PERIYAR UNIVERSITY
SALEM – 11

M.Sc., (BIOINFORMATICS)
Semester System
(Effective from the academic year 2021-2022)
# M.Sc. BIOINFORMATICS

**CHOICE BASED CREDIT SYSTEM**

(For the students admitted from the year 2021-22 onwards)

<table>
<thead>
<tr>
<th>Part</th>
<th>Course Code</th>
<th>Course</th>
<th>Hrs.</th>
<th>Credit</th>
<th>Marks</th>
<th>CIA</th>
<th>EA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Semester I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core I</td>
<td>21PBI01</td>
<td>Introductory Bioinformatics</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Core II</td>
<td>21PBI02</td>
<td>Cell and Molecular Biology</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Core III</td>
<td>21PBI03</td>
<td>Biomolecules</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Core IV</td>
<td>21PBI04</td>
<td>Programming in C and C++</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Elective-I</td>
<td>21PBIE01</td>
<td>Mathematics and Biostatistics</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Practical-I</td>
<td>21PBIP01</td>
<td>Lab Course – I Cell and Molecular Biology Practical</td>
<td>6</td>
<td>4</td>
<td>40</td>
<td>60</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td><strong>Semester II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core V</td>
<td>21PBI05</td>
<td>Genomics and Proteomics</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Core VI</td>
<td>21PBI06</td>
<td>Relational Database Management systems</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Core VII</td>
<td>21PBI07</td>
<td>Structural Biology</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Elective-II</td>
<td>21PBIE02</td>
<td>Genetic and Genetic engineering</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Value Edu</td>
<td></td>
<td>Human Rights</td>
<td>2</td>
<td>1</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Practical-II</td>
<td>21PBIP02</td>
<td>Lab Course- II Programming Lab</td>
<td>6</td>
<td>4</td>
<td>40</td>
<td>60</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Practical-III</td>
<td>21PBIP03</td>
<td>Lab Course- III Sequence and Structural Analysis Lab</td>
<td>6</td>
<td>4</td>
<td>40</td>
<td>60</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td><strong>Semester III</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core VIII</td>
<td>21PBI08</td>
<td>Molecular Modelling and Drug Design</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Core VIX</td>
<td>21PBI09</td>
<td>Omics and Systems Biology</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Core X</td>
<td>21PBI10</td>
<td>Perl, Python and Web Programming</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Core XI</td>
<td>21PBI11</td>
<td>Big Data Analysis and Next Generation Sequencing</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Elective-III</td>
<td>21PBIE03</td>
<td>Data Warehousing and Data Mining</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Practical-IV</td>
<td>21PBIP04</td>
<td>Lab Course-IV Programming in Perl, Python, and Web</td>
<td>6</td>
<td>4</td>
<td>40</td>
<td>60</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td><strong>Semester IV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core XII</td>
<td>21PBI12</td>
<td>Research Methodology</td>
<td>4</td>
<td>4</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Add-On</td>
<td></td>
<td>PHP Scripting Language</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Project</td>
<td>21PBI13</td>
<td>Project Work</td>
<td>14</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

**PROJECT WORK**

Dissertation submission: 120

CIA: 30

Viva-voce: 50

Total Credits: 91

Total Marks: 2200
SEMESTER - I
INTRODUCTORY BIOINFORMATICS

UNIT I
Introduction to Bioinformatics; Computers in Biology to understand biological system; Basic commands of Windows, Unix and Linux operating systems; Concept of open resources in Bioinformatics.

UNIT II
**Sequence analysis:** Biological background for sequence analysis; Sequence alignment; Global, Local, Pairwise and Multiple sequence analysis; Algorithm for alignments; Database searching; Tools for sequence alignment.

UNIT III
**Biological Databases:** Database concepts; Introduction to data types and source; Protein sequence and structural databases; Nucleic acid database; Genome databases; Specialized databases; Carbohydrate databases; clinically relevant drug-drug interaction databases; Information retrieval from Biological databases; Entrez system.

UNIT IV
**Cheminformatics:** Introduction; Chemoinformatics tools; Chemical structure representation (SMILES and SMARTS); Chemical databases: CSD, ACD, WDI, Chembank, PUBCHEM, Chemical structure file formats; Structural isomers; Structure visualization.

