Performance Comparison of Different Methods used for Stars Extraction:

Step 1: Tagging the Stars in the sentence

Methods Used:

1. Naïve Method – Find all the Nouns and Proper Nouns and remove those which do not lie in the dictionary
2. Stanford NER Tagger – The sentences were tagged with Stanford NER Tagger
3. ML Based Approach – Create own model using contextual features for tagging.

Performance:

|  |  |  |
| --- | --- | --- |
|  | Precision | Recall |
| Naïve Method | 0.93 | 0.78 |
| Stanford NER | 0.99 | 0.52 |
| ML Method | 0.89 | 0.98 |

Step 2: String Matching Using Modified Levenshtien Distance

Algorithm Description: The tagged parts were matched with the database to infer the actual star. A metric based on combination of string matching and Levenshtien Distance was used to find the star in the database.

Performance:

|  |  |  |
| --- | --- | --- |
|  | Precision | Recall |
| Naïve Method | 0.76 | 0.70 |
| Stanford NER | 0.98 | 0.49 |
| ML Method | 0.91 | 0.89 |
| Ground Truth | 0.96 | 0.90 |