Name : Imran Khan
I-D: 14281
Subject: NLP

Answer no.1:
Natural Language Processing (NLP) is a part of Artificial Intelligence (AI) in which natural human language can be processed and understood by machines in order to enable data-driven insights to be developed which were unusable before because of the 'casual' Nature of the data - for example, spoken speech or a speech written in heavy accent. Simply, NLP is used to help computers overcome the ambiguity of English or other languages because more and more data is in informally written and spoken formats such as emails and social media Communication, therefore, NLP helps the computer to understand, comprehend and transform this data into meaningful information.

Applications:
1) Speech recognition
Speech recognition software programs that allow us to decode human voice. These include mobile telephony, home automation, hands-free computing, virtual assistants, video games, and more. This technology is being used for other methods of input such as typing, clicking or any other way to select text. for example voice assistants (Cortana, Google Assistant, Siri etc.).
2) Automatic Summarization
Going back to the amount of text data we have every day, information overload can be a real problem but now we have an automated summarization. It is the process of making a short, accurate and fluent summary of a long text document. The most important benefit of using abstract is that it reduces reading time. Here are some APIs you can try: Eileen Text Analysis, Rain Cloud Summarization, ML Analyzer, Text Summary, Text Summary.

Challenges:
1) Breaking the sentences
Formally known as "sentence limit elimination", this process of breaking is no longer difficult to achieve, but nonetheless, is an important process, especially in the case of highly structured data which Contains serious information. Breaking requests should be smart enough to separate paragraphs into appropriate sentence units. However, highly complex data is not always available in easily identifiable sentence formats. These data can be in the form of tables, graphics, notifications, page breaks, etc., which is why the machine needs to take proper action to make sense of the way a person would refer to an interpreting text.
2) Context determination
One of the most important challenge in the entire NLP process is to train a machine to draw context from the discussion within a document. Consider the following two sentences:
"I enjoy working in a bank."
"I enjoy working near a river bank."
The context of these sentences is completely different.

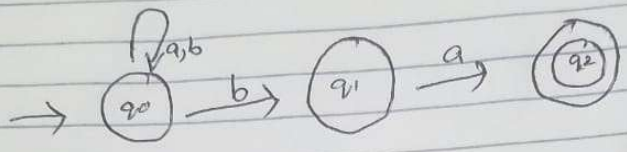
Answer no.2)
a) abc

b) ab abc ac acb an ab ac
c) abc abcc abc
d) ab abc abcc abc
e) $!

Answer no.3A)

Ans 3 :- (A)

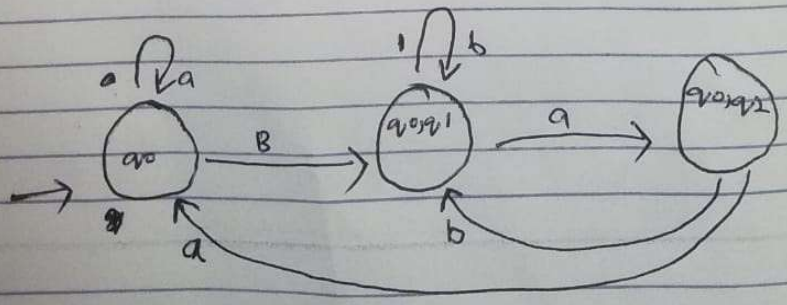
NFA :-



E/q	a	b
q0	q0	q0,q1
q1	q2	–
q2	–	–

DFA :-

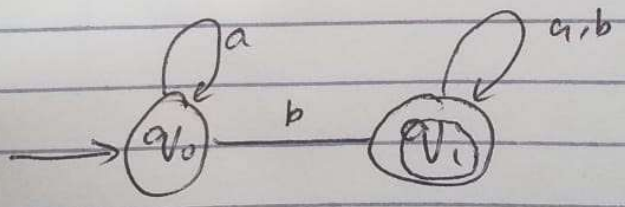
E/q	a	b
q0	q0	{ q0,q1 }
{ q0,q1 }	{q0,q2}	{q0,q1}
{ q0,q2 }	q0	{q0,q1}



Answer no. 3B)

Ans 3, (B)

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



Answer 4 a)

It is defined as the process of assigning a part of a speech to a given word. This is commonly called POS tagging. Simply put, POS tagging is the act of labeling each word in a word with the appropriate part of your speech. We already know that some parts of speech include nouns, verbs, adjectives, adjectives, pronouns, combinations and their subcategories.

