INDIAN SCHOOL AL WADI AL KABIR

Woksheet No:1	ARTIFICIAL INTELLIGENCE (417) CLASS X CHAPTER 4: NATURAL LANGUAGE PROCESSING

1 Mark questions

1. The _____ domain of Artificial Intelligence, that is focused on enabling computers to understand and process human languages.
   a. Data Science b. Computer Vision c. **Natural Language Processing** d. None of these

2. Which of the following is an application of Natural Language Processing?
   a. Automatic Summarization b. Sentiment Analysis c. Text classification **d. All of these**

3. _____are software applications that mimic written or spoken human speech for the purposes of simulating a conversation or interaction with a real person
   a. Face filter b. Google lens c. **chatbots** d. Facial recognition

4. _____ work around a script which is programmed in them
   a. smart bots b. chat bots c. **script bots** d. robots

5. _____refers to the grammatical structure of a sentence
   a. Syntax b. Semantics c. Object detection d. Normalization

6. Pick the odd one out.
   a. Stemming **b. Localization** c. Lemmatization d. removal of stop words.

7. _____ is a term used for any word or number or special character occurring in a sentence.
   a. corpus b. statement c. **token** d. stop word

8. Which algorithms result in two things, a vocabulary of words and frequency of the words in the corpus?
   a. Sentence segmentation b. Tokenisation **c. Bag of words** d. Text normalisation

9. Which feature of NLP helps in understanding the emotions of the people mentioned with the feedback?
   (a) Virtual Assistants (b) **Sentiment Analysis** (c) Text classification (d) Automatic Summarization

10. What will be the output after stemming the word studies?
    a. study b. studie c. **studi** d. none of these

2 Mark questions

11. Write note on Natural language processing

Natural Language Processing (commonly called NLP) takes in the data of Natural Languages which humans use in their daily lives and operates on this. Natural Language Processing, or NLP, is the sub-field of AI that is focused on enabling computers to understand and process human languages.

12.Write notes on text classification and virtual assistants.

Text classification makes it possible to assign predefined categories to a document and organize it to help you find the information you need or simplify some activities. For example, an application of text categorization is spam filtering in email.

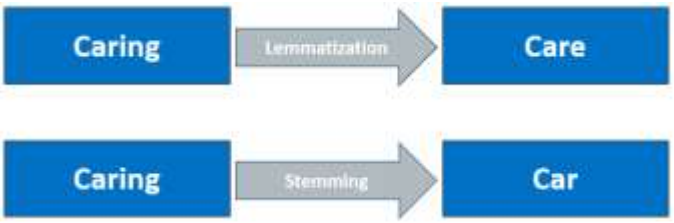
Virtual Assistants: Accessing our data, helps us in keeping notes of our tasks, make calls for us, send messages and a lot more. With the help of speech recognition, these assistants can not only detect our speech but can also make sense out of it. Ex: Google Assistant, Cortana, Siri, Alexa, etc

13. Differentiate Stemming and Lemmatization

Stemming: is the process in which the affixes of words are removed and the words are converted to their base form which is known as **Stem**. Stemming does not consider if the stemmed word is meaningful or not It just removes the affixes hence it is faster.

Lemmatization: is the process in which the affixes of words are removed and the words are converted to their base form which is known as **lemma**. Lemmatization makes sure that lemma is a word with meaning and hence it takes a longer time to execute than stemming.

Eg:



14. Identify any two stop words which should not be removed from the given sentence and why?

Get help and support whether you're shopping now or need help with a past purchase. Contact us at abc@pwershel.com or on our website www.pwershel.com

Stopwords in the given sentence which should not be removed are:
@, . (fullstop),_(underscore) , 123(numbers) These tokens are generally considered as stopwords, but in the above sentence, these tokens are part of email id. removing these tokens may lead to invalid website address and email ID. So these words should not be removed from the above sentence.
(1 mark for identifying any two stop words from the above, and 1 mark for the valid justification.)

15. Write note on Bag of Words Algorithm.

Bag of Words is a Natural Language Processing model which helps in extracting features out of the text which can be helpful in machine learning algorithms.

As we put The normalized corpus which we have got after going through all the steps of text processing, into the bag of words algorithm, the algorithm returns to us the unique words out of the corpus and their occurrences in it. Thus, we can say that the bag of words gives us two things:

1. A vocabulary of words for the corpus

2. The frequency of these words (number of times it has occurred in the whole corpus).

4 Marks Questions

16. Samiksha, a student of class X was exploring the Natural Language Processing domain. She got stuck while performing the text normalization. Help her to normalize the text on the segmented sentences given below:

Document 1: Akash and Ajay are best friends

Document 2: Akash likes to play football but Ajay prefers to play online games

1. Tokenisation

Akash, and, Ajay, are, best, friends, Akash, likes, to, play, football, but, Ajay, prefers, to, play, online, games

2. Removal of stopwords

Akash, Ajay, best, friends, Akash, likes, play, football, Ajay, prefers, play, online, games

3. Converting text to a common case

akash, ajay, best, friends akash, likes, play, football, ajay, prefers, play, online, games

4. Stemming/Lemmatisation

akash, ajay, best, friend, akash, like, play, football, ajay, prefer, play, online, game

(1 mark for each step; 1*4=4)

17. Differentiate script bot and smart bot.

Script-bot	Smart- bot
Script bots are easy to make	Smart-bots are flexible and powerful
Script bots work around a script which is programmed in them	Smart bots work on bigger databases and other resources directly
Mostly they are free and are easy to integrate to a messaging platform	Smart bots learn with more data
No or little language processing skills	Coding is required to take this up on board
Limited functionality	Wide functionality

18. We, human beings, can read, write and understand many languages. But computers can understand only machine language. Do you think we might face any challenges if we try to teach computers how to understand and interact in human languages? Explain.

Yes, we might face any challenges if we try to teach computers how to understand and interact in human languages.
The possible difficulties are:
1. Arrangement of the words and meaning - the computer has to identify the different parts of a speech. Also, it may be extremely difficult for a computer to understand the meaning behind the language we use.
2. Multiple Meanings of a word - same word can be used in a number of different ways which according to the context of the statement changes its meaning completely.
3. Perfect Syntax, no Meaning - Sometimes, a statement can have a perfectly correct
syntax but it does not mean anything. For example, take a look at this statement:
Chickens feed extravagantly while the moon drinks tea.
This statement is correct grammatically but does this make any sense? In Human language, a perfect balance of syntax and semantics is important for better understanding.
(1 mark for Yes and 1 mark each for the points on possible difficulties)

****