**WORKSHEET-1**

**NLP (Solutions)**

**ANS 1)** A (Lexical Processing), B (Syntactic processing), D (Semantic processing)

**ANS 2)** A (Spam- Ham classification), C (Chat-Bot building), D(Question- Answering system)

**ANS 3)** A(Breaking the text in to words), B (Removing Stopwords)

**ANS 4)** A (word\_tokenize()), B (sent\_tokenize())

**ANS 5)** A [“I”, “love”, “#”, “food”, “#”, “pasta”]

**ANS 6)** B [“I”, “love”, “#food”, “#pasta”]

**ANS 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.

**ANS 8)** A (spam-ham classifier building)

**ANS 9)** A (It takes in to consideration of only the words present in the text and not the order of the words.)

**ANS 10)** B (2, 5)

**ANS 11)** A (The importance of a word in a document becomes more if it is present exclusively only in this document)

**ANS 12)** A (0.2)

**ANS 13)** B (.89)

**ANS 14)** B (The tf-idf score of “vapour” is less than tf-idf score of “Bangalore” in document 1)

**ANS 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 in this document.)