Department of Computer Science (BS-SE)
IQRA NATIONAL UNIVERSITY PESHAWAR

NAME: M OMAR MASOOD
ID: 14305
SUBJECT: Natural Language Process
DEPARTMENT: BS (SE) 5th Semester
SUBMITTED TO: MAM Aasma

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?

ANS: Natural Language Processing usually shortened as NLP, is a branch of artificial intelligence that deals with the interaction between computers and human using the natural language. The ultimate objective of NLP is to read decipher, understand and make sense of the human language in a manner that is valuable.

Following are the two Application of NLP:

(1)

Machine Translation

- Translating a text from one language to another

(2)

Speech recognition

- Recognizing a spoken language and transforming it into a text

Following are the 2 challenges Of NLP:

(1)

Ambiguity

- One word/sentence can have different meanings
  - Fall
    - The third season of the year
    - Moving down towards the ground or towards a lower position
  - The door is open.
    - Expressing a fact
    - A request to close the door

(2)

Semantics

- The astronomer loves the star.
  - Star in the sky
  - Celebrity

(http://en.wikipedia.org/wiki/Star#/media/File:Starsinthesky.jpg)

(http://www.businessnewsdaily.com/2023-celebrity-hiring.html)

Question No. 2:
(5)

Specify the text strings using the below regular expressions:

a. /a(bc)
   - Given string:

ANS: ab abc ac acb a0b a2b a42c A87d

b. /[abc]
   - Given string:

ANS: ab abc ac acb a0b a2b a42c A87d

c. /abc+
   - Given string:

ANS: ab abc abcc babc

d. /abc*
   - Given string:

ANS: ab abc bcca babc

e. /[^a-z A-Z 0-9]

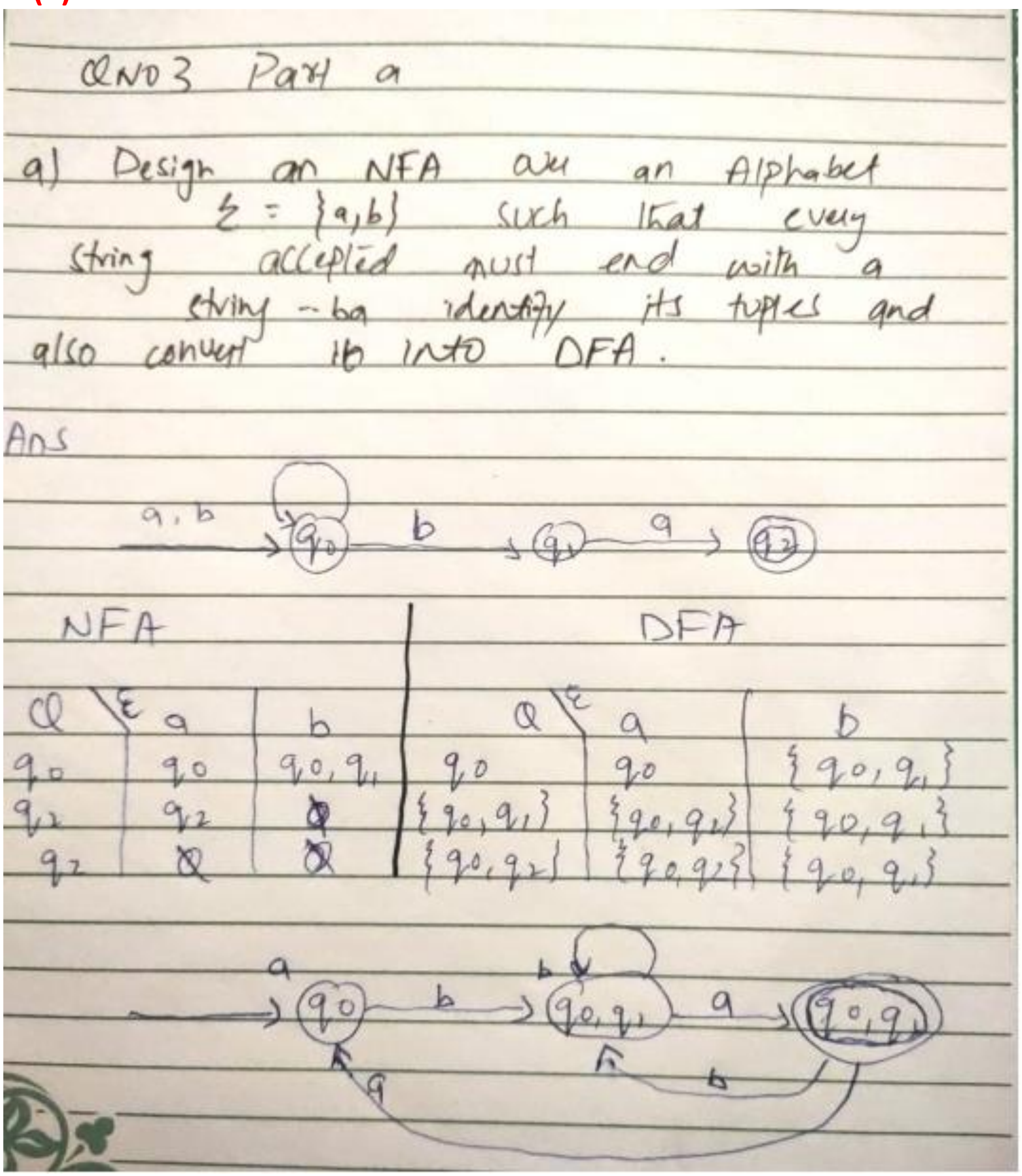
- Given string:

