

Cognitive Computing and NLP

Assignment 3

Problem Statement : Write a program that reads several news articles or blogs online or from a text file and tells the user what category the article is about

Objectives :

- 1> Study and explore IBM Watson NLU API
- 2> To learn concepts of text analysis and sentiment analysis with classification

Theory :

IBM Watson NLU api: This Watson API is used to analyze various features of text content at scale. Provide text, raw HTML or a public URL and IBM Watson NLU will give you results for the features you request.

Features of text analysis:

- 1) **Categories**: Returns a five-level taxonomy of the content. The top three categories are returned.
- 2) **Concept**: Returns high-level concepts in the content.
- 3) **Emotion**: Detects anger, disgust, fear, joy or sadness, that is conveyed in the content.
- 4) **Entities**: Identifies people, activities, organizations and other entities.
- 5) **Keywords**: Return important keywords in the content.
- 6) **Metadata**: Returns the information from the document, including authors name, title etc.
- 7) **Relation**: Recognizes when two entities are related and identifies the type of relation.

- 8) Semantic roles : Parses sentences into subject, action and object form.
- 9) Sentiment : Analyze the general sentiment of given content or the sentiment toward a specific target phrases
- 10) Syntax : Returns information about the tokens and phrases in the input text.
- 11) Summarization: Returns a summary of the text

Platform : IBM Watson cloud, JSON and 64-bit
open source Linux

Input : Text document, news articles, online blogs.

Output : Classification of input files in different categories.

Conclusion : Hence, learned the concepts of text analysis and sentiment analysis and classified the input document in respective categories.

FAQ

Question 1

Sentiment Analysis : It refers to the use of NLP, text analysis, computational linguistics and biometrics to systematically identify, extract, quantify and study affective states and subjective information.

Question 2

- 1) **Rule based system :** Classifies text into organized by using a set of hand crafted rules.
- 2) **ML based :** Instead of hand crafted rules, it makes classification based on past observation. Algo: Naive Bayes, SVM
- 3) **Hybrid :** Combines ML based classifier with a rule based system.