NLP- Project Review-1

Team ID- 5

Team members-
18BCE2522 Shashank Shukla
18BCE2526 Mihir Agarwal
Slot- G2+TG2

Our title for NLP project is **Sentence extraction in recognition textual entailment task**. The problem statement is Solving RTE (Recognizing textual entailment) problem using sentence extraction to cover semantic variation by extracting subject, predicate and object from each sentence.
Dataset used is Third Pascal Recognizing Textual Entailment Challenge (RTE-3) dataset. It has approximately has 800 pairs text (T) and hypothesis (H) with labels as True or False showing whether T entails H or not.

For sentence extraction, following methodology is followed-

- Preprocessing- Formation of parse tree using Stanford NLP library.
- Sentence extraction- Used part of speech and parse tree for sentence extraction.
- Part of sentence extraction- Subject, Predicate and Object- Used part of speech tag and syntactic parse tree.
- Feature extraction and classifier- TF-IDF is used for word weighing and feature table is formed.

After feature extraction, any classification algorithm can be used to classify whether the text and hypothesis are entailed or not. Naive Bayes and WEKA can also be used which is used in the reference paper.