Much of POS tagging comes under rule-based POS tagging, Stochastic POS tagging and transformation-based tagging.

For example, suppose if the preceding word of a word is an article then word must be a noun.

POS tag ambiguity:

Tagging is a type of classification that can be defined as the automatic assignment of tokens. The descriptor here is called a tag, which can represent a part of speech, word information and so on.

Answer 4 b)

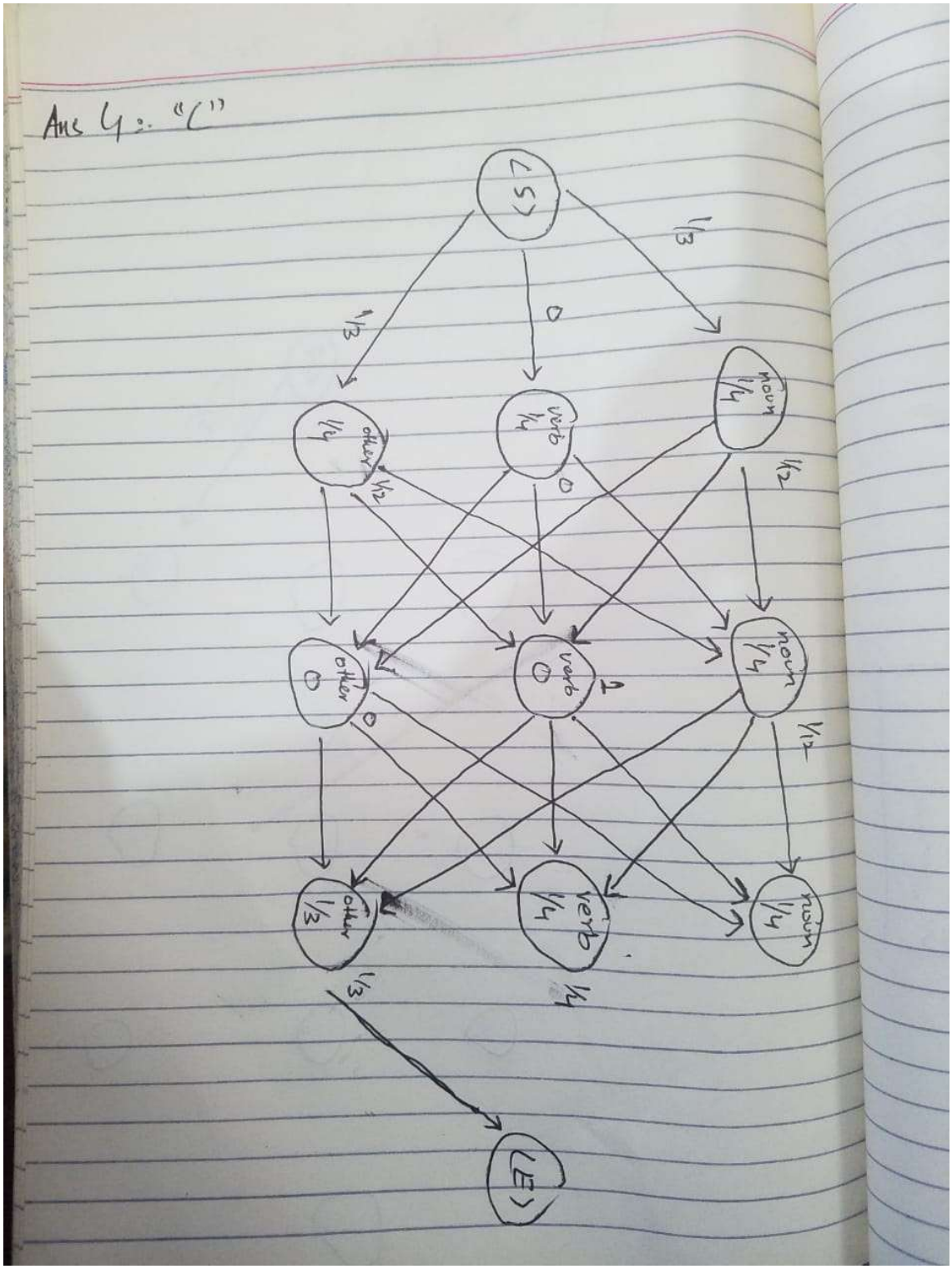
Closed classes:

