ID:14269 , NAME: QAZI BILAL
SUB:NPL , DEP: BS(SE)

Time: 6 Days	**Total Marks: 30**
Instructor: Aasma Khan	**Date: 20th 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?

ANS: 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.

- The following are 2 applications of NLP.

(1):

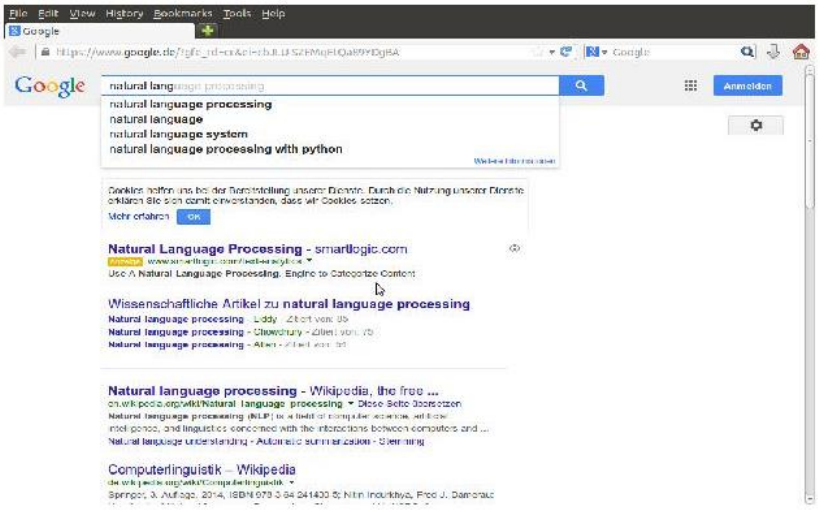
Spell and Grammar Checking

- Checking spelling and grammar
- Suggesting alternatives for the errors

(2):

Word Prediction

- Predicting the next word that is highly probable to be typed by the user



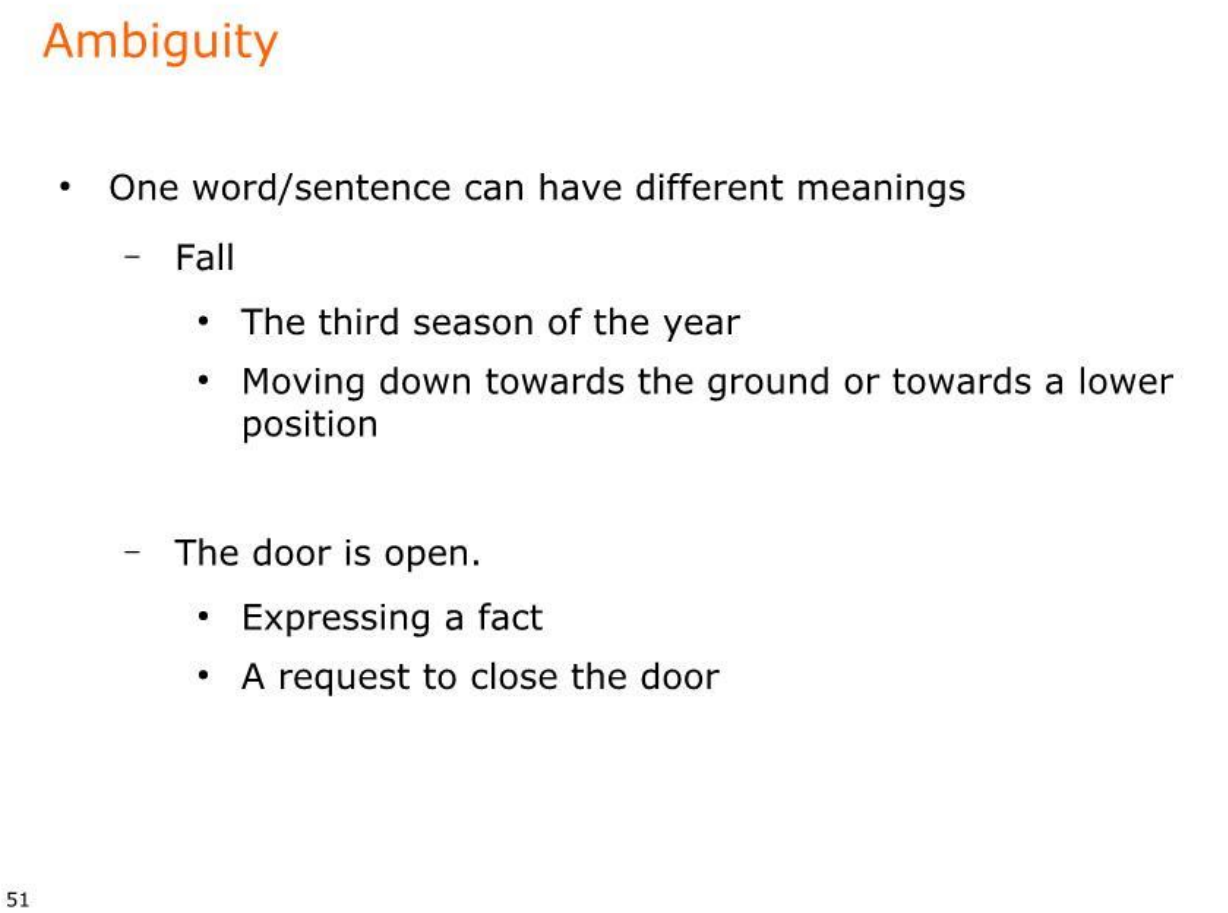
- The following are 2 challenges of NLP:

(1):

Paraphrasing

- Different words/sentences express the same meaning
  - Season of the year
    - Fall
    - Autumn
  - Book delivery time
    - When will my book arrive?
    - When will I receive my book?