UNIT V
**Medical and Pharmacy Informatics:** Introduction to Pharmacy Informatics, Medical transcription; Role of informatics to enhance the services provided by pharmaceutical care givers. Health Information Systems Architecture, Health data management, Medical coding, Telemedicine and Telehealth, Ethics in medical informatics, Pharmacy systems and automation, Informatics applications in pharmacy, survey and evaluation of on-line resources.

References:


CELL AND MOLECULAR BIOLOGY

Unit–I

Cell and Sub cellular Organelles: Organization of prokaryotic and eukaryotic cells. Composition of membranes, endocytosis and exocytosis, membrane transport types and function. Structure and functions of mitochondria and chloroplast (Mitochondria – structure, biogenesis; Chloroplast – structure, biogenesis; Molecular events of electron transport chain, ATP synthesis, photosynthesis and photorespiration), endoplasmic reticulum, ribosomes, Golgi, lysosomes, vacuoles, peroxisomes, glyoxisomes, cytoskeleton.

Unit–II


Unit–III

Unit–IV


Unit-V


**Reference Books**

BIOMOLECULES

UNIT I
Introduction to Biomolecules – Structure and function of DNA, RNA, protein, carbohydrate and lipid. Folding pathways: Principles of protein folding, role of chaperons,

UNIT II

UNIT III
Structure and Stability of Proteins: Peptide bond; Ramachandran Plot; Calculation of conformation of different states of proteins; Properties of α-helix and β-sheets; Secondary structure prediction and determination; connectivities. Tertiary structure: Determination of state of tertiary structure; characteristic balance in rigidity and flexibility; Domain concept (α-,β-,α/β- and α+β-domains) and interacting motifs. Quaternary structure: Geometry, Symmetry and intermolecular interfaces.

UNIT IV
Structure and Stability of Carbohydrates: configuration and Conformation; Step Up and Step down reactions; Conversion of Aldos to Ketos and vice versa. Conformational stability ratings, Anomeric effect, Reverse anomeric effect; Structure and stability of polysaccharides, glycoproteins, glycolipids and proteoglycans. Structure and stability of Nucleic acids (DNA and RNA), topological structure, fine structure of DNA and its organization in genome.

UNIT V
Stereochemistry: Configuration and conformation and stability; Elements of symmetry, Chirality; RS-, EZ-, DL- and dl- system of nomenclature; Stereo-specific and stereo-selective reactions; Determination of relative configuration by asymmetric synthesis. Stereochemistry of proteins and nucleic acids.

Recommended Books:
- Problem Approaches in Biochemistry. Wood and Hood.
- Biochemistry by Voet and voet
• Biochemistry by Zubay
• Ernest Eliel, 1996. Stereochemistry of carbon compounds, Prentice Hall
PROGRAMMING IN C AND C++

UNIT I

UNIT II
Control Statements - Conditional and Unconditional Control Statement – Loopy Statement: while, do–while, for – nested loop – Arrays

UNIT III

UNIT IV
Introduction to C++- Basic concepts of OOPS-Introduction to C++,C vs C++-data types, variables, constants, operators and statements in C++ - Conditional and looping statements

UNIT V

Recommended Books:
- Byron Gottfried.1998. Programming with C (Schaum's Outline Series ) - Tata
- Robert Laffore , Object oriented programming with c++ - Waite series.
- Larry. Wall, Programming Perl - Tom Christiansen, Orielly Publications.
MATHEMATICS AND BIOSTATISTICS

UNIT I

UNIT II

UNIT III
Data Representation: Types of Numerical data, Tables and Graphs, Measures of central tendency: Arithmetic Mean, Weighted arithmetic mean, Median and Mode - Geometric mean and Harmonic mean. Measures of Dispersion: Range, Interquartile range, Average deviation, Standard deviation and Coefficient of variation, Lorenz curve. F ratio prediction and analysis of variance ANOVA.

UNIT IV
Linear Correlation and Regression: Types of Correlation, Methods of studying Correlation: Scatter diagram, Karl Pearson’s Coefficient of Correlation, Spearman’s Rank Correlation. Regression lines, Equations and Coefficients.

