Submitted By	**Hooria Khan Orakzai**
ID	**14263**
Department	**BS(Software-Engineering)**
Semester	**5th**
Subject	**Natural Language Processing**
Submitted To	**Mam Aasma Khan**

Department of Computer Science (BS-SE)

IQRA NATIONAL UNIVERSITY PESHAWAR

Final Examination (Spring- 2020)

Natural Language Processing

Semester: 5th

Time: 6 Hours 9 AM-3 PM **Total Marks: 50**

Instructor: Aasma Khan **Date: 22nd June, 2020**

Note: Attempt all Questions.

Question No. 1: (5)

Explain objectives of NLP? Write the name of 2 Applications of NLP with example? Write the name of 2 Challenges of NLP with example?

Answer: Objective of NLP is to read, decipher, understand, and make sense of the human languages in a manner that is valuable.

- **2 Applications of NLP:**

  Speech Recognition

  Sentiment Analysis

  - **Speech Recognition:**

    Speech recognition is simply the ability of a software to recognize speech. Anything that a person says, in a language of their choice, must be recognized by the software.

    Example:

    The example of speech recognition is Siri in iPhone and Google Assistant in Android mobiles.

  - **Sentiment Analysis:**

Sentiment Analysis (also known as opinion mining or emotion AI) is a sub-field of NLP that tries to identify and extract opinions within a given text across blogs, reviews, social media, forums, news etc

Example:

1) "I really like the new design of your website!" → Positive.
2) "I'm not sure if I like the new design" → Neutral
3) "The new design is awful!" → Negative

Question No. 2: (5)

Specify the text strings using the below regular expressions:

a. /a(bc)
   - Given string: ab abc ac acb a0b a2b a42c A87d
   - Answer: ab abc ac acb a0b a2b a42c A87d

b. /[abc]
   - Given string: ab abc ac acb a0b a2b a42c A87d
   - Answer: ab abc ac acb a0b a2b a42c A87d

c. /abc+
   - Given string: ab abc abcc babc
   - Answer: ab abc abcc babc

d. /abc*
   - Given string: ab abc abcc babc
   - Answer: ab abc abcc babc

e. /[^a-z A-Z 0-9]
   - Given string: a89 opx cfff $1!
   - Answer: a89 opx cfff $1!

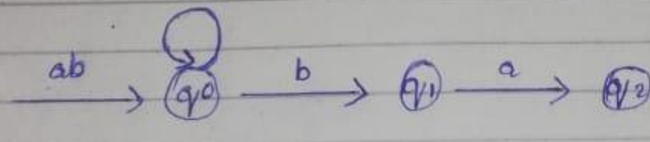
Question No. 3: (10)

a) Design an NFA over an alphabet ∑= {a, b} such that every string accepted must end with a string --ba. Identify its tuples and also convert it into DFA. (05)

b) Design an NFA for the regular expression : $a^* b(a+b)^*$ (05)

Name: Hooria Khan Orakzai
ID: 14263.

Ques No 3 (Part A)

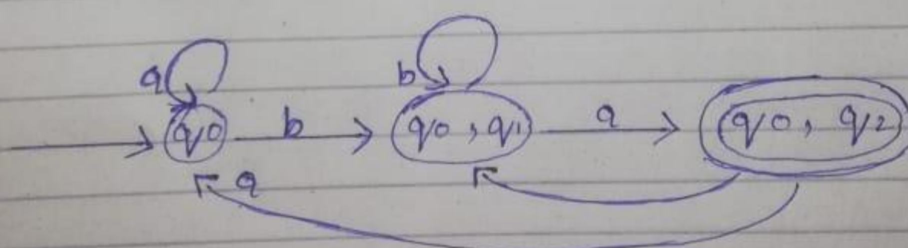


NFA

Q \ Σ	a	b
q0	q0	q0, q1
q1	q2	∅
q2	∅	∅

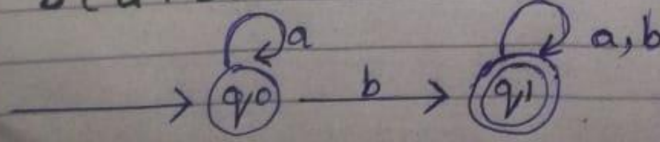
DFA

Q \ Σ	a	b
q0	q0	{q0, q1}
{q0, q1}	{q0, q2}	{q0, q1}
{q0, q2}	{q0, q2}	{q0, q1}



(PART B)

$a^* b (a+b)^*$



Question No. 4: (15)

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

POS Tagging:

It is a process of converting a sentence to forms – list of words, list of tuples (where each tuple is having a form (word, tag)). The tag in case of is a part-of-speech tag, and signifies whether the word is a noun, adjective, verb, and so on

Types Of POS Tagger:

POS-tagging algorithms fall into two distinctive groups: Rule-Based POS Taggers and Stochastic POS Taggers.