(2):



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

ANS:

- Phonology

Phonology concerns the use of sounds in a particular language. English makes use of about 45 phonemes – contrastive sounds, eg. /p/ and /b/ are contrastive because pat and bat mean different things. (Note the use of [x] for a phone and /x/ for the related phoneme.) In Vietnamese these two sounds are not contrastive, so a Vietnemese second language learner of English is likely to have trouble producing and hearing the distinction. Phonemes are not always pronounced the same, eg. the [p] in /pat/ is different from that in /tap/ because the former is aspirated, but aspiration or no aspiration /pat/ still means pat (in English). Some of this allophonic variation in the pronunciation of phonemes in different contexts is a consequence of co-articulation but some is governed by language or dialect specific rules. For example, in American English, speakers are more likely to collapse want to to 'wanna' in an utterance like who do you want to meet? than

British English speakers, who will tend to say 'wanta'. However neither group will phonologically reduce want to when who is direct object of want e.g. who do you want to resign? Phonology goes beyond phonemes and includes syllable structure (the sequence /str/ is a legal syllable onset in English), intonation (rises at the end of questions), accent (some speakers of English pronounce grass with a short/long vowel) and so forth.

- Morphology

Morphology concerns the structure and meaning of words. Some words, such as send, appear to be 'atomic' or monomorphemic others, such as sends, sending, resend appear to be constructed from several atoms or morphemes. We know these 'bits of words' are morphemes because they crop up a lot in other words too – thinks, thinking, reprogram, rethink. There is a syntax to the way morphemes can combine – the affixes mentioned so far all combine with verbs to make verbs, others such as able combine with verbs to make adjectives – programable – and so forth. Sometimes the meaning of a word is a regular, productive combination of the meanings of its morphemes – unreprogramability. Frequently, it isn't or isn't completely eg. react, establishment.

Question No. 2: (10)

a) What do you mean by regular expressions?

- Regular expression:

Regular expressions 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 oxp cfff

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

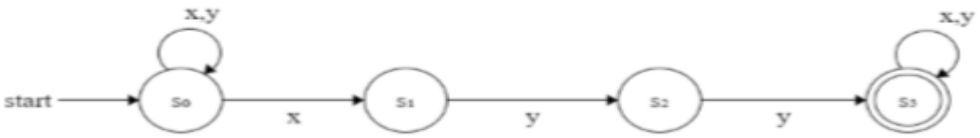
       OUTPUT: price of cat $1

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 ∑={x,y} such that every string accepted must have a substring --xyy-- ? identify its tuples and also convert it into DFA.
NFA



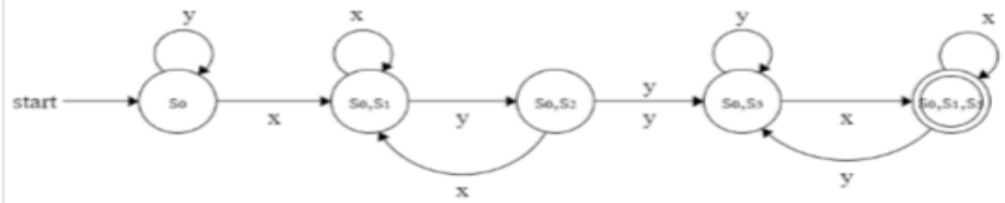
FSA Transition Table:

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

DFSA Transition Table:

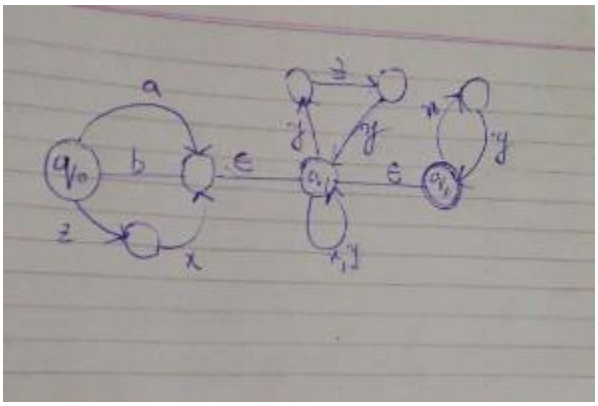
Q	X	Y
→ s0	{s0 , s1}	s0
{s0,s1}	{s0,s1}	{s0,s2}
{s0,s2}	{s0,s1}	{s0,s3}
{s0,s3}	{s0,s1,s3}	{s0,s3}
{s0,s1,s3}	{s0,s1,s3}	{s0,s3}

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))= $\frac{P(\text{The green jungle})}{P(\text{The green})} = \frac{2}{6} = 0.333$

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)ddgz = $\frac{P(\text{The green park})}{P(\text{The green})} = \frac{0}{6} = 0$

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