

CS839 –Project Stage 1 Report

Team Members:

- Sreyas krishna Natarajan(snatarajan6@wisc.edu)
- Sugadev Chellakkannu(chellakkannu@wisc.edu)
- Vignesh Thirunavukkarasu(thirunavukka@wisc.edu)

Entity Type:

- We extracted the Person names (cricket players) for the articles got from the CricBuzz repository(webpage for the Cricket Sport).
- person names are marked up in the documents using `<person>...</person>` tags
- Eg. `<person>Virat Kohli</person>` is the captain of India

Documents

- **Overall:**
Number of Documents: 305
Number of Mentions: 1817
- **Set I:**
Number of Documents: 202
Number of Mentions: 1212
- **Set J:**
Number of Documents: 103
Number of Mentions: 605

Classifier:

- Initial
- Final

Precision, Recall, F1 on Set J

Additional Information:

We have **35 features in total**. We increased our accuracy by adding additional features by analysing the documents