POS Tag Ambiguity:

Common parts of speech in English are noun, verb, adjective, adverb, etc. The POS tagging problem is to determine the POS tag for a particular instance of a word. The main problem with POS tagging is ambiguity. In English, many common words have multiple meanings and therefore multiple POS. The job of a POS tagger is to resolve this ambiguity accurately based on the context of use.

For Example:

People jump high
People noun/verb
Jump noun/verb
High noun/adjective
List of all possible tag for each word

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

Open Classes:

An open class is one that commonly accepts the addition of new words open classes (like nouns, verbs and adjectives) acquire new members constantly. Open classes normally contain large numbers of words

Close Classes:

An closed class is one to which new items are very rarely added, a closed. Open classes (like nouns, verbs and adjectives) acquire new members constantly. 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

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

ANSWER:

hi	P1(h1)	P2(h2)	P3(h3)	P4(h4)	P5(h5)
Noun	1/3x1/5=1/15	1/5x1/4x1/15=1/300	0x1/150=0	0	1/5x1/4x1/16875
Verb	0	2/5x1/4x1/15=1/150	1/5x1/4x1/150=1/3000	0	0
Other	0	0	1/5x1/3x1/150=1/2250	1/3x2/5x1/2250=1/16875	0

Question No. 5: (15)

a) 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.

P(T)=PWIT)=tt(ti/ti-1)-P(wi)tiP1(ti/ti-1)=P(Wi/ti)

Le corpus :^ John got many NLP books found all very interesting

POS tagged

^N V 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 Probabilties P(N/V)=# (^=N)/#^

	^	N	V	A	R	.
^	0	1	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

Lexical probability:

	John	Got	Many	Nlp	Books
^	0.5	0.3	0.2	0.1	0
N	0.5	0.3	0.2	-	-
V	0.5	0.3	0.2	-	-
A	0.5	-	-	-	-

R	0.5	-	-	-	-
.					

P(John/^) = P(Wi=John/ti=^)

=# (John,^)/ #^

\# (m y,^)/#^ = 0.5x2/2 = 0.5

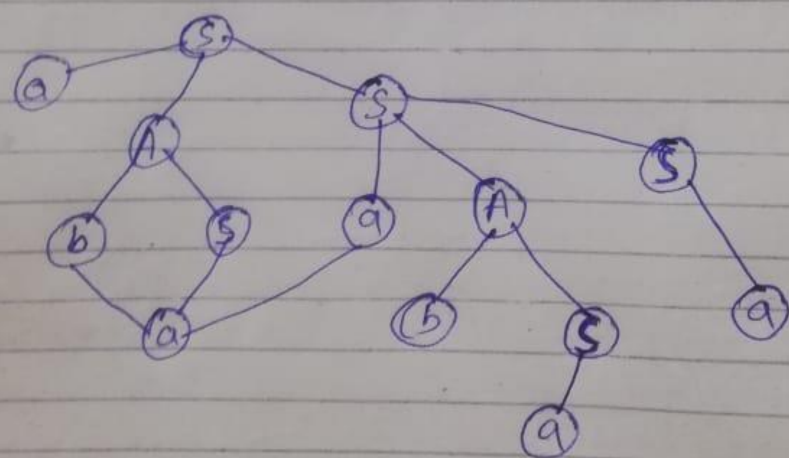
b) Find the CFG of the string "abaabaa" using the production rules

S→a, S→aAS, A→bS

Name: Hooria Khan Orakzai
ID: 14263

QUES No 5 (PART B)

S -> aAS
S -> abSS (A -> bS)
S -> abaS (S -> a)
S -> abaaAS (S -> aAS)
S -> abaabSS (A -> bS)
S -> abaabaS (S -> a)
S -> abaabaa (S -> a)



Good Luck ☺