ANS: 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.

ANS(a):


QNo 3 Part a

a) Design an NFA over an Alphabet Σ = {a,b} such that every string accepted must end with a string --ba identify its tuples and also convert it into DFA.

Ans

NFA

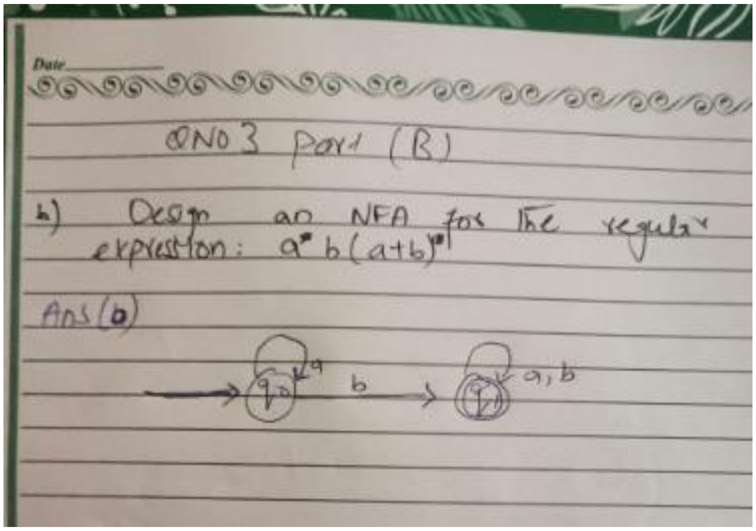
Q \ Σ	a	b
q0	q0	q0, q1
q1	q2	∅
q2	∅	∅

DFA

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

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

ANS(b):



Question No. 4:

(15)

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

ANS(a): 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 Bank1 on the Bank2 on the river Bank3 for transaction Bank1 is verb the other two banks are noun. People jump high People

Eg:

Noun/Verb Jump Noun/Verb High Noun/Adjective List of all possible tags for each word.

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

ANS(b):

