Register Number: ……………………..

B.TECH DEGREE EXAMINATIONS: NOV/DEC 2012

Seventh and Fifth Semester

BIOTECHNOLOGY

CSE 139 Bioinformatics

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. An example for a high throughput biological experiment leading to large data sets is
   a) Protein purification b) Microarray
   c) Carbohydrate isolation d) chromatography
2. Federated approach for data integration involves
   a) A persistent database b) Distributed data sources with data dictionary
   c) Data cleaning and mining d) Flat file approach
3. Genbank file format is an example of
   a) Xml b) Semi-structured text
   c) Smbl d) Tab delimited text
4. Relational data model relies on queries based on
   a) Relational algebraic operators b) Character similarity/identity
   c) ontologies d) Pattern comparison
5. Data warehousing approach DOES NOT involve use of
   a) Relational databases b) Storage in central repository
   c) Wrappers to query data sources d) Prior transformation/cleaning of data
6. The browsing approach to extract data from databases cannot be used for
   a) Large data sets b) Small data sets
   c) Redundant data sets d) obtaining related data sets
7. Which of the following is NOT a specific biological tool
   a) BLAST b) Clustal W
   c) Phylip d) Hidden markov model
8. An optimal plan in query optimization involves
   a) Returns answer in least time/space b) Creates new data
   c) Reformats existing data d) Returns characteristics of data
9. What is the internal programming language of SRS

a) Perl b) Python
c) icarus d) Java

10. For analysis tools to be incorporated into SRS they should be
    a) Launched from a unix command line b) Launched for a windows command line
    c) Written in python/ruby d) Launched for a MacOS command line

PART B (10 x 2 = 20 Marks)

11. Explain the diversity of biological data.
12. What is N2 factor with respect to programs querying databases
13. What are the disadvantages of storage of biological data in spreadsheets and semi-structured text files
14. Write a short note on the information driven life science discovery process
15. Explain the NCBI Entrez system of browsing driven database solutions
16. Distinguish declarative and procedural methods to access data in Genbank
17. What is a homolog? Can BLAST tool be used to derive this information for a sequence?
18. Explain query optimization
19. What are the advantages of using xml to represent biological data
20. Write a short note on the SRS server at the European bioinformatics Institute (EBI)

PART C (5 x 14 = 70 Marks)

21. a) Explain the nature of biological data, its sources and diversity

(OR)

b) Explain the challenges in storing and managing data from various databases

22. a) Explain the data warehousing and federated approach to data integration

(OR)

b) Explain the role on meta data analysis and ontologies in data organization

23. a) Explain the data integration models of the follwing data banks (a) swiss prot and other ExPASy databases (b) NCBI Entrez (c) AllGenes (d) DiscoveryLink

(OR)

b) Highlight the database integration issues that needs to be addressed to handle complex queries like the following: "Retrieval of all human EST tags that by BLAST are >60% identical over >50 amino acids to mouse channel genes expressed in central nervous system (CNS) tissue"

24\. a) Explain the considerations the need to be optimized for query processing for biological data resources

(OR)

b) Explain the importance of data visualization in designing a biological information system

25\. a) Why is xml increasingly used in biological databases? Explain how xml is integrated in the SRS system. Which support features of SRS are xml are supported by SRS. (4+5+ 5)

(OR)

b) Discuss the data integration capabilities of Sequence Retrieval System (SRS)

****