- determiners: a, an, the
- pronouns: she, he, I
- prepositions: on, under, over, near, by, …

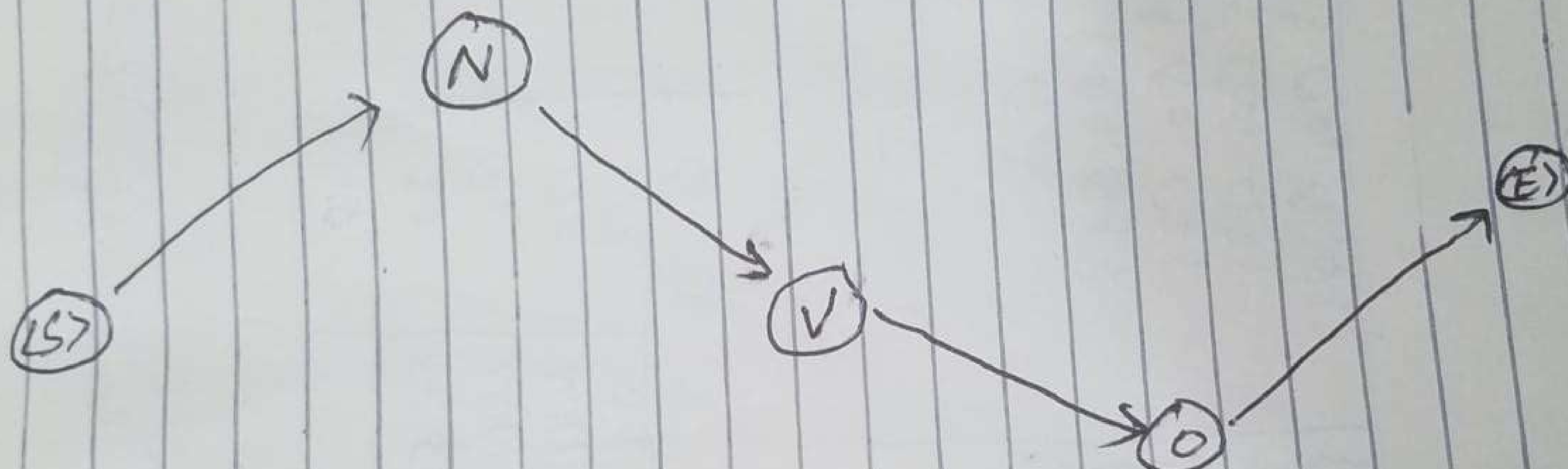
Open classes:

- Nouns, Verbs, Adjectives, Adverbs.

Answer 4 c)


Ans 4: "C"
<S>
1/3
1/3
0
noun 1/4
verb 1/4
other 1/4
1/2
other 0
verb 0
noun 1/4
other 1/3
verb 1/4
noun 1/4
1/3
1/4
<E>

3



Answer 5 a)