UNIT V
Probability: The concept of probability, Sample space, Independent events, Mutually exclusive events, Addition law of probability, Conditional probability, Bayes theorem.
Probability Distributions: Expected value and Variance, Binomial distribution, Poisson distribution, Normal distribution, Analysis of variance, Chi squared distribution, F- Test, Student’s t distribution.
Theory of Sampling: The purpose of sampling, Principles of sampling, Methods of samplings, Techniques of non-probability sampling, Size of Sample, Sampling and Non-Sampling errors.

Recommended Books:
• Balaguruswamy. “Numerical Methods”; Tata Mcgraw Hill.
LAB COURSE – I
CELL AND MOLECULAR BIOLOGY – PRACTICAL

1. Isolation of lymphocytes.
3. Microscopic examination of blood cells, epithelial cells, plant cells, tissue types.
4. Cell counting and viability
5. Estimation of DNA by diphenylamine method.
7. DNA electrophoresis in agarose gel, determination of molecular weight.
8. Restriction enzyme digestion and ligation.
9. SDS-PAGE of proteins and Western hybridization (demonstration)
10. Real-time qPCR (demonstration).
SEMESTER – II
GENOMICS AND PROTEOMICS

UNIT I

UNIT II
Restriction mapping, FISH, Sequence tagged site. Sequencing genomes-chain termination, chemical degradation, pyro sequencing. Sequence assembly - shot gun, clone contig methods. Human genome project- mapping phase- sequencing.

UNIT III

UNIT IV

UNIT V
Analytical proteomics-analytical protein and protein separation techniques- 1D SDS-PAGE, isoelectric focusing, 2D SDS-PAGE, image analysis of 2D gels-MALDI-TOF, HPLC-protein digestion techniques. Protein identification and analysis- Mass spectrometry-tandem mass spectrometry-peptide mass finger printing- SALSA algorithm-protein arrays.
Recommended Books:

RELATIONAL DATABASE MANAGEMENT SYSTEMS

Unit-I

Unit-II

Unit-III
The Structured Query Language: Important Facts - Advantages of SQL - Overview of Oracle’s SQL Environment - Basic Concepts in a Typical Oracle Database Environment - Database Creation - Database Management - Tablespace Creation - Tablespace Management - Table Creation - Dropping or Modifying a Table - SQL Data Manipulation Statements: Insertion of Data - Update Operations - Deletion of Data - Commit and Rollback Operations - Basic Syntax for Queries - Simple Queries

Unit-IV

Unit-V
**Recommended Books:**

STRUCTURAL BIOLOGY

UNIT I
Nucleic acids, Amino acids, Adenylates, Carbohydrates, Lipids, Cofactors, Vitamins, and Hormones. General characteristics of nucleic acid structures (A, T, G, C, U), forces and stabilizing geometries, glycosidic bond, rotational isomers. Stabilizing ordered forms of DNA (A, B and Z), base pairing types, base stacking, tertiary structure of DNA (Supercoiled DNA), Melting of the DNA double helix (Hyperchromicity), Interaction with small ions and small molecules.

Unit II
Composition and levels of protein structure, Conformational analysis and forces that determine protein structures, geometries, phi, psi, omega angles, Ramachandran or steric contour diagram, allowed chi angles of side chains in proteins, hydrogen bonding, disulphide bonds, hydrophobic interactions, vanderwaals forces, potential energy calculations, alpha helices, beta sheets, helix to coil transition.

Unit III
General features and thermodynamic aspects of protein folding, folding kinetics, protein-ligand interactions, Scatchard plot, Relationship between the primary, secondary, and tertiary structure of proteins. Structural families and classes, multifunctional domains. Overview and Challenges in structural bioinformatics, structural genomics: High – throughput structural genomics using x-ray crystallography method

Unit IV

Unit V
Association of macromolecules, molecular conjugates, supramolecular interactions, protein-protein interactions, protein-nucleic acid interactions, lipid/membrane-protein interactions. Molecular mechanics and dynamics (Newtonian and Monte Carlo simulations), theoretical
principles and its importance towards *insilico* simulations, results of molecular dynamics calculations and their implications to biological function.

