## BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI (END SEMESTER EXAMINATION)

CLASS: BRANCH	B.TECH : CS/IT/ECE/CE	SEMESTER : VII SESSION : MO/2022	
TIME:	SUBJECT: BE316 BIOINFORMATICS ALGORITHMS (OE III) 3:00 Hours	FULL MARKS: 50	
INSTRUC 1. The o 2. Atten 3. The n 4. Befor 5. Table	CTIONS: Juestion paper contains 5 questions each of 10 marks and total 50 marks. Inpt all questions. Inissing data, if any, may be assumed suitably. The attempting the question paper, be sure that you have got the correct question p tes/Data hand book/Graph paper etc. to be supplied to the candidates in the examin	aper. Nation hall.	
Q.1(a) Q.1(b) Q.1(c)	What are the tuples and attributes in the relational database? Explain with an example. Discuss the content and file formats of NCBI, PDB, and KEGG databases. Using the information from various databases and bioinformatics tools, how can you contribute to the fight against COVID-19 diseases?		[2] [3] [5]
Q.2(a) Q.2(b) Q.2(c)	How does binary search work? Explain with a hypothetical example of at least ten numbers. Heap sort is not a stable algorithm. Demonstrate with an example. Why do we need sorting and searching algorithms for bioinformatics research? Explain with suitable examples.		[2] [3] [5]
Q.3(a)	<ul> <li>belect the correct option of the regular expressions that would fit for the sequence EWILKDF, also explain your choice.</li> <li>E-M-x-[ILV]-x{2}-F</li> <li>[EN]-W-x-[ILV]-[RKH]-x-F</li> <li>[EN]-W-x{2}-[ILV]-F</li> <li>F-W-1-[ILMV]-x-K-[FA]</li> </ul>		[2]
Q.3(b)	Considering the below figure, calculate the sensitivity and selectivity.		[3]



Q.3(c) Considering the following sequences, calculate the scores of all the possible motifs of length 4 using [5] Gibbs's sampler algorithm for the outlier sequence.

Seq-1: ACCGTGGTGT Seq-2: TGGCACAAGC Seq-3: GCCGATAGTC Seq-4: AGTGGCGAAC

## Seq-5: CCTGTGGTCA

Note:

- 1. Consider sequence-2 as an outlier sequence and only perform the calculation for the first iteration.
- 2. Select the random motif of length 4 in all the sequences.

- Q.4(a) Explain how to sequence length and its type (DNA Vs. Protein) affects the significance of alignment. [2]
- Explain the three main classifications of multiple sequence alignment methods with one example Q.4(b) [3] each briefly.
- Generate the alignment for the following sequences using the Needleman-Wunsch algorithm. Use Q.4(c) [5] the BLOSUM-62 AA substitution matrix (given below) for match and mismatch data and consider gap penalty -8.

Sequences: HEAGAWGHEE and PAWHEAE.



- Q.5(a) Discuss the significant differences between unsupervised machine learning algorithms and deep [2] learning. [3]
- Q.5(b) Explain the following terms of the Genetic algorithm.
  - a) Breeding
  - Children b)
  - Parents c)
  - d) Chromosome
  - Crossover e)
  - Mutation f)
- Q.5(c) Discuss the applications of bioinformatics in biotechnology research by explaining the research [5] articles discussed in the class presentation.

Note: You can only discuss the examples presented by other groups.

## :::::28/11/2022:::::M