

# Analysis of Articles Based on Parameters Dictated by Human Behavior Towards Literature

Bonani Hazarika  
Manisha Digra  
Manisha Dudi

## 1 Idiom Detection

### 1.1 Requirements:

1. input.txt
2. idiom\_list.txt : Idiom database

### 1.2 Files:

1. newline.py : Rearranges the text removing extra spacing.
2. idiom\_final.py : Detects idioms in input file.
3. percent.py : Finds the percentage occurrence of idioms in input file.
4. simile.py : Detects similes and calculates its percentage occurrence.

## 2 Ngram

### 2.1 Requirements:

1. input.txt

### 2.2 Files:

1. newline.py : Rearranges the text removing extra spacing.
2. give\_input.py : Removal of stopwords, duplicates, punctuations and integers and tokenizes into words.
3. getngrams.py : Receives ngram data corresponding to the words in the input file and provide output in the form of csv file.
4. ngram\_maxfreq.py : Calculates the year in which a particular word had maximum usage.
5. ngram\_result.py : Counts the number of words corresponding to 1800s and 1900s and gives the maximum.

## **3 Metaphor**

### **3.1 Requirements:**

1. input.txt

### **3.2 Files:**

1. newline.py : Rearranges the text removing extra spacing.
2. Similarity.py : Gives subject-object pairs. Calls “GetDependencyParse.py” and “parse.py”
3. lch\_sim.py : Calculate the lch similarity.
4. Hyper\_Pair : Detects metaphor and returns final metaphor percentage in the input text.

## **4 Final percentage**

1. rating.py : Gets percentages from all the above three sections and provides the analysis.