**Recommended Books:**
- Principles of protein structure by G Schulz and R H Schrimer, Springer Verlag.
- Principles of nucleic acid structure by Sanger, Springer Verlag.
- Introduction to Protein Science by Arthur M Lesk, Oxford University Press.
GENETICS AND GENETIC ENGINEERING

Unit–I

Unit–II

Unit–III

Unit–IV

Unit–V
Recommended Books

LAB COURSE II
PROGRAMMING IN BIOINFORMATICS

MS-Office and HTML.
- Working with MS-Office Packages – One Exercise each in Word, Excel, Power Point and Access.
- Working with HTML Tags and HTML Forms. Creating HTML Pages (At least five different pages to be created using all tags learnt).
- Basic commands in MS-DOS and command line execution in LINUX.

C Programming
- To print the size of char, int, float and double data types.
- To convert any given temperature in a) Centigrade scale to Fahrenheit scale & b) Fahrenheit to Centigrade scale.
- To find the area and circumference of a circle with any given radius.
- To find the volume of a cylinder and cone of any given radius and height.
- To compute the pH of a given solution for a given a) H⁺ ion concentration and b) OH⁻ ion concentration.
- To find the average molecular weight of a double-stranded DNA of length n.
- To calculate the Body Mass Index (BMI) value.

C++ programming
- Write a C++ program to find the sum for the given variables using function with default arguments.
- Write a C++ program to swap the values of two variables and demonstrates a function using call by value.
- Write a C++ program to swap the values of two variables and demonstrates a function using Call by reference using reference type (&).
- Write a C++ program to swap the values of two variables and demonstrates a function using Call by reference using pointer (*).
- Write a C++ program to swap the values of two dynamically allocated variables and release the memory after swapping. (use new & delete operators)
- Write a program to find the largest, smallest & second largest of three numbers. (use inline function MAX and MIN to find largest & smallest of 2 numbers)
LAB COURSE III
LAB IN SEQUENCE AND STRUCTURAL ANALYSIS

1) Data retrieval using ENTREZ (sequence, structure and specialized databases)
2) Database file formats
3) Gene finding (Genscan)
4) Protein prediction
5) Sequence search
6) Sequence alignment
7) Phylogenetic tree construction
8) Sequence comparison
9) Proteome analysis using tools
10) Exon finding
11) Genome homology
12) Molecular visualization using Rasmol
13) Model building using space filling/ball and stick models
14) Find the molecular weight of a DNA with n base pairs in length
15) Find the molecular weight of a given dephosphorylated oligonucleotide sequence
16) Find the molecular weight of a given DNA sequence, after checking for phosphorylation
17) Computing amino acid composition of a given protein sequence.
18) Computing base composition of a given nucleotide sequence.
19) Small molecule building, (ISIS DRAW and CHEMSKETCH)
20) Homology Modeling (SWISS-MODEL, SPDBV)
21) Model structure refinement (SPDBV)
22) Model validation (What Check and ProCheck)
23) Docking using (AUTODOCK or AMBER)
24) 2D and 3D QSAR along with CoMFA and CoMSIA
SEMESTER-III

MOLECULAR MODELING AND DRUG DESIGN

UNIT I

Unit II
Computational quantum mechanics: one electron atom, poly electronic atoms and molecules, Hartree Fock equations; calculating molecular properties using ab initio and semi empirical methods. Molecular mechanics: general features of molecular mechanics force field, bond stretching, angle bending, torsional terms, non-bonded interactions; force field parametrisation and transferability.

Unit III

Unit IV
Molecular Dynamics and Monte Carlo simulation: Introduction – Use of single Model time steps, multiple steps; Setting up MD, energy conservation in MD Simulation; Monte Carlo Simulation. BIOSUITE.

Unit V
Drug design: General approach to discovery of new drugs – Structure based drug design, Ligand -based drug design; Lead discovery- Lead optimization– lead modification – Physiochemical principles of drug action – Drug stereo chemistry –Drug action; Molecular Docking, AUTODOCK and HEX; SAR studies; Pharmacophore identification – QSAR, CoMFA.

Recommended Books:


PERL, PYTHON AND WEB PROGRAMMING

UNIT I

Programming in PERL: Introduction, Basic Operators and Control Structures, Scalars, Lists, Hashes, Pattern Matching and Regular Expressions, Subroutines, Text and String Processing

UNIT II

Advance Programming in PERL: File handling; Object oriented programming:- Classes, Objects, Methods.

