KATHMANDU UNIVERSITY
End Semester Examination
July, 2018

Level: BE                                                    Course: COMP 473
Year: IV                                                     Semester: II

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

Exam Roll No:                        Time: 30 mins                        FM: 10

Registration No:                                                          Date:

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

SECTION "A"
[20 Q x 0.5 = 10 marks]

Tick the most appropriate answer.

1. Regular expressions are used to represent which language
   A. Recursive language
   B. Context free language
   C. Regular language
   D. All of the above

2. Which of the following statement is true?
   A. Every language that is defined by regular expression can also be defined by finite automata
   B. Every language defined by finite automata can also be defined by regular expression
   C. We can convert regular expressions into finite automata
   D. All of the above

3. The RE in which any number of 0's is followed by any number of 1's followed by any number of 2's is:
   A. (0+1+2)*
   B. 0*1*2*
   C. 0* + 1 + 2
   D. (0+1)*2*

4. How many tuples is a Finite State Machine comprised of?
   A. 4
   B. 5
   C. 6
   D. Unlimited

5. The basic limitation of finite automata is that:
   A. It can't remember arbitrary large amount of information
   B. It sometimes recognizes grammars that are not regular
   C. It sometimes fails to recognize regular grammar
   D. All of the above

6. How many different lexemes are there in the following 5-tuple?

 <man, men, girls, girl, mouse>
   A. 1
   B. 2
   C. 3
   D. 4

7. In the English language inflectional morphemes can be:
   A. Prefixes, Suffixes and Infixes
   B. Suffixes only
   C. Infixes only
   D. None of the above

8. How many tokens are there in the following text?
*They picnicked by the pool, then lay back on the grass and looked at the stars.*

 A. 17
 B. 15
 C. 16
 D. 18

9. Given the following n-gram orders and the corresponding perplexity measures, which of the following is a better model?

| N-gram order | Unigram | Bigram | Trigram |
|---|---|---|---|
| Perplexity | 962 | 170 | 109 |

   A. Unigram   B. Bigram   C. Trigram   D. None of the mentioned

10. Select the correct part-of-speech for the word in italics in the text below:
He whispered *very* quietly in the library.
A. Verb   B. Conjunction  C. Adverb D. Preposition

11. Which of the following is not a stop word in English?
A. these   B. also   C. into   D. capital

12. Pointing to a tree, young Ramal says, "bird flied away". Ramal's interesting, but inaccurate use of the "ed" word ending shows that he is trying to learn _____rules of language.
A. syntactical   B. semantic C. pragmatic   D. morphological

13. Top down parser generates:
A. Leftmost derivation
B. Rightmost derivation
C. Rightmost derivation in reverse
D. Leftmost derivation in reverse

14. An example of a "relatively useless" question is:
A. Why are people so unkind?
B. Where can I learn to swim?
C. Why didn't you wait for me?
D. What would you like to do today?

15. What do discourse analysts study?
A. Forms of communication other than talk
B. The way discourse "frames" our understanding of the social world
C. The rhetorical styles used in written and oral communication
D. All of the above

16. Querying of unstructured textual data is referred to as:
A. Information access   B. Information update    C. Information manipulation   D. Information retrieval

17. The process of designing and implementing software to transparently manage different linguistic and cultural conventions without additional modifications is:
A. Software internationalization    B. Software Localization    C. Globalization D.None of the mentioned.

18. Which of the following statement is/are true for Word2Vec model?

A. The architecture of word2vec consists of only two layers – continuous bag of words and skip-gram model

B. Continuous bag of word is a shallow neural network model

C. Skip-gram is a deep neural network model

D. Both CBOW and Skip-gram are deep neural network models

19. Which of the following models can be used for the purpose of document similarity?

A. Training a Word2Vec model on the corpus that learns context present in the document

B. Training a bag of words model that learns occurrence of words in the document

C. Creating a document-term matrix and using cosine similarity for each document

D. All of the above

20. Collaborative Filtering and Content Based Models are the two popular recommendation engines, what role does NLP play in building such algorithms.

A. Feature extraction from text

B. Measuring Feature Similarity

C. Engineering Features for vector space learning model

D. All of the above

## Answer Keys to Objective Questions

1. C
2. D
3. B
4. B
5. A
6. C
7. B
8. D
9. C
10. C
11. D
12. D
13. A
14. A
15. D
16. D
17. A
18. C
19. D
20. D

Level: BE                                                            Course: COMP 473
Year: IV                                                             Semester: II
Time: 2 hrs 30 mins                                                  FM: 40

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

## SECTION "B"
[6 Q. x 4 = 24 marks]

Attempt ANY SIX questions

1. A sentence can easily have more than one parse tree that is consistent with a given CFG. How do PCFGs and non-probability-based CFGs differ in terms of handling parsing ambiguity? [4]

2. The preposition "of" is semantically ambiguous and, in fact, has many different possible meanings. Consider the following four noun phrases:
   *house of brick*
   *bowl of cherries*
   *deck of cards*
   *height of the horse*
   For each of the noun phrases, describe the meaning that is intended. [4]

3. Noun phrase co-reference resolution includes **pronoun** resolution, **proper noun** resolution, and **common noun** resolution. Which of the three would you expect to be the most difficult to handle computationally? Explain why. [4]

4. Define precision and recall. Explain their relationship to each other diagrammatically. [4]

5. The following shows a simple context-free grammar (CFG) for a fragment of English.

```
S -> NP VP          Adj -> angry        Vbe -> is
VP -> Vbe Adj       Adj -> big          N -> dog
NP -> Det N         Adj -> former       N -> cat
N -> Adj N          P -> at
Adj -> Adj PP       P -> on
PP -> P NP          Det -> the
```

Show the parse tree that this grammar would assign to "the dog is angry at the cat." [4]

6. Briefly discuss "lexical" and "syntactic" ambiguity with suitable examples. [4]

7. What is the "naïve" assumption behind Naïve Bayes and does this assumption hold for language-based classification tasks? [4]

[2 Q. x 8 = 16 marks]

Attempt *ANY TWO* questions

8. What is anaphora resolution?  Explain the following terms with suitable examples: [2+6]

   a. Reference   b. Referring expression   c. Referent  d. Coreference

9. Define Localization. Shed light on the scope and key concepts of software localization. [2+6]

10. Explain the vector-space model in Information Retrieval. How are term-frequency (TF) and inverse-document frequency (IDF) employed to retrieve the relevant documents? [4+4]