Performance Comparison of Different Methods used for Movies Extraction:

Step 1: Tagging the Movie in the sentence

Methods Used:

1. Naïve Method – Find all the Nouns and Proper Nouns.
2. Rule Based Approach – Rule based Chunks were created to identify movies
3. ML Based Approach – Create own model using contextual features for tagging.

Performance:

|  |  |  |
| --- | --- | --- |
|  | Precision | Recall |
| Naïve Method | 0.614 | 0.685 |
| Rule Based Method | 0.842 | 0.691 |
| ML Method | 0.845 | 0.751 |

Step 2: String Matching Using Modified Levenshtien Distance

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

Performance:

|  |  |  |
| --- | --- | --- |
|  | Precision | Recall |
| Naïve Method | 0.516 | 0.723 |
| Rule Based Method | 0.732 | 0.725 |
| ML Method | 0.762 | 0.779 |
| Ground Truth | 0.864 | 0.961 |