**NLP CS563: Midsem Quiz**

|  |  |  |
| --- | --- | --- |
| **Name**: M. Maheeth Reddy | **Roll No.**: 1801CS31 | **Date**: 23-Feb-2022 |

**---------------------------------------------------------------------------------------------------------------------------------**

**Ans 1**: **a) PoS Tagging**

**Words Tags Words Tags**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| With | Others |  | tactics | Noun |
| Wriddiman | Noun |  | to | Others |
| Saha | Noun |  | get | Verb |
| gaining | Verb |  | an | Others |
| support | Noun |  | interview | Noun |
| from | Others |  | the | Others |
| the | Others |  | Indian | Adjective |
| cricket | Noun |  | cricket | Noun |
| fraternity | Noun |  | board | Noun |
| after | Others |  | BCCI | Noun |
| putting | Verb |  | is thinking | Verb |
| out | others |  | of | Others |
| a | others |  | hiring | Verb |
| WhatsApp | Noun |  | an | Others |
| chat | Noun |  | agency | Noun |
| showing | Verb |  | to | Others |
| an | Others |  | deal | Verb |
| unnamed | Adjective |  | with | Others |
| journalist | Noun |  | such | Adjective |
| using | Verb |  | cases | Noun |
| strong | Adjective |  | in | Others |
| arm | Noun |  | its | Others |
|  |  |  | ecosystem | Noun |

**b) Named Entity Tags**

With Wriddhiman(**Person**) Saha(**Person**) gaining support(**Others**) from the cricket(**Others**) fraternity(**Others**) after putting out a WhatsApp(**Organization**) chat(**Others**) showing an unnamed journalist(**Person**) using strong-arm tactics(**Others**) to get an interview(**Others**), the Indian(**Location**) cricket(**Others**) board(**Organization**) (BCCI(**Organization**)) is thinking of hiring an agency(**Organization**) to deal with such cases(**Others**) in its ecosystem(**Location**).

**Table for Named Entity Tags**:

**Words Tags Words Tags**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| With | Others |  | tactics | Others |
| Wriddiman | Person |  | to | Others |
| Saha | Person |  | get | Others |
| Gaining | Others |  | an | Others |
| support | Others |  | Interview | Others |
| from | Others |  | the | Others |
| the | Others |  | Indian | Location |
| cricket | Others |  | cricket | Others |
| fraternity | Others |  | board | Organization |
| after | Others |  | BCCI | Organization |
| putting | Others |  | is thinking | Others |
| out | Others |  | of | Others |
| a | others |  | hiring | Others |
| WhatsApp | Organization |  | an | Others |
| chat | Others |  | agency | Organization |
| showing | Others |  | to | Others |
| an | Others |  | deal | Others |
| unnamed | Others |  | with | Others |
| journalist | Person |  | such | Others |
| using | Others |  | cases | Others |
| strong | Others |  | in | Others |
| arm | Others |  | its | Others |
|  |  |  | ecosystem | Location |

**c) *I saw a man with a telescope***

The given example “I saw a man with a telescope” has Syntactical Ambiguity. Syntactical Ambiguity is basically the possession of more than one possible structure for the same string of words.

Here, the sentence “I saw a man with a telescope” has two meanings:

Meaning 1: I saw a man who has a telescope.

Meaning 2: I saw a man using a telescope.

There is an ambiguity in structure syntax regarding the telescope: It is not clear about who is having the telescope: The speaker? or the man?

This is the reason the ambiguity is called Syntactical Ambiguity.

**Ans 2**:

**Vanila PCFG can’t resolve all kinds of syntactic ambiguities**. The reason is

1. In the vanilla PCFG algorithm the probability of the productions inferred are not specific and very general.
2. This is because the probabilities of productions are not reliant on words or concepts.
3. For example, there is a preference for attaching prepositional phrases to nominals.

Hence, vanilla PCFG cannot resolved ambiguities that require semantics to resolve.

**Example**: Consider the following sequences:

1. “write with book”
2. “write with pen”

We know that sequence 1 is incorrect, but vanilla PCFG will consider both as correct.

**How to resolve such ambiguities?**

We need to lexicalize PCFG productions to solve this problem:

1. The productions must be specialized to specific words by including their head words in the respective LHS non-terminals (for example we can have VP-ate).
2. The head words in this case can be the verb for a verb phrase, main noun for the noun phrase, preposition for the preposition phrase.
3. For lexicalized productions we can include the head word and the POS of each non-terminal for the non-terminal’s symbol. For solving this efficiently we can use the Collins’ parser.

**Ans 3**:

**a)**

Coreference resolution is an NLP task that involves determining all referring expressions that point to the same real-world entity. A referring expression (i.e., a mention) is either a noun phrase (NP), a named entity (NE), or a pronoun, which refer to an entity in the real world known as the referent

Coreference resolution typically requires a pre-processing pipeline comprising a variety of NLP tasks (e.g., tokenization, lemmatization, named entity recognition, part-of-speech tagging). Historically, these tasks are addressed before training the coreference resolution model (in a pre-processing stage) and, consequently, errors made by pre-processing models impact coreference resolution models, which typically assume that the information provided from this pre-processing stage is correct.

Named entity recognition (NER) is a subtask of information extraction that seeks to locate and classify named entities in text into predefined categories such as the names of persons, organizations or locations . Identifying the token-level boundaries of mentions (such as named entities) is a necessary step toward obtaining mention clusters from natural text; thus, NER is tightly related to coreference resolution.

Inversely, if we have details about coreference resolution then it can help us in NER, because if we know that two words are correlated then their NER result would also be same. So if we know NER result for one such word, we know its result for all words correlated to it.

**b)**

In local context information, features like local content, person prefix and corporate suffix, etc. are used. However, there are contexts where in the the prefixes and suffixes occur in non-local places.

**Example**:

1. Elon Musk is the CEO of SpaceX Inc.

2) Elon Musk was named to be the CEO of key player of the fast paced space exploration, SpaceEx Inc.

In both these contexts, the initcaps are the same where Elon Musk is linked to SpaceX. However, the placement of the elements makes it difficult for performing NER with local context only. In this case, modelling of non-sequential global contextual features will give better performance.