other issues- Machine Translation: Language similarities and differences - The transfer metaphor - The interlingua idea: Using meaning - Direct translation - Using statistical techniques – Usability and system development.

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text & References:

- Daniel Jurafsky& James H.Martin, “Speech and Language Processing", 2nd Edition, Pearson Education, 2009.
- James Allen, "Natural Language Understanding", 2nd Edition, Pearson Education, 2008.
- Manning, Christopher D and HinrichSchütze, “Foundations of Statistical Natural Language Processing”, Cambridge, 1st Edition, MA: MIT Press, 1999.

AGENT BASED INTELLIGENT SYSTEMS

Course Code: DSE4309 **Credit Units: 03**

Course Objective:
This course provides students basic knowledge of employing intelligent agents in solving complex problems and gives the awareness of the building blocks of agents and working of different types of agents. It also analyses the reasons for uncertainty and ability to design agents to handle them.

Course Contents:

Module-I: Introduction
Definitions – History – Hybrid Intelligent Agents – Agents vs Multi Agent Systems– Structure – Environment – Basic Problem Solving Agents – Complex Problem Solving Agents – Formulating Search Strategies – Intelligent Search.

Module-II: Concepts for Building Agents
Situated Agents: Actions and Percepts - Proactive and Reactive Agents: Goals and Events- Challenging Agent Environments: Plans and Beliefs - Social Agents - Agent Execution Cycle.

Module-III: Knowledge Based Agents
Knowledge Representation – Logic – First Order Logic – Reflex Agent – Building a Knowledge Base – General Ontology – Interference – Logical Recovery.

Module-IV: Planning Agents
Situational Calculus – Representation of Planning – Partial Order Planning – Practical Planners– Conditional Planning - Preplanning Agents.

Module-V: Agents and Uncertainty
Acting under uncertainty – Probability – Baye's Rule – Belief Networks – Utility Theory - Decision Network- Value of Information – Decision Theoretic Agent Design.

Module-VI: Higher Level Agents
Learning Agents – General Model – Inductive Learning – Learning Decision Tree – Reinforcement Learning – Knowledge in Learning – Communicative Agents – Types of Communicative Agents – Future of AI.

Examination Scheme:

Components	A	CT	S/V/Q	HA	EE
Weightage (%)	5	10	8	7	70

CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, EE: End Semester Examination; Att: Attendance

Text and References:

- Stuart Russell and Peter Norvig, "Artificial Intelligence: A Modern Approach", 3rd Edition, Prentice Hall, 2010.
- Lin Padgham, Michael Winikoff, "Developing Intelligent Agent Systems: A Practical Guide", 1st Edition, John Wiley & Sons, 2004.
- ZiliZhang,Chengqi Zhang, "Agent-Based Hybrid Intelligent Systems: An Agent- Based Framework for Complex Problem Solving", 1st Edition, Springer-Verlag New York, LLC , 2004.
- Ngooc Thanh Nguyaaen, Lakhmi C. Jain, "Intelligent Agents in the Evolution of Web and Applications", 4th Edition, Springer, 2009.

Syllabus - Fourth Semester

PROJECT-DISSERTATION-II

Course Code: DSE4437 **Credit Units: 15**

GUIDELINES FOR DISSERTATION

Research experience is as close to a professional problem-solving activity as anything in the curriculum. It provides exposure to research methodology and an opportunity to work closely with a faculty guide. It usually requires the use of advanced concepts, a variety of experimental techniques, and state-of-the-art instrumentation.

Research is genuine exploration of the unknown that leads to new knowledge which often warrants publication. But whether or not the results of a research project are publishable, the project should be communicated in the form of a research report written by the student.

Sufficient time should be allowed for satisfactory completion of reports, taking into account that initial drafts should be critiqued by the faculty guide and corrected by the student at each stage.

The File is the principal means by which the work carried out will be assessed and therefore great care should be taken in its preparation.

In general, the File should be comprehensive and include

A short account of the activities that were undertaken as part of the project;

A statement about the extent to which the project has achieved its stated goals.

A statement about the outcomes of the evaluation and dissemination processes engaged in as part of the project;

Any activities planned but not yet completed as part of the DISSERTION, or as a future initiative directly resulting from the project;

Any problems that have arisen that may be useful to document for future reference.