(4) Ans $P(T) = P(W|T) = \pi (t_i/t_{i-1}) - P(t_i/t_{i-1})$

$= P(W_i/t_i)$

Corpus = ^John got many AdLi books found all very interesting

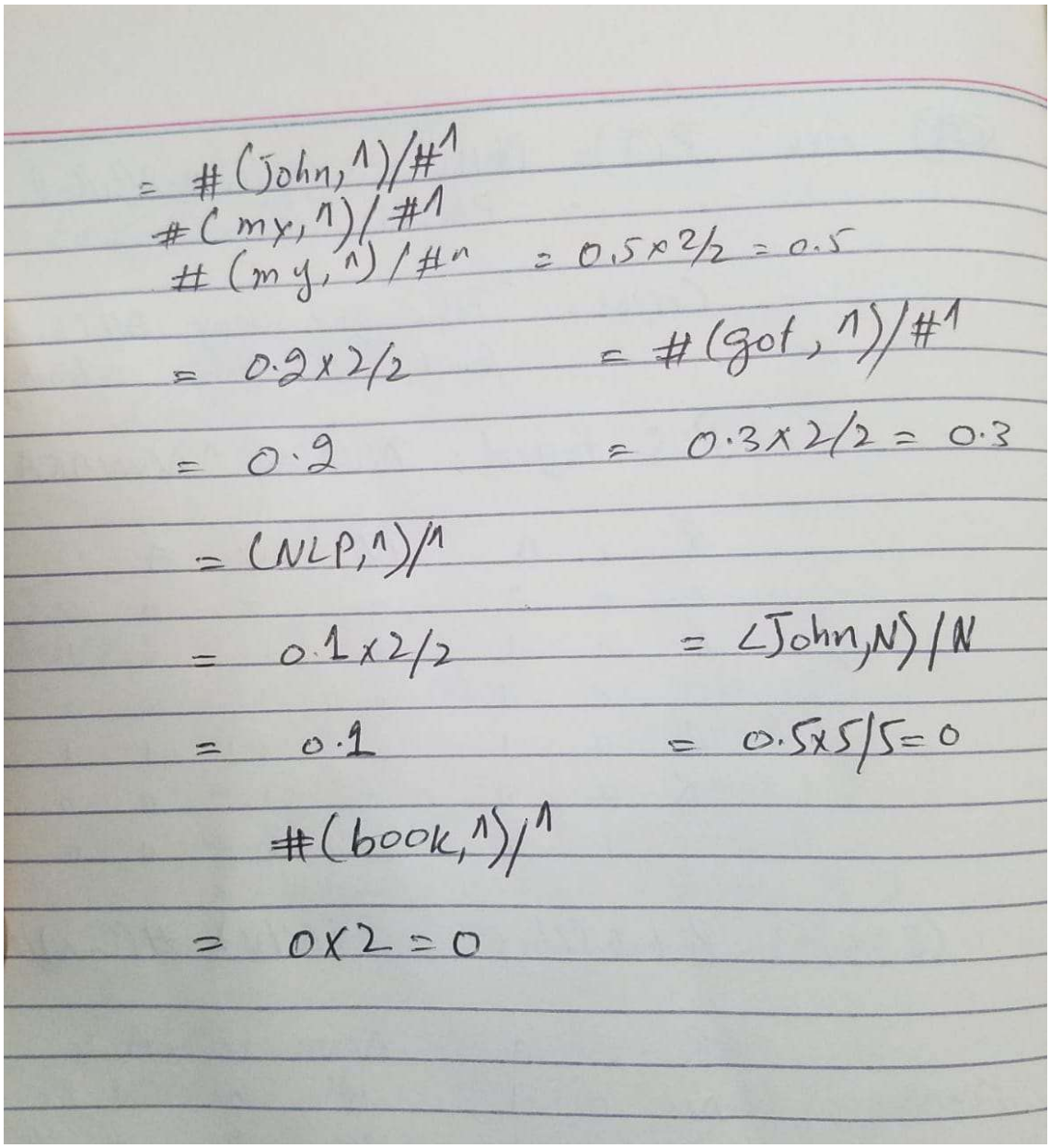
POS tagged: NVNN. ^NVNARA

	R	.	^	N	V	A
^	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

Probabilities $P(N/V) = \#(^{\wedge}=N)/\#^{\wedge}$

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

Books	John	Got	Many	N2P	
^	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	-	-	-	-
.	-	-	-	-	-


= #(John, ^)/#^
#(my, ^)/#^
#(my, ^)/#^
= 0.2 x 2/2
= 0.2
= (NLP, ^)/^
= 0.1 x 2/2
= 0.1
#(book, ^)/^
= 0 x 2 = 0
= 0.5 x 2/2 = 0.5
= #(got, ^)/#^
= 0.3 x 2/2 = 0.3
= (John, N)/N
= 0.5 x 5/5 = 0

Answer 5 b)

5.
Ans (B)

$S = aAS$ (Rule: 2)
$= abS$ (Rule: 3)
$= abaAS$ (Rule: 1)
$= abaabSS$ (Rule: 2)
$= abaabaS$ (Rule: 3)
$= abaabaa$ (Rule: 3)

Left most derivation tree to obtain the String 'w' as follows;