COURSE CODE (CREDITS): 20B1WCI732 (2)

MAX. MARKS: 35

COURSE NAME: From Graph to knowledge Graph

COURSE INSTRUCTORS: Ravindara Bhatt

MAX. TIME: 2 Hours

Note: All questions are compulsory.

1. Find names in the text and classify them by type {ORG, PER, LOC, MISC} for Named Entity Recognition (NER) task given below.

   The European Commission said on Thursday it disagreed with German advice. Only France and Britain backed Fischler'sproposal. "What we have to be extremely careful of is how other countries are going to take Germany's lead", Welsh National Farmers ' Union chairman John Lloyd Jones said on BBC radio.[5 Marks]

2. Given the following set of words {*high, low, verb, semisupervised, automatic, noun*}, fill in the blanks for these statements. Note that some words in the set may be used more than once (and every word is used at least once):

   a) Rule-based IE techniques are expected to have _____ precision and _____ recall.

   b) Bootstrapping is an example of a(n) _____ technique for IE.

   c) _____ extractors have _____ cost but usually lead to semantic drift.

   d) Supervised extractors have _____ cost, but have _____ precision.

   e) Semisupervised extractors require a _____ amount or training data.

   f) One of the concrete subproblems in IE from natural-language text is defining the domain of interest. One simple approach to do it automatically is by setting any _____ phrase as a candidate entity and any _____ phrase as a candidate relation.
   [5 Marks]

3. Which of the following is/are true? Justify your answer?

   a) Ontologies are machine-readable and Interoperable

   b) Ontologies can be built in a modular fashion

   c) For every graph G, $\chi(G) \leq n(G) - \alpha(G) + 1$, where $\alpha(G)$ is the independence number of G and $\chi(G)$ is the chromatic number.

   d) If G is a connected graph, then $\chi(G) \leq 1 + a(G)$, where $a(G)$ is the average of the vertex degrees in G and $\chi(G)$ is the chromatic number.

   e) Knowledge Graph uses Ontologies for Multilevel Relationships
   [1 ×5 = 5 Marks]

4. Table A provides information on books, authors, and publishers. Identify classes, relationships, and attributes. Combine classes and relationships and view ontology in a graph format. Using ontology as a framework add in real data about individual books, authors, publishers, and locations to create a knowledge graph.
   [5 Marks]

Table A (Books, Publishers, Authors)

Books

Title	Author	Publisher	Year Published	Followed By
To Kill a Mockingbird	Harper Lee	J. B. Lippincott Company	1960	Go Set a Watchman
Go Set a Watchman	Harper Lee	HarperCollins, LLC: Heinemann	2015	
The Picture of Dorian Gray	Oscar Wilde	J. B. Lippincott & Co.	1890	
2001: A Space Odyssey	Arthur C. Clarke	New American Library, Hutchinson	1968	

Publishers

Name	City	Country
J. B. Lippincott & Company	Philadelphia	United States
HarperCollins, LLC	New York City	United States
Heinemann	Portsmouth	United States
New American Library	New York City	United States
Hutchinson	London	United Kingdom

Authors

Name	Country of Birth
Harper Lee	United States
Oscar Wilde	Ireland
Arthur C. Clarke	United Kingdom

5. You work as a manager for a chip manufacturer, and you currently have 3 people on the road meeting clients. Your salespeople are in Jaipur, Pune and Bangalore, and you want them to fly to three other cities: Delhi, Mumbai and Kerala. The table B below shows the cost of airline tickets in INR between the cities. Apply the Hungarian method to the Table B in order to minimize fare?   **[5 Marks]**

Table B (Cost Matrix)

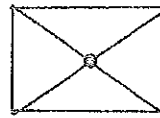|	Delhi	Kerala	Mumbai
Jaipur	1500	4000	4500
Pune	2000	6000	3500
Bangalore	2000	4000	2500

Figure 1

6.

(a) Assign weights (1, 1, 2, 2, 3, 3, 4, 4) to the edges (**Figure 1**) in two ways: one way so that the minimum-weight spanning tree is unique, and another way so that the minimum-weight spanning tree is not unique.

(b) Develop an algorithm to test whether a graph is bipartite. The graph is given by its adjacency matrix or by lists of vertices and their neighbors.   **[2.5 + 2.5 Marks]**

7. Ordinarily, focused crawlers (and many other types of crawlers as well) take as traditional inputs such as some starting (or seed) URLs and, possibly, a topic description (e.g., a list of keywords). Suggest at least two innovative mechanisms can be used for representing a user's intent. What domains would they be particularly useful for?   **[5 Marks]**