**UT2 Question bank**

**Module 4 – Syntax Analysis (Topics - HMM , Viterbi)**

1. Given a corpus with the following sentences having three tags - Noun, Verb, and Modal

Nitish loves CampusX.

Can Nitish google campusX?

Will Ankita google CampusX

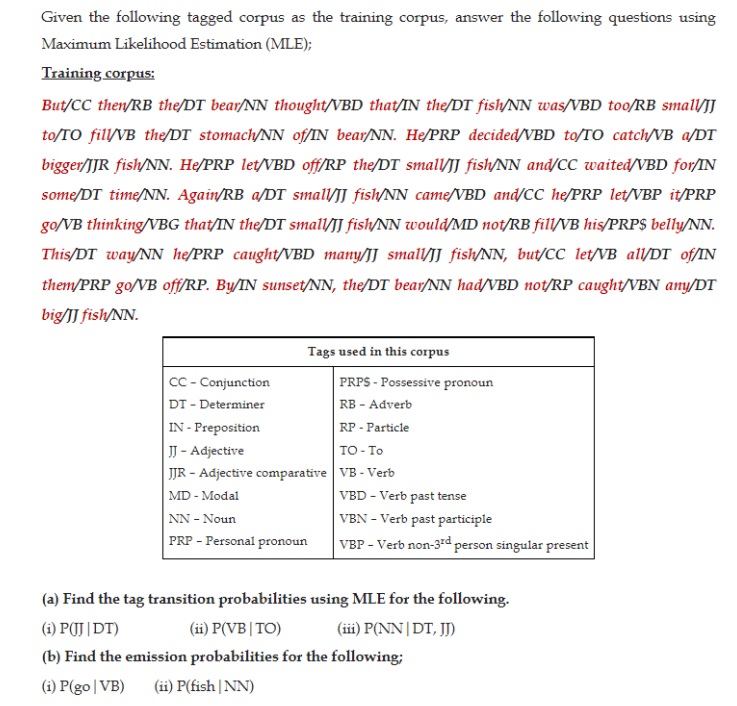
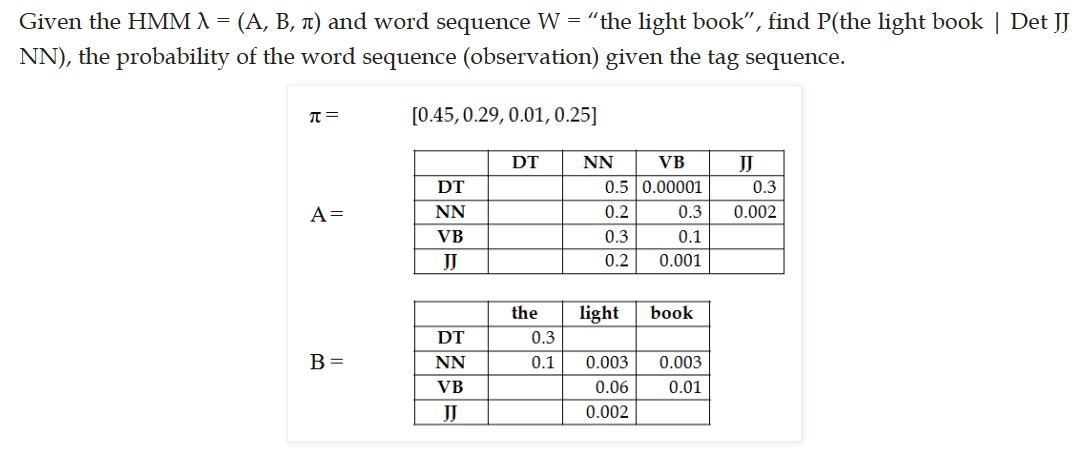
Ankita loves will

Will loves google

1. Assign correct tagging to the given corpus.
2. Construct Emission probability and Transition probability Matrix.

Apply HMM model to find the probability of P(will will google campus?| M N V N).

M- Modal, N-Noun, V- Verb.

1. 
2. Given the training corpus

*Mary Jane can see will*

*The spot will see Mary*

*Will Jane spot Mary?*

*Mary will pat Spot*

1. Construct emission probability and transition probability matrix.
2. Apply Viterbi algorithm to find correct tagging for the test sentence. “**Will can spot Mary”.**

**Pg 46**

(similar questions will be asked)

**Module 5 – Semantic Analysis**

1. We have a document database of five documents with the following content.

D1: “Information Retrieval System”

D2: “Information Storage”

D3: “Digital Speech Synthesis System”

D4: ”Speech filtering”

D5: “Speech retrieval”

1. Calculate Term frequency matrix for each term in a corpus.
2. Calculate inverse document frequency (IDF) for each term.
3. Calculate TF-IDF for the term “Speech” and explain how the term “**Speech**” is related to all the documents. Which document is most relevant to the term “Speech”?
4. Explain the following terms with a suitable example: Synonym, Antonym, Hyponym , Hypernym, Meronym.

* **Identify** the relationship between the following synsets. Justify your answer.

1. Couch – Sofa
2. Car – wheel
3. Meal – breakfast
4. **I left my heart—and my suitcase.**
5. Mammal – Dog
6. Given two documents

**A : “Jupiter is the largest planet”**

**B: “Mars is the fourth planet from the earth”. Represent both documents in a vector form using ….**

1. the traditional word count method and estimate the similarity between the two documents using cosine similarity.
2. the TF-IDF method and estimate the similarity between the two documents using cosine similarity.
3. Analyze and comment on the results of methods (a) and (b)
4. Are there any other similarity measures available? If yes, explain those measures and compare two documents using this new similarity measure for (a).
5. what is semantic analysis? why semantic analysis difficult? Explain various approaches to semantic analysis.
6. What is WordNet? How is “sense” defined in WordNet? Explain with example.
7. What do you mean by word sense disambiguation (WSD)? Discuss knowledge based WSD.

**Module 6:**

1) Define discourse & pragmatic analysis . Discuss reference resolution problem in detail.

2) Discuss following referring expressions with suitable examples w.r.t reference phenomena

Indefinite NPs , Definite NPs, Pronouns, Demonstratives and anaphora.

3) Explain three types of referents that complicate the reference resolution problem.

4) Write a note on Syntactic and Semantic Constraints on Coreference.

5) Discuss various Preferences in Pronoun Interpretation with suitable example.

**Applications: (Practical based questions will be asked)**

1. Sentiment Analysis
2. Text Summarization