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

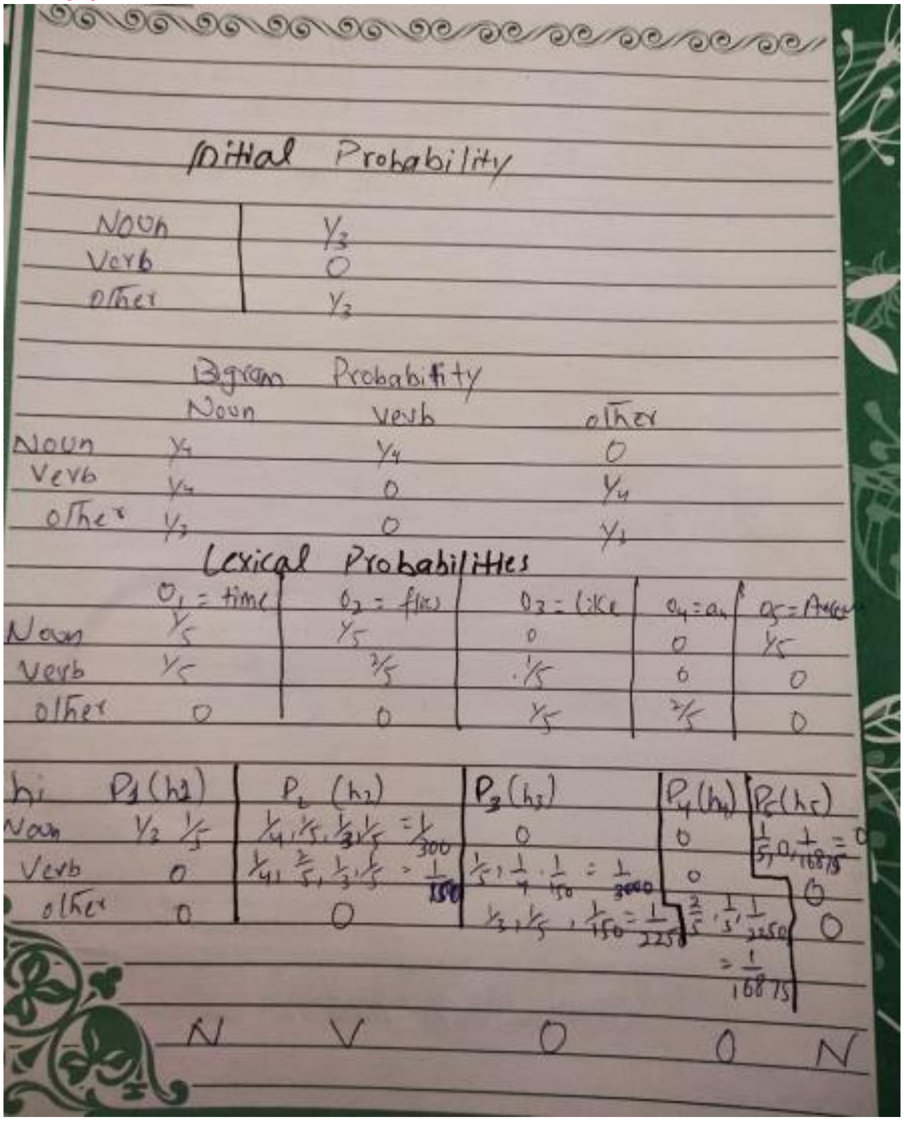
c) Apply Viterbi Algorithm on the below given bigram and lexical probabilities;

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

ANS(c):



Initial Probability

Noun	1/3
Verb	0
other	1/3

Bigram Probability

	Noun	verb	other
Noun	1/4	1/4	0
Verb	1/4	0	1/4
other	1/3	0	1/3

Lexical Probabilities

	O_1 = time	O_2 = flies	O_3 = like	O_4 = an	O_5 = 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

hi	$P_1(h_1)$	$P_2(h_2)$	$P_3(h_3)$	$P_4(h_4)$	$P_5(h_5)$
Noun	1/3 1/5	1/4, 1/5, 1/3, 1/5 = 1/300	0	0	1/5, 0, 1/16875 = 0
Verb	0	1/4, 2/5, 1/3, 1/5 = 1/150	1/5, 1/4, 1/150 = 1/3000	0	0
other	0	0	1/3, 1/5, 1/150 = 1/2250	2/5, 1/3, 1/2250 = 1/16875	0

N V O O N

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.

ANS(a):

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,^)/#^

=0.2x2/2

=0.2

=(NLP,^)/^

=0.1x2/2

=0.1

(book,^)/^

=0x2=0.

= 0.5x2/2 = 0.5

= # (got ,^)/ # ^

= 0.3x2/2 =0.3

= <John,N)/N

=0.5x5/5=0

b) Find the CFG of the string “abaabaa” using the production rules S→a, S→aAS, A→bS

ANS(b):