UNIT III

BioPERL Programming: General Bioperl classes, Sequences (Bio::Seq Class, Sequence Manipulation), Features and Location Classes (Extracting CDS), Alignments (AlignIO), Analysis (Blast, Genscan), Databases (Database Classes, Accessing a local Database)

UNIT IV

Python Programming: Overview, Data structures, Control Flow, Modules, Basic I/O, Exception Handling, Regular Expressions, File Manipulation, Classes, Standard library

UNIT V

WEB Programming: Common Gateway Interface:- HTML form elements, GET, POST & HEAD Method, CGI Environment Variables, Handling forms, Passing Parameters via CGI, Debugging CGI programs.

Recommended Books:
- D. Tisdall, Beginning Perl for Bioinformatics, O'Reilly, 2001
OMICS AND SYSTEMS BIOLOGY

UNIT I

Introduction and scope of proteomics: Components of a complex mixture and protein sequencing: MALDI TOF MS, QTrap, MS/MS, 2D Gel electrophoresis and protein microarrays. qRT PCR and proteomics. Proteomic approach for clinical studies; Protein biomarker discovery and validation- Body fluid profiles, blood disease profiles, diabetes profiles, infectious diseases.

Unit II

Protein arrays: Basic principles, computational methods for identification of polypeptides, Bioinformatics-based tools for analysis of proteomics data (Tools available at ExPASy proteomics server); databases (such as Inter Pro) Protein-protein interactions; database such as STEINGS and DIP; PPI Modeling in biological systems.

Unit III


Unit IV


Unit V

Recommended Books:

- Julio Collado-Vides (Editor), Ralf Hofestadt (Editor) 2002. Gene Regulation and Metabolism: Postgenomic Computational Approaches, MIT Press.
BIG DATA ANALYSIS AND NEXT GENERATION SEQUENCING

UNIT I

R-Statistical package: Essentials of R-package and libraries, mathematical operations, string operations, Data structures: vectors, data framers, lists, matrices, control loop: if else, while for loops. File Input/Outpur operations. R plots and the graphics library. Overview of statistical packages and Bioconductor libraries in R. Data representation: qualitative and quantitative data types, Tabulation and visual display of data, plotting line plot, scatter plot, frequency histograms, pie-chart, heat map and 3D plots.

UNIT II

Epigenomics: Different platforms of NGS, Overview of metagenomics principles, methylation of DNA and genetics, histone modifications, ChIP-chip ChIP-seq- techniques. Impact of transcriptome on biology, volume of data produced and imported repositories.

UNIT III

Microarray data analysis: Gene expression analysis, statistical method, relative merits of various platforms, Mapping algorithm such as Burro-wheeler, Measuring gene, IncRNA, siRNA, for RNA-seq NGS data. Sequence assembly concepts and challenges in assembling short reads. Hamiltonian cycle and de Brjin.Gene prediction and annotation.

UNIT IV

Identification of genetic variants from genome sequence: SNPs, SNVs, translocation, copy number variation. Concept behind genome-aide association studies. Introduction to various applications. Concepts and algorithms to measure transcriptional regulation; methylation and alternative splicing; small RNA analysis, validation of whole-genome database.

UNIT V

Data analysis interpretation: gene expression analysis, differential expression analysis, Allele-specific expression, prioritizing genetic variants, non-synonym variants (SIFT, Polyphen), Synonyms variants, regulatory variants, statistical methods on rare variants, statistical considerations, Hidden Markov model annotating histone markers.
Recommended Books:

- Norman Maltoff-UC Davis. 2009. The Art of R Programming
- Hadley Wickham.2013. Advanced R Programming
DATA WAREHOUSING AND DATA MINING

Unit-I
Overview and Concepts: Need for data warehousing, Basic elements of data warehousing, Planning and Requirements: Project planning and management, Collecting the requirements. Architecture And Infrastructure: Architectural components, Infrastructure and metadata.

Unit-II
Data Design And Data Representation: Principles of dimensional modeling, Dimensional modeling advanced topics, data extraction, transformation and loading, data quality. Information Access and Delivery: Matching information to classes of users, OLAP in data warehouse, Data warehousing and the web. Implementation and Maintenance: Physical design process, data warehouse deployment, growth and maintenance.

Unit-III

Unit-IV
Advanced Topics: Spatial mining, temporal mining. Visualization : Data generalization and summarization-based characterization, Analytical characterization: analysis of attribute relevance, Mining class comparisons: Discriminating between different classes, Mining descriptive statistical measures in large databases Data Mining Primitives, Languages, and System Architectures: Data mining primitives, Query language, Designing GUI based on a data mining query language.

