

Compiler mini project

Title – Part of Speech (POS) tagging.

Features –

1. Used stemming for verbs
2. Input to POS tagger is in devnagari script
3. Identifies verb,noun,pronoun,adjective,adverb,preposition and conjunction (POS support)
4. Simple and complex sentence identification
5. POS tagging for Marathi language

Regular expressions used -

1. `(.)*` {return NOUN};
2. For common marathi words-

खुप	{return ADVERB;}
की	{return CONJUNCTION;}
मग	

Productions used -

1. Simple or compound sentence

```
sentence:  simple_sentence { printf("Parsed a simple sentence.\n"); }  
          |  compound_sentence { printf("Parsed a compound sentence.\n"); }  
          ;
```

2. Structure for simple sentence

```
simple_sentence: subject object VERB  
              ;
```

3. Structure for compound sentence

```
compound_sentence:  
    |  compound_sentence CONJUNCTION simple_sentence  
    |  simple_sentence CONJUNCTION simple_sentence  
    ;
```

4. Subject

```
subject:  NOUN  
          |  PRONOUN  
          ;
```

5. Object

```
object:
```

| object ADVERB
| object POSSESSIVE_ADJECTIVE
| object ADJECTIVE
| object NOUN
| PRONOUN
| PREPOSITION
;