

Total No. of printed pages = 4

CSE 1818 PE52

Roll No. of candidate

12/6/22

--	--	--	--	--	--	--	--	--	--

BINA CHOWDHURY CENTRAL LIBRARY
(GMIT & GIPS)
Azare, Hathhawapara,
Guwahati-781017

2022

B.Tech. 8th Semester End-Term Examination

CSE

SPEECH AND NATURAL LANGUAGE PROCESSING

(New Regulations 2017 – 2018)

(New Syllabus 2018 – 2019)

Full Marks – 70

Time – Three hours

The figures in the margin indicate full marks
for the questions.

Answer question No. 1 and any *four* from the rest.

1. Answer the following with best suitable answer (MCQ/ Fill in the blanks) :

(10 × 1 = 10)

(i) The Add-1 smoothing for bigrams is (where N is the total number of tokens in the training set and |V| is the size of the vocabulary)

(a) $P_{Laplace}(W) = (C(W) + 1) / (N + |V|)$

(b) $P_{Laplace}(W_n | W_{n-1}) = C(W_{n-1} W_n) / C(W_{n-1})$

(c) $P_{Laplace}(W_n | W_{n-1}) = (C(W_{n-1} W_n) + 1) / (C(W_{n-1}) + |V|)$

(d) None of this

(ii) The different ambiguities of NLP are

(a) Lexical Ambiguity (b) Syntactic ambiguity

(c) Semantic ambiguity (d) All of these

[Turn over

- (iii) Which are the following actions for lemmatization
- (a) was, am, are, is → be
 - (b) playing, played, plays → play
 - (c) study, studying, studied → study
 - (d) car, car's, cars, cars' → car
- (iv) Computational vs computer is an example of _____ morphology.
- (a) Derivational
 - (b) Inflectional
 - (c) Complex
 - (d) Compound
- (v) Consider the statement "The students went to class". Assign POS tags for the statement.
- (a) DT NN VB P NN
 - (b) DT NN NN P NN
 - (c) NN NN VBG P NN
 - (d) DT NN VB P DT
- (vi) Which of the following model is used for speech recognition?
- (a) Speech model
 - (b) Hidden Markov Model
 - (c) Finite State Transducers Model
 - (d) Grammar Model
- (vii) In the POS tagging problem, what is the output of the Viterbi algorithm?
- (a) Probability of word sequence given a particular tag sequence
 - (b) Optimal transition and observation probabilities for HMM
 - (c) Probability of the best tag sequence given a word sequence
 - (d) None of the above
- (viii) In this sentence: "...no benefits justify the risk of nuclear weapons... I will explain why nuclear technology has a future on our planet despite its dangers". Which type of lexical cohesion can you find?
- (a) Hyponymy
 - (b) Synonymy
 - (c) Antonymy
 - (d) None of the above
- (ix) _____ is/are Python library (ies) to make programs that work with natural language.
- (a) Gensim
 - (b) spaCy
 - (c) CoreNLP
 - (d) NLTK

BINA CHOWDHURY CENTRAL LIBRARY
(GIT & SIPO)
Azara, Haldinowasara,
Gurukul - 781017

(x) Mention the following regular expression that can be used to identify the date(s) present in the text? "The next meetup on SNLP will be held on 2022-04-01, previously it happened on 01/04, 2022"

(a) $\backslash d\{4\}-\backslash d\{2\}-\backslash d\{2\}$

(b) $(19|20)\backslash d\{2\}-(0[1-9]|1[0-2])-[0-2][1-9]$

(c) $(19|20)\backslash d\{2\}-(0[1-9]|1[0-2])-([0-2][1-9]|3[0-1])$

(d) None of the above

2. (a) "The Finite-State transducer (FST) is used to do morphological recognition". Discuss with examples by using two-level morphology. (7)

(b) Differentiate between with example (preferable your own language with gloss)

(i) free and bound morpheme

(ii) Derivational and inflectional morphology. (8)

3. (a) How the rare and unknown words can potentially affect the performance of many NLP systems. (6)

(b) Implement the Bag of Words algorithm using Python. (9)

4. (a) Define Probabilistic Context-Free Grammar (PCFG). Suggest some improvement on different issues of PCFG. (2+6)

(b) A simple PCFG is as follows:

$S \rightarrow NP VP$ 1.0 $NP \rightarrow NP PP$ 0.4

$PP \rightarrow P NP$ 1.0 $NP \rightarrow$ astronomers 0.1

$VP \rightarrow V NP$ 0.7 $NP \rightarrow$ ears 0.18

$VP \rightarrow VP PP$ 0.3 $NP \rightarrow$ saw 0.04

$P \rightarrow$ with 1.0 $NP \rightarrow$ stars 0.18

$V \rightarrow$ saw 1.0 $NP \rightarrow$ telescopes 0.1

Calculate the probability of the sentence "astronomers saw the stars with ears" with the above information. (7)

BINA CHOWDHURY CENTRAL LIBRARY
(GIMT & GIPS)
Azara, Halkhowapara,
Guwahati - 781017

5. (a) Discuss the classical approaches to machine translation with the help of the Vauquois triangle. (6)
- (b) What is Neural Machine Translation? Briefly explain a pipeline of NMT with an example. (9)
6. (a) What is coreference resolution? Illustrate the steps for pronoun resolution by example. (2+6)
- (b) What is discourse structure? Analyze the recovering different issues on discourse structure. (2+5)
7. (a) Discuss Hidden Markov Model (HMM) considering POS tagging with a proper diagram. (8)
- (b) Optimize the HMM with Viterbi Algorithm. (7)
8. Write short notes on any *three*: (3 × 5 = 15)
- (a) Python Libraries used for NLP
- (b) Transfer Learning for NLP
- (c) NLTK
- (d) Accuracy, Precision and Recall.

BINA CHOWDHURY CENTRAL LIBRARY
(GIMT & GIPS)
Azara, Halkhowrapara,
Guwahati - 781017