Department of Computer Science (BS-SE)

IQRA NATIONAL UNIVERSITY PESHAWAR

Sessional (Fall- 2020)

Natural Language Processing

Name	Hassan Saad
ID	14286

Semester: 5th

Deadline: 10th June 2020 **Total Marks: 20**

Instructor: Aasma Khan **Date: 9th May, 2020**

Note: Attempt all Questions.

Question No. 1: **(15)**

a) Explain Part of Speech Tagging (POS) and explain POS tag ambiguity with two examples.
(03)

- **POS Tagging**:
  Pos tagging is a process that attaches each word in a sentence with a suitable tag from a given set of tags.
  The given set of tags is called target
  e.g POS TAGS
  NN- Noun ; e.g Dog- NN
  VM- Main verb; e.g Run VM
- **Types of POS Tagger**: POS-tagging algorithms fall into two distinctive groups: Rule-Based POS Taggers and Stochastic POS Taggers.
- **POS tag ambiguity:**
  In English post tag ambiguity. A $Bank_1$ on the $Bank_2$ on the river $Bank_3$ for transaction $Bank_1$ is verb the other two banks are noun.
  People jump high
  People Noun/Verb
  Jump Noun/Verb
  High Noun/Adjective
  List of all possible tags for each word

b) State difference between open vs. closed classes.
(02)

Open Classes:

Open classes (like nouns, verbs and adjectives) acquire new members constantly. Open classes normally contain large numbers of words

Close Classes:

Closed class is one to which new items are very rarely added such as (pronouns and conjunctions. Closed classes normally contain small numbers of words

c) Apply Viterbi Algorithm on the below given bigram and lexical probabilities; (10)

Initial Probabilities	
Noun	1\3
Verb	0
Other	1\3

Bigram Probabilities			
	Noun	Verb	Other
Noun	1\4	1\4	0
Verb	1\4	0	1\4
Other	1\3	0	1\3

Lexical Probabilities					
	O1=time	O2=flies	O3=like	O4=an	O5=arrow
Noun	1\5	1\5	0	0	1\5
Verb	1\5	2\5	1\5	0	0
Other	0	0	1\5	2\5	0

Question No. 2: **(05)**

Apply Bayesian theorem over the below given string:
^John got many NLP books. ^He found them all very interesting.
Where for lexical probabilities assume John=0.5, got=0.3, many=0.2, NLP=0.1 and books=0.

Good Luck ☺

ANSWER NO 2:

POS Tags:

^N V A N N. ^ N V N A R A.

Recording Numbers:

	^	N	V	A	R	.
^	0	2	0	0	0	0
N	0	1	2	1	0	1
V	0	1	0	1	0	0
A	0	1	0	0	1	1
R	0	0	0	1	0	0
.	1	0	0	0	0	0

Bigram Probability:

Bigram Probability= P(x|y)=P(a.b)/P(a)

Now putting all the values in Bayes theorem i.e.

	^	N	V	A	R	.
^	0	2	0	0	0	0
N	0	1/5	2/5	1/5	0	1/5
V	0	1/2	0	1/5	0	0
A	0	1/3	0	0	1/3	1/3
R	0	0	0	1	0	0
.	1	0	0	0	0	0

P(T)=P(W/T)=TTP(Ti-/Ti-1)x P(Wi/Ti)
Where lexical probability is given i.e.
John=0.5, got=0.3, many=0.2, NLP=0.1 and books=0.