Unit-V
DBMS- Introduction, overview and types. Relational and transactional Database. Relational database-Introduction to relational DB, Data Definition-Manipulationcontrol-Objects, Views, sequences and Synonyms. Data Abstraction; Data Models; Instances & Schemes; E-R Model - Entity and entity sets; Relations and relationship sets; E-R diagrams; Reducing E-R Diagrams to tables. Network Data Model: Basic concepts; Hierarchical Data Model: Basic Concepts; Multimedia Databases - Basic Concepts and Applications; Indexing and Hashing; Text Databases; Introduction to Distributed Database Processing, Data Security. ORACLE and SQL- introduction and functions in DBMS.

Recommended Books:
- Dunham, M.H.“Data Mining Introductory and Advanced Topics”, Pearson Education.
- Kamber, H., Kaufmann, M. “Data Mining Concepts and Techniques”.
- Xiaohua Hu and Yi Pan 2007. Knowledge Discovery In Bioinformatics, John Wiley & Sons, Inc
Implement the following concepts using Perl:

1. Program to perform control statements and until, for, for each, while operations.
   a) To find greatest of 3 numbers b) To check whether a number is prime or not
   c) Generating fibonacci series. d) Palindrome.
2. To perform matrix arithmetic operations
3. Functions and subroutines
   c) Translate DNA to RNA. d) DNA to protein by using Perl module.

Implement the following concepts using Python:

1. Program to perform control statements and until, for, for each, while operations.
2. String Operations
3. Load, Extract and Retrieve features from GenBank and FASTA Files
4. Split a Sequence File into a Set of smaller files
5. Sequence file formats conversion
6. Manipulations of sequence files

Implement the following concepts Web programming:

1. Write Perl / Python programs to demonstrate Write Perl / PHP programs to display
2. Write Perl / Python programs to display Server, Cookie, Session information and Digital Clock
SEMESTER- IV

RESEARCH METHODOLOGY

UNIT I
Research: Definition, types, approaches, significance, research methods, criteria of good research, Literature collection and citation, research design: Basic principles of experimental design. Result analysis and interpretation, Report writing and manuscript preparation.

UNIT II

UNIT III

UNIT IV
Algorithms: Protein and Nucleic acid sequence Algorithms: Sequence Databases, Use of the algorithms BLAST, Multiple sequence alignments and Clustering algorithms. Phylogeny: Evolutionary trees; Biological networks: Pathway analysis. Protein structure analysis: Protein structure databases; Protein structure comparison; Fold recognition; 3D-1D Profiles; Threading and Comparative structure modeling.

UNIT V
Computational applications: Introduction to internet – firebox, flock, mozilla, netscape., Clustal W, Wingene, oligo primer analysis, SPDBV, Swiss-prot, PDB, restriction analysis, primer design, data mining methods for sequence analysis, web-based tools for sequence searches, motif analysis and presentation.

Recommended Books:
SEMESTER- IV
Add-On (Internship Course)
PHP SCRIPTING LANGUAGE

UNIT - I
Basic of Coding in PHP: Mixing PHP and HTML - Introducing Variables and Operators - PHP Variables.

UNIT - II
Displaying Dynamic Content - Sending E-Mail - Using File System - Uploading Files to Website.

UNIT - III
Establishing a connection - Creating a Database Table - Inserting Data into the Table - Selecting and Displaying Data.

UNIT - IV
System Planning - Adding Contacts Modifying Contacts - Deleting Contacts - Working with Contacts.

UNIT - V
Managing a Simple Mailing List: Mailing List Software - Developing Subscription Mechanism, Mailing Mechanism. Creating Custom Logs and Reports.

RECOMMENDED BOOKS
Dissertation submission: 120
CIA: 30
Viva-voce: 50
Total: 200

Question Paper Pattern

THEORY - Question Paper Pattern [EA] (Total Marks: 75)

Time: Three Hours

Part - A (15 x 1 = 15 Marks)
Answer ALL Questions Multiple Choice

Part - B (2 x 5 = 10 Marks)
Answer ANY TWO Questions out of Five

Part - C (5 x 10 = 50 Marks)
Answer ALL Questions
Either (or) Type Five questions
(One question from Each Unit)