Name: **Enrolment No:**	

UPES

End Semester Examination, May 2023

Course: Computational Linguistics & NLP **Semester: VI**

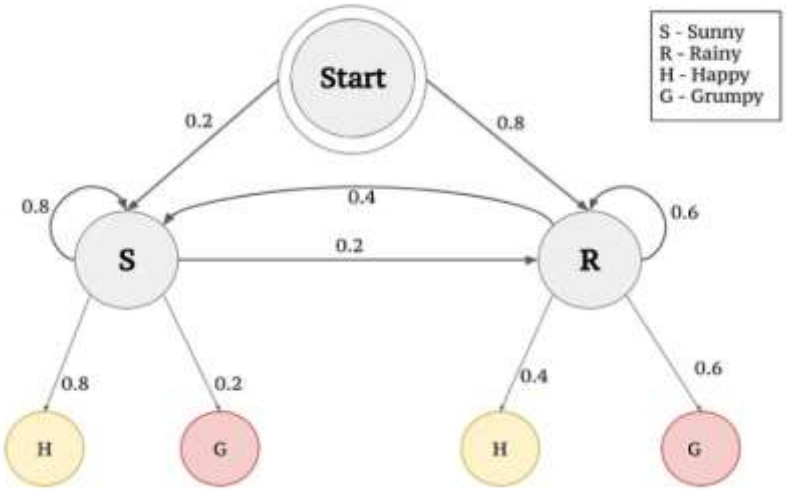
Program: BTech CSE AIML **Time 03 hrs.**

Course Code: CSEG3024 **Max. Marks: 100**

Instructions: All Questions are compulsory. Calculators are allowed. Any missing values can be assumed with proper mention.

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Describe Naïve Bayes and how can it be used for text classification with proper working example.	**4**	**CO1**
Q 2	Assume we have a corpus of three documents: Document 1: "I love cats." Document 2: "I love dogs." Document 3: "I have a black cat." Represent these documents using the BoW model, which involves converting them into vectors of word frequencies.	**4**	**CO2**
Q 3	Explain stochastic POS Taggers?	**4**	**CO2**
Q 4	If the value of Inverse Document Frequency of a word w is 10 times the word w', what does it signify?	**4**	**CO3**
Q 5	Differentiate between morphological segmentation and machine translation?	**4**	**CO2**
	SECTION B **(4Qx10M= 40 Marks)**		
Q 6	Explain in detail various approaches of parsing with examples.	**10**	**CO2**
Q 7	Read the below corpus and answer the questions that follow: <s> I am Sam </s> <s> Sam I am </s> <s> I do not like green eggs and ham </s> a) Calculate the probability of **<s> I I am not </s>** using bigram. b) What will the next most probable word that will follow **<s> I ?**	**10**	**CO1**
Q 8	Differentiate between word2vec and BERT. Which of them to be used in what cases? Explain with examples.	**10**	**CO2**
Q 9	Suppose we have two documents A and B, represented as sets of words: Document A: {"apple", "orange", "banana", "grape", "cherry"} Document B: {"orange", "banana", "kiwi", "grape", "watermelon"}	**10**	**CO1**

	Calculate the cosine similarity and Jaccard similarity between these two documents.		
SECTION-C (2Qx20M=40 Marks)			
Q 10	Write about the following in short: a) Summarization of CDS reports b) AUTOTUTOR architecture c) Scripted dialogue and its use d) Sentiment Analysis	**20**	**CO3**
Q 11	Input text: "THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG". What will the total message size if this input text is sent: a) without any compression b) with Huffman text compression or, Consider the below figure which elaborates how a person feels on different climate:  Consider the sequence of emotions: H,H,G,G,G,H for 6 consecutive days. Find out the more likelihood of the series.	**20**	**CO1,CO 2**