

**MID TERM Assignment (Fall- 2020)**  
**Name: HOORIA KHAN ORAKZAI**  
**ID: 14263**  
**DEPARTMENT: BS(SOFTWARE-ENGINEERING)**  
**SUBJECT: Natural Language Processing**

---

Time: 6 Days  
Instructor: Aasma Khan

Total Marks: 30  
Date: 20<sup>th</sup> April, 2020

---

Note: Attempt all Questions.

Question No. 1: (08)

- a) Briefly explain NLP? Write the name of 2 Applications of NLP with example? Write the name of 2 Challenges of NLP with example? (05)

**Answer:**

✚ **Natural language processing:**

(NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (**natural**) languages, in particular how to program computers to process and analyze large amounts of natural language data. A definitive goal of NLP is to peruse, interpret, comprehend, and understand the human dialects in a way that is important. Most NLP procedures depend on AI to get significance from human dialects.

✚ **Two Application of NLP:**

- Interactive Voice Response (IVR) applications utilized in call centre to react to specific clients' solicitations.
- Machine interpretation (MT), procedure of deciphering one source language or content into another dialect, is one of the most significant utilizations of NLP
- Personal associate applications, for example, Google-Assistant, Siri, Cortana, and Alexa

✚ **Challenges:**

- **Syntax & Ambiguity:**

**Examples:**

I saw a man with a telescope.

- Who had the telescope?

### Semantics:

#### Examples:

The astronomer loves the star

- Star in the sky
- Celebrity

b) Define Phonology and Morphology with the help of example? (03)

### Phonology:

Phonology is the study of how sounds are organized and used in natural languages. It's a field of linguistics which study the distribution of sound in a language as well as the interaction between those different sounds.

#### Example:

An example of phonology is the study of different sounds and the way they come together to form speech and words - such as the comparison of the sounds of the two "p" sounds in "pop-up".

### Morphology

Morphology is the study of morphemes; a morpheme is defined as "the smallest unit of meaning in a language." All words, since they have meaning, have at least 1 morpheme, but a word can have several morphemes.

#### Examples:

For example the word "cat" has just one morpheme but the word "cats" has 2, as the -s denotes plurality.

**Question No. 2:** (10)

a) What do you mean by regular expressions?

#### Answer

### Regular Expressions:

Regular Expression also called regex. It is a very powerful programming tool that is used for a variety of purposes such as feature extraction from text, string replacement and other string manipulations. A regular expression is a set of characters, or a pattern, which is used to find sub strings in a given

string. for ex. extracting all hashtags from a tweet, getting email id or phone numbers etc. from a large unstructured text content.

b) Specify the text strings using the below regular expressions:

1. `/[a-fA-F0-9]`

a. Given string: a89opxcfff

OUTPUT: a89 opx cff

2. `/[abc]`

a. Given string abc ac acb a0b a2b a42c A878

OUTPUT: abc ac acb a0b a2b a42c A878

3. `a(b|c)`

a. Given string abc aa acbaob

OUTPUT: abc aa acbaob

4. `/abc*`

a. Given string ab abc abcc babc abc abcc babc

OUTPUT: ab abc abcc babc abc abcc babc

5. `/abc+`

a. Given string ab abc abcc babc abc abcc babc

OUTPUT: ab abc abcc babc abc abcc babc

6. `/[^a-z A-Z]`

a. Given string Price of cat \$1

OUTPUT: Price of cat \$1

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

a. Given string: a89 opx cfff \$1!

OUTPUT: a89 opx cfff \$1!

8. `/a(bc)`

a. Given string: ab abc ac acb a0b a2b a42c A87d

OUTPUT: ab abc ac acb a0b a2b a42c A87d

9. /a[bc]

a. Given string abc ac acb a0ba2b

OUTPUT: abc ac acb a0ba2b

10.a|b|c

a. Given string: ab abc ac acb a0b a2b a42c A87d

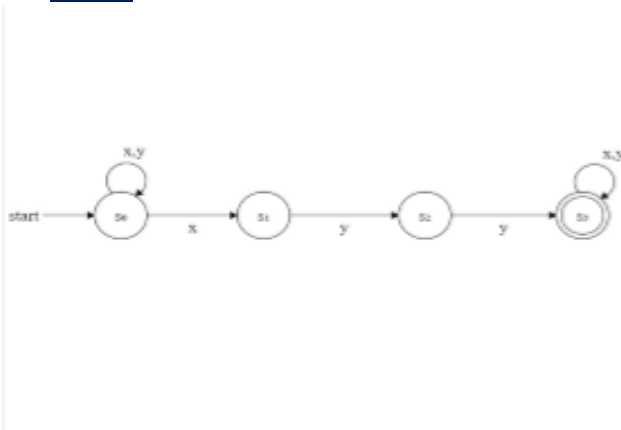
OUTPUT: ab abc ac acb a0b a2b a42c A87d

Question No. 3:

(05)

a) Design an NFA over an alphabet  $\Sigma=\{x,y\}$  such that every string accepted must have a substring --xyy-- ? identify its tuples and also convert it into DFA.

