# CSL7340 - Natural Language Processing

Major Exams - Time 2 hours

Full marks - 20

***Each question carries 5 marks with part marking specified. You can use hand-drawn diagrams or text, if you wish to. But all submissions of scanned content have to be within specified time only.***

1. Consider the passage in quotes extracted from a scientific paper to answer the following questions - “***Ivermectin is an inhibitor of the COVID-19 causative virus (SARS-CoV-2) in vitro. A single treatment with it is able to effect ~5000-fold reduction in virus at 48 h in cell culture***.”

(a). What are all the relations that can be extracted from the above passage? (1.5)

(b). Explain how a relation extraction system can be designed to extract all the above relations. (2)

(c). If you store all the above-mentioned relations in a Knowledge-graph - can your system handle the question “What is the effect of a single dose of Ivermectin in reducing COVID-19 spread?” If it can - mention how it can do so. If it can’t then state why it can’t. (1.5)

1. If you have to design a chatbot for a pharmaceutical company that answers customer queries related to medicines, their prices, availability and alternatives -

(a). what are the different kinds of knowledge base designs that you will choose to store knowledge for the chatbot and why? Explain with examples. (2.5)

(b). If you had to design an intent-classifier for the chatbot from scratch, how will you design it? (2.5)

1. (a). What is the relevance of “vanishing gradient” problem to NLP? Explain with example. (2)

(b). Explain the roles of LSTM and transformer based language models in tackling the problem - indicating the advantages and limitations of each. (3)

1. “Frequent n-grams in a repository are often useful domain terms.”

(a). Present a method to identify all frequent n-grams from a repository without doing phrase detection or POS tagging. (2)

(b). “Learning embeddings for these terms can be more useful than learning embeddings for the component words” - cite one example task from a scientific domain of your choice to support the statement. (1)

(c). Present the outlines of a method that can be deployed to learn such embeddings. (2)