**Abstract:**

In order to enable users to quickly view a summary about a topic searched , a system is proposed in which given an input query topic, the best relevant tweets are retrieved by the system using speech acts and from which the noise removal is done and then an abstract summary is formed. This requires the use of summarization algorithms from tweets and it also requires the formation of sentences based on information got from tweets.

Domain:Machine Leaning ,Natural Language Processing

**Problem Statement:**

To build a system that helps to get a topic oriented summarization from tweets and gains information from them .In this system we plan on using basic speech acts like question, suggestion, comment, suggestion, miscellaneous to find the tweets containing information and using relevance between the tweets

**Applications:**

It could be used to generate abstractive summaries for news articles , technical articles, evaluative text, and briefings.

**Literature Review:**

1. **Automatic Twitter Topic Summarization With Speech Acts [Renxian Zhang, Wenjie Li, Dehong Gao, and You Ouyang]**

This paper proposes a speech act-based approach to Twitter topic summarization.The approach to topic summarization is employed in three Core modules : recognizing speech acts in tweets, extracting speech act-guided key words/phrases,and generating abstractive summaries

**2.Sequential Summarization: a Full View of Twitter Trending Topics [Dehong Gao, Wenjie Li, Xiaoyan Cai, Renxian Zhang and You Ouyang]**

Sequential summarization, which aims to provide a serial of chronologically ordered short sub-summaries for a trending topic in order to provide a complete story about the development of the topic while retaining the order of information presentation. Two approaches, i.e., stream-based and semantic-based approaches, are developed to detect the important subtopics within a trending topic. Then a short sub-summary is generated for each subtopic. In addition, we propose three new measures to evaluate the position-aware coverage, sequential novelty and sequence correlation of the system-generated summaries.

**3. Comparing Twitter Summarization Algorithms for Multiple Post Summaries[David Inouye\* and Jugal K. Kalita+]**

In this paper , a Frequency based summariser (involving Sum Basic & hybrid TF-IDF) was proved to provide better results than the other 6 summariser’s used. Eight different summarizers :random, most recent, MEAD, TextRank, LexRank, cluster,Hybrid TF-IDF and SumBasic.A threshold was used to find out similarity.

**System Design:**

**1)Data Collection:**

Using Twitter Api to get the tweets which are used for our Topic Summarisation

**2)Preprocessing**

**a)Removal of Stop Words:**

Stop words are removed

**b)Translation of cases**

All words are translated into lower cases for better results.

**3)Speech act Recognition**

**Feature Set Design:**

**a)Word based**

eg:Question:What,Who ,Why etc..

Getting the word based features on the speech acts

**b)Symbol based**

?-Question and #xyz would be used to get more information

**Classification algorithm**

**a)Data Preparation**

A suitable data set for a supervised classification algorithm is used.

A Classification algorithm based on machine learning is used.

**4)Speech act guided keyword/phrase extraction**

**a)Noise Resistant Phrase Extraction**

In this module we extract the key phrases which are of importance to the topic

**b)POS based Phrase Extraction**

**c)Phrase/Word Ranking**

Assign a rank to a phrase and see its importance.

**5)Topic Summarisation**

**a)Hash tag info:**

Apply a method to use hash tag for information related to the topic by splitting it up .

**b)Template Design:**

Insert the information so far gathered and putting them in slots of a template .

**c)Summarization Algorithm:**

A summarization algorithm is used to summarize all the sentences so far extracted from the text.

**Final Work to be done:**

Given a topic a user should view an abstract summary of the Topic.