Project Description

1. Problem description

Our project is a sentiment detection system, which tag each sentence positive, negative or neutral.

2. Proposed Solution and Implementation Details

- **Corpus**
  Collect 100 to 1000 sentences and manually assign each sentence to sentiment tag, which is positive, negative, and neutral

- **Baseline system (naïve strategy)**
  Look for certain keywords that are indicators of certain sentiment (positive, negative or neutral)

- **Improvement strategy**
  - **Lexical Features**
    - **Word tokens:** In baseline system, we focus on single word. By improvement strategy, we focus on all related words in a sentence.
      Example: He is not happy. The word "happy" is positive but two words "not" "happy" make the sentiment to be negative.
    - **Lemmatization:** To reduce inflectional forms and sometimes derivationally related forms of a word to a common base form.
      Example: The word "hated" cannot be tagged as negative. After lemmatization, the word "hate" could be tagged.
  - **Syntactic Features**
    - **POS tagging:** POS tags could be for identifying and treating differently the different meaning of polysemous words.
      Example: The word "loathing" has different tags. In "Fear and Loathing in Las Vegas", "loathing" is a noun, which makes the sentiment of this sentence to be neutral.
    - **Syntactic parsing:** If the sentiment word is in PP and connected with nominal, which will not affect the sentiment of the sentence. If the sentiment word in PP and connected with VP, which will affect the sentiment of the sentence.
      Example: "This is a book of happiness." In this sentence, "book" is a nominal and "of happiness" is a PP.

"He returned with regret." In this sentence, "return" is a VP and "with regret" is a PP.

- **Semantic Features**
  - **Thematic relations:** Used to express the role that a noun phrase plays with respect to the action or state described by a sentence's verb.
    Example: The word "hate". If the action (hate) has connection with the noun phrase, which makes the sentiment of the sentence to be negative.
  - **Semantic relations:** tokens in a sentence have same concept and relationship.
    Example: Synonymy. If a token has a positive synonymy, which makes the sentence to be positive. Vice versa.

3. Programming tools

Python 3.5, NLTK (WordNet, treebank), The Stanford Parser

4. Architectural diagram