26-3-18

DOON UNIVERSITY, DEHRADUN

Mid Semester Examination, 8ᵗʰ Semester, 2018

Academic Year 2017-18 (Even Semester)

School of Technology

Integrated M.C.A

TMC-554 Natural language Processing

Time Allowed 2.00 Hours   *Maximum Marks: 30*

SECTION : A

Attempt all the questions.All questions carry equal marks.(2+2+2)

Q.1)Define conditional probablity and bayes theorem with example.

Q.2)Discuss Challenges with Natural Language Processing.

Q.3)Explain the role of "types and features" in sementic representation.

SECTION : B

Attempt all the questions.All questions carry equal marks.(6+6)

Q.1)Explain viterbi algorithm.Find the most likely tag sequence using viterbi algorithm.

Q.2)Write short notes on:(2+2+2)

1. Noun Phrases 2.Verb Phrases 3.Transitivity and Passives

SECTION : C

Attempt all the questions.All questions carry equal marks. (6+6)

Q.1)You have a corpus consisted of 500 sentences and have words in only four categories: N, V, ART, and P, including 850 nouns, 300 verbs, 550 articles, and 305 prepositions for a total of 2005 words.For the sentence "flies like a flower",calculate the transition probablities using lexical probablities and *HMM* model.

Category	Count at i	Pair	Countat i,i+1	Bigram	Estimate
<start>	500	<start>,ART	275	....	....
<start>	500	<start>,N	75	....	....
ART	550	ART,N	550	......	......
N	850	N,V	300	......	......

N	850	N,N	250	....	......
N	850	N,P	100	......	......
V	300	V,N	150	....	....
V	300	V,ART	300	....	....
P	305	P,ART	250	....	......
P	305	P,N	150	....	......

Bigram probabilities from the generated corpus

	N	V	ART	P	Totel
flies	21	23	0	0	44
fruit	50	5	1	0	56
Like	12	30	0	31	73
a	1	0	201	0	202
the	1	0	300	2	303
flower	55	15	0	0	70
flowers	45	16	0	0	61
birds	64	1	0	0	65
others	601	210	48	282	1141
Totel	850	300	550	305	2005

Some corpus word counts

OR

Define Natural language Processing and discuss the current issues and future scope of NLP in the field of:

1. Social media.

2. Management.

Q.2) Given sentences:

I) "Humans are social animals and language is our primary tool to communicate".

II)" Natural Language Processing is the science of teaching machines how to understand the language we humans speak and write".

(a) What will be the part-of-speech tags for each word in document? Clearly indicate your PoS tag set. (1+1)

(b) What will be the content of document after "stop" word removal? (1+1)

(c) What will be the content of document after "stemming" each word in the document after stop word removal? (1+1)