**Department of Computer Engineering**

**Name of the Student: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Roll Number: \_\_\_\_\_\_\_**

**SAP ID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Class: \_\_\_\_\_\_\_\_\_\_\_\_\_ Division:\_\_\_\_\_\_\_\_\_\_ Batch:\_\_\_\_\_\_\_\_\_\_**

**Subject:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**DATE OF PERFORMANCE:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ DATE OF SUBMISION:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**A.Y. 2024-2025**

|  |  |
| --- | --- |
| **Course:** Web Intelligence | **Course Code:** DJ19CEC801 |
| **Class:** B. Tech – C1 and C2 | **Semester:** VIII |

**ASSIGNMENT NO: 01** (Based on Module 3)

|  |  |  |
| --- | --- | --- |
| Que. No. | Questions | CO Mapped |
| 1. | What is a wrapper? Explain how wrapper induction is different from automatic extraction? |  |
| 2. | Explain the following terms:   1. Types of data rich pages 2. Landmarks 3. Wildcards |  |
| 3. | Differentiate between landmark refinement and topology refinement in web intelligence |  |
| 4. | Write a short note on: Instance-based wrapper learning |  |
| 5. | With the help of a suitable example, explain string edit distance and tree edit distance. |  |
| 6. | What is a DOM tree? Build a corresponding DOM tree representation for the following script:    Enlist the use cases of DOM trees. |  |
| 7. | How “Center star method” and “partial tree alignment method” tackle the problem of pairwise alignment? Explain with a suitable example in web intelligence. |  |