Teaching and Examination scheme for M.Sc. Final Bioinformatics

Course No	Course Name	Lecture hrs/week	Exams hrs	Max Marks
BIM05	Computer Programming in object oriented languages	3	3	100
BIM06	Biological data banks, Data mining and data security, and database management	3	3	100
BIM07	Biostatistics & Computer Application	3	3	100
BIM08	Bio programming	3	3	100
	Combined Practical			
	1. Experimental	12	12	120
	2. Project work and Record			50
	3. Viva –voice			30
		TOTAL		**600**

M.Sc Final Bioinformatics, 2013

Paper V- Computer Programming and Object Oriented Languages

Min Pass Marks 36 Duration 3Hrs Max Marks 100
Note: Attempt any five questions, taking atleast one question from each section. Each question carries mark

Section A

Fundamental of computing, introduction to operating system: WINDOWS, NT, UNIX/LINUX operating systems.
Comparative advantages of security (hacking/cracking), installation, portability and programming of these operating systems.
Computer viruses.
Elements of programming in 'C'-Pointers, pointers to the functions, macro and programming in 'C', graphs, Data structures-Linked list; Stack; Queue; Binary Trees, Threaded binary tree, File Handling in 'C'; exception handling in 'C'.

Section B

An introduction to JAVA programming. Object oriented programming in java. Java basics, working with objects. Arrays, conditionals and loops.
Creating classes and applications in java, java applets basic; graphics, fonts and colour. Simple animation, images and sound. Managing simple events and interactivity. Creating user interfaces with AWT.
Windows, networking and other titbits, modifiers, Access control and class design. Packages, Interfaces, Exceptions, multithreading, streams as I/O using native methods and libraries. Java programming tools. Working with data structures and java image filters.

Section C

Introduction to application development using Visual Basic, working with code and forms, variables, procedures and controlling program executor standard controls.
Data access using data control.
Connecting to oracle database using Visual Basic.
Structured Query Language (SQL): Constraints, types of SQL commands, data correlation, introduction to index, types of index.

Paper VI - Biological Databanks, Data Mining, Data Security and Database Management

Min Pass Marks 36 Duration 3Hrs Max Marks 100
Note: Attempt any five questions, taking atleast one question from each section. Each question carries mark.

Section A

Data warehousing, data capture, data analysis. Introduction to nucleic acid and protein sequence databanks; Gene bank, EMBL, nucleotide sequence databanks, AIDS virus sequence databanks, rDNA databanks. Protein sequence databanks: NBRF-PIR, SWISSPROT, signal peptide databanks. Database similarity alignment –NEEDLEMAN and Wunsch, Smith Waterman algorithms. Multiple sequence alignment-CLUSTAL, PRAS, Patterns, motifs and profiles: Prosite, blocks, Prints –S, Pfam etc. Primer design.
Data security: Science and study of methods of protecting data, discretionary and mandatory access controls, secure database design, data integrity, secure architecture, secure transaction processing, information flow protocols, interference controls and anding. Security models for rational and object oriented database. Security of databases in a distributed environment. Statistical database security. Prerequistes: INFS 762 and INFS 614.

Section B

Database system concepts and architecture. Data models and scheme and instances, Database independence, database language, interface and structure.
Data modelling using entity relationship model: ER model concept, notation for diagrams, mapping constraints, keys. Concept of super key, candidate keys, primary key, Generalization, Aggregation, Reducing ER diagrams to tables.
Relational data model and language: Concepts, integrity constraints, keys, domain constraints, referential integrity, assertions, triggers, foreign key, relational algebra and calculus.
Example database design: Functional dependencies, normal forms, first, second and third functional personal normal forms, BCNF, multivalued dependencies, fourth and fifth normal forms, Steps in database design.