NFA



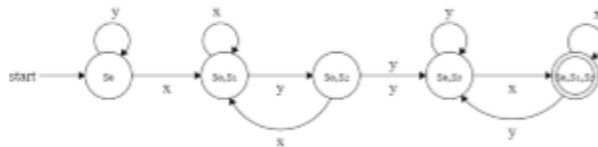
NFA TRANSITION TABLE :

S	X	Y
s0	s0 , s1	s0
s1	Null	s2
s2	Null	s3
s3	s1	s3

DFSA TRANSITION TABLE:

Q	X	Y
→ s0	{s0, s1}	s0
{s0,s1}	{s0,s1}	{s0,s2}
{s0,s2}	{s0,s1}	{s0,s3}
{s0,s3}	{s0,s1,s3}	{s0,s3}
{s0,s1,s3}	{s0,s1,s3}	{s0,s3}

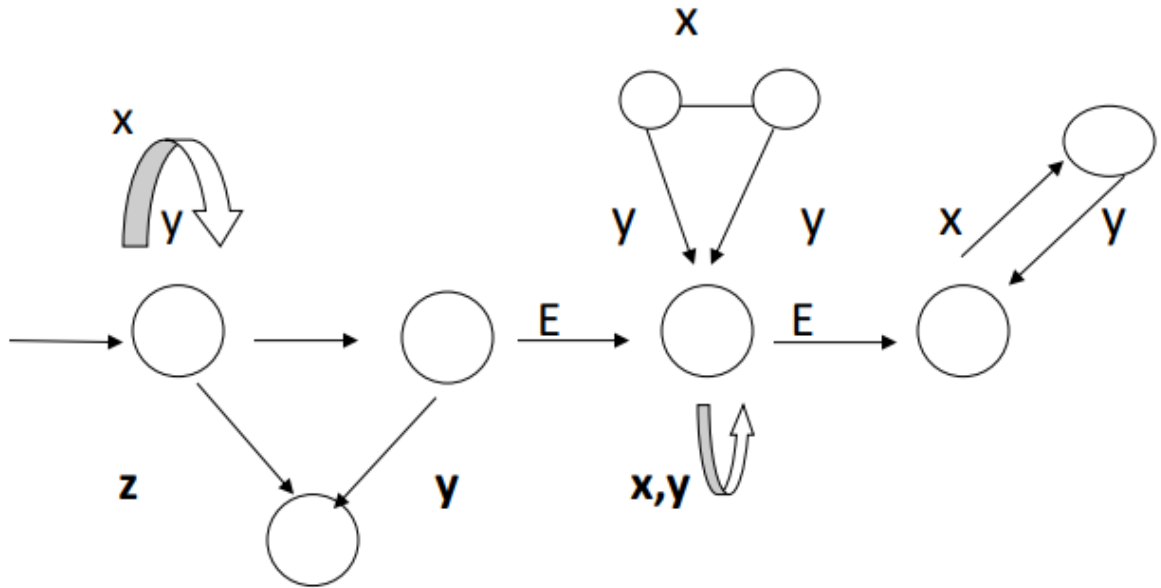
**DFA:**



**Question No. 4:**

(05)

a) Design an NFA for the regular expression :  $(x+y+zx)((yxy)^*+(x+y)^*)(xy)^*$



Question No. 5:

(02)

Find the Maximum Likelihood Estimation of the below according to the given corpus using conditional probability:

<s> The green eyes </s>                      <s> The green jungle </s>  
 <s> The green jungle </s>                      <s> The green eyes </s>  
 <s> The green park </s>                        <s> The green eyes </s>

i. **P(jungle | The green)**

$$\frac{P(\text{Jungle The green})}{P(\text{The Green})} = \frac{2}{6} = 0.3333$$

ii. **P(eyes | The green)**

$$\frac{P(\text{The Green eyes})}{P(\text{The Green})} = \frac{3}{6} = 0.5$$

iii. **P(park | The green)**

$$\frac{P(\text{The Green Park})}{P(\text{The Green})} = \frac{1}{6} = 0.17$$

iv. **P(sea | The green)**

$$\frac{P(\text{The Green Sea})}{P(\text{The Green})} = \frac{0}{6} = 0$$

$$P(S) = \frac{2}{6} \cdot \frac{3}{6} \cdot \frac{1}{6} \cdot \frac{0}{6} = 0$$

Answer