WORKSHEET-1 ANSWER KEY

NLP

1. (A) Lexical Processing
   (B) Syntactic processing
   (D) Semantic processing
2. (A) Spam- Ham classification
3. (A) Breaking the text in to words
   (B) Removing Stopwords
4. (A) word_tokenize()
   (B) sent_tokenize()
5. (A) [“I”, “love”, “#”, “food”, “#”, “pasta”]
6. (B) [“I”, “love”, “#food”, “#pasta”]
7. (A) They provide no useful information, especially in applications such as spam detector or search engine.
   (B) Since the frequency of stopwords is very high, removing stopwords results in a much smaller data.
   (C) removing stopwords results in faster computation.
8. (A) spam-ham classifier building
9. (A) It takes in to consideration of only the words present in the text and not the order of the words.
10. (B) (2,5)
11. (A) The importance of a word in a document becomes more if it is present exclusively only in this document.
12. (C) 0
13. (D) 0.159
14. (C) tf-idf of both “vapour” and “Bangalore” are equal to zero.
15. (A) The bow model gives equal importance to all the words while tf-idf model gives more importance to those words in a document which occurs exclusively only I this document.