- **Report Layout**

The report should contain the following components:

- **Title or Cover Page**

The title page should contain the following information: Project Title; Student's Name; Course; Year; Supervisor's Name.

- **Acknowledgements** (optional)

Acknowledgment to any advisory or financial assistance received in the course of work may be given.

- **Abstract**

A good"Abstract" should be straight to the point; not too descriptive but fully informative. First paragraph should state what was accomplished with regard to the objectives. The abstract does not have to be an entire summary of the project, but rather a concise summary of the scope and results of the project

- **Table of Contents**

Titles and subtitles are to correspond exactly with those in the text.

- **Introduction**

Here a brief introduction to the problem that is central to the project and an outline of the structure of the rest of the report should be provided. The introduction should aim to catch the imagination of the reader, so excessive details should be avoided.

- **Materials and Methods**

This section should aim at experimental designs, materials used. Methodology should be mentioned in details including modifications if any.

- **Results and Discussion**

Present results, discuss and compare these with those from other workers, etc. In writing these section, emphasis should be given on what has been performed and achieved in the course of the work, rather than discuss in detail what is readily available in text books. Avoid abrupt changes in contents from section to section and maintain a lucid flow throughout the thesis. An opening and closing paragraph in every chapter could be included to aid in smooth flow.

Note that in writing the various secions, all figures and tables should as far as possible be next to the associated text, in the same orientation as the main text, numbered, and given appropriate titles or captions. All major equations should also be numbered and unless it is really necessary never write in "point" form.

- **Conclusion**

A conclusion should be the final section in which the outcome of the work is mentioned briefly.

- **Future prospects**

- **Appendices**

The Appendix contains material which is of interest to the reader but not an integral part of the thesis and any problem that have arisen that may be useful to document for future reference.

- **References / Bibliography**

This should include papers and books referred to in the body of the report. These should be ordered alphabetically on the author's surname. The titles of journals preferably should not be abbreviated; if they are, abbreviations must comply with an internationally recognised system.

Examples

For research article

Voravuthikunchai SP, Lortheeranuwat A, Ninrprom T, Popaya W, Pongpaichit S, Supawita T. (2002) Antibacterial activity of Thai medicinal plants against enterohaemorrhagic*Escherichia coli* O157: H7. *ClinMicrobiol Infect*, **8** (suppl 1): 116–117.

For book

Kowalski,M.(1976) Transduction of effectiveness in *Rhizobium meliloti*. SYMBIOTIC NITROGEN FIXATION PLANTS (editor P.S. Nutman IBP), **7**: 63-67

ASSESSMENT OF THE DISSERTATION FILE

Essentially, marking will be based on the following criteria: the quality of the report, the technical merit of the project and the project execution.
Technical merit attempts to assess the quality and depth of the intellectual efforts put into the project.
Project execution is concerned with assessing how much work has been put in.
The File should fulfill the following ***assessment objectives:***

Range of Research Methods used to obtain information

Execution of Research

Data Analysis
Analyse Quantitative/ Qualitative information
Control Quality

Draw Conclusions

Examination Scheme:

Dissertation	50
Viva Voce	50

Total 100

ata, leading to production of a structured report.

Selecting the Dissertation Topic
It is usual to give you some discretion in the choice of topic for the dissertation and the approach to be adopted. You will need to ensure that your dissertation is related to your field of specialization.

Deciding this is often the most difficult part of the dissertation process, and perhaps, you have been thinking of a topic for some time.

It is important to distinguish here between 'dissertation topic' and 'dissertation title'. The topic is the specific area that you wish to investigate. The title may not be decided until the dissertation has been written so as to reflect its content properly.

Few restrictions are placed on the choice of the topic. Normally we would expect it to be:
- relevant to business, defined broadly;
- related to one or more of the subjects or areas of study within the core program and specialisation stream;
- clearly focused so as to facilitate an in-depth approach, subject to the availability of adequate sources of information and to your own knowledge;
- of value and interest to you and your personal and professional development.