Section C

Query processing and optimization. Transaction processing concepts, concurrency control techniques, Locking technique. Time stamping and concurrency control. Recovery.
Distributed database system. Fragments of relation. Optimization transmission cost by semi joins.
Distributed concurrency control. The optimistic approach. Management of deadlocks and crashes.
Advanced topics in databases like temporal database, spatial database, data mining, data warehousing and its applications.

Paper VII- Biostatistics, Computer applications

Min Pass Marks 36 **Duration 3Hrs** **Max Marks 100**

Note: Attempt any five questions, taking atleast one question from each section. Each question carries mark

Section A

Collection, organization and representation of data: Collection of data. Primary and secondary data. Sampling and sampling design-Census method, sample method, random and non-random sampling. Size of sample. Tabulation and graphics representation. Measure of central tendency and dispersion: Mean, Median, and Mode.

Measure of dispersion: Range, Standard deviation, Lorenz curve.

Skewness and Kurtosis: Objective and measures of skewness. Karl person's coefficient of skewness. Bowley's coefficient of skewness. Kelley's measure of skewness. Kurtosis.

Correlation analysis: Types of correlation-partial and negative correlation, linear and non-linear correlation, methods of studying correlation-scatter diagram, Graphic method, Karl pearson's coefficient of correlation. Correlation of grouped data, Rank correlation, Cocurrent deviation method, Partial and multiple correlation.

Regression analysis: Regression line, Regression equations- of X on Y and Y on X. Regression in a bivariate grouped frequency distribution. Multiple regression.

Section B

Probability theory: Types of probability-Mathematical, posterior and axiomatic probability. Theorems of probability-Addition and Multiple theorem.

Theoretical distribution: Bionomial, Poission and Normal distribution.

Sampling and test of significance: Steps in tests of hypothesis. Sampling distribution. Standard error. Test of significance for attributes. Test for number of success and proportion of success. Test of significance for variables. (large samples)-tests of differences between means of two samples and between two standard deviations. Tests of significances for variables (small samples)-students t-distribution to test the difference between means of two samples, and test the significance of an observed correlation coefficient. Variance ratio test (or F-test).

Chi-square tests and goodness of fit: Characteristics of x^2 test, use of X-test, Yates correlation.

Analysis of variance: One way and two way classification. Multivariate analysis.

Section C

Introduction to Macro and Micro computers, Attachments and peripherals. Hardware and Software. Application of computers in statistical data processing.

Software packages for statistical analysis: SAS, MINITAB, BMDP, SPSS, S-Plus, MATLAB. Academic and research software-XGobi, XLisp-Stat, ExplorN, MANET.

Pitfalls of data analysis by employing statistics: Problem with statistics, Source of bias, Problem with interpretation.

.

Paper VIII - Bioprogramming

Min Pass Marks 36 **Duration 3Hrs** **Max Marks 100**

Note: Attempt any five questions, taking atleast one question from each section. Each question carries mark.

Section A

Biological database: Codd rules, data normalization. Biological database importance and functioning. Types of biological databases, Micro biological databases, Primary sequence databases, Databases of carbohydrates, RNA, Genome, virological, organism.

Sequence database: Nucleotide sequence database, Protein sequence database, The EMBL nucleotide sequence databases, Structure databases.

Bioperl: General introduction to Bioperl documentation, Bioperl classes.

Sequences: The Bioperl sequence to class, Format converter, Sequence classes, building mechanism summary.

Section B

Feature and location classes: Codd feature, Codd reading, extracting CDS, tag system, location, graphical view of features. Sequence analysis tools.

Alignment IO, simple alignment, Codd reading protol 12 dna.

Analysis: Blast, running blast, parsing blast, biotools, BPlite family passers, position specific iterative blast, b12 sequence, bast 2 sequences, blast internal class structure.

Section C

Genscane.

Perl reminders, UML. Perl reminders to use bioperl modules, references, file handle and streams, exceptions, Getopt: Std, classes, BEGIN block.

Perl reminders for a further advanced understanding of bioperl modules, modules, compiler instructions, tie.

Solutions: Sequences, alignments, analysis, and databases.