Planning the Dissertation
This will entail following:
- Selecting a topic for investigation.
- Establishing the precise focus of your study by deciding on the aims and objectives of the dissertation, or formulating questions to be investigated. Consider very carefully what is worth investigating and its feasibility.
- Drawing up initial dissertation outlines considering the aims and objectives of the dissertation. Workout various stages of dissertation
- Devising a timetable to ensure that all stages of dissertation are completed in time. The timetable should include writing of the dissertation and regular meetings with your dissertation guide.

The Dissertation plan or outline
It is recommended that you should have a dissertation plan to guide you right from the outset. Essentially, the dissertation plan is an outline of what you intend to do, chapter wise and therefore should reflect the aims and objectives of your dissertation.

There are several reasons for having a dissertation plan
- It provides a focus to your thoughts.
- It provides your faculty-guide with an opportunity, at an early stage of your work, to make constructive comments and help guide the direction of your research.
- The writing of a plan is the first formal stage of the writing process, and therefore helps build up your confidence.
- In many ways, the plan encourages you to come to terms with the reading, thinking and writing in a systematic and integrated way, with plenty of time left for changes.
- Finally, the dissertation plan generally provides a revision point in the development of your dissertation report in order to allow appropriate changes in the scope and even direction of your work as it progresses.

Keeping records
This includes the following:

- Making a note of everything you read; including those discarded.

- Ensuring that when recording sources, author's name and initials, date of publication, title, place of publication and publisher are included. (You may consider starting a card index or database from the outset). Making an accurate note of all quotations at the time you read them.
- Make clear what is a direct a direct quotation and what is your paraphrase.

Dissertation format

All students must follow the following rules in submitting their dissertation.

- Front page should provide title, author, Name of degree/diploma and the date of submission.
- Second page should be the table of contents giving page references for each chapter and section.
- The next page should be the table of appendices, graphs and tables giving titles and page references.
- Next to follow should be a synopsis or abstract of the dissertation (approximately 500 words)
- Next is the 'acknowledgements'.
- Chapter I should be a general introduction, giving the background to the dissertation, the objectives of the dissertation, the rationale for the dissertation, the plan, methodological issues and problems. The limitations of the dissertation should also be hinted in this chapter.
- Other chapters will constitute the body of the dissertation. The number of chapters and their sequence will usually vary depending on, among others, on a critical review of the previous relevant work relating to your major findings, a discussion of their implications, and conclusions, possibly with a suggestion of the direction of future research on the area.
- After this concluding chapter, you should give a list of all the references you have used. These should be cross - references with your text. For articles from journals, the following details are required e.g.

  Draper P and Pandyal K. 1991, The Investment Trust Discount Revisited, Journal of Business Finance and Accounting, Vol18, No6, Nov, pp 791-832.

  For books, the following details are required:
  Levi, M. 1996, International Financial Management, Prentice Hall, New York, 3rd Ed, 1996

- Finally, you should give any appendices. These should only include relevant statistical data or material that cannot be fitted into the above categories.

The Layout Guidelines for the Dissertation

- A4 size Paper
- Font: Arial (10 points) or Times New Roman (12 points)
- Line spacing: 1.5
- Top and bottom margins: 1 inch/ 2.5 cm; left and right margins: 1.25 inches/ 3 cm

Guidelines for the assessment of the Dissertation

While evaluating the dissertation, faculty guide will consider the following aspects:

- Has the student made a clear statement of the objective or objective(s).
- If there is more than one objective, do these constitute parts of a whole?
- Has the student developed an appropriate analytical framework for addressing the problem at hand.
- Is this based on up-to-date developments in the topic area?
- Has the student collected information / data suitable to the frameworks?
- Are the techniques employed by the student to analyse the data / information appropriate and relevant?
- Has the student succeeded in drawing conclusion form the analysis?
- Do the conclusions relate well to the objectives of the project?
- Has the student been regular in his work?
- Layout of the written report.

Assessment Scheme:

Continuous Evaluation: (Based on Abstract, Regularity, Adherence to initial plan, Records etc.)		40%
Final Evaluation: Based on,		60%
Contents & Layout of the Report,	20	
Conceptual Framework,	05	
Objectives & Methodology and	05	
Implications & Conclusions	10	
Viva & Presentation	20	