



## **Tweet categorization and image retrieval using proposed feature extraction method with subspace clustering algorithms**

**M. Vadivukarassi<sup>1\*</sup>, N. Puviarasan<sup>2</sup>, P. Aruna<sup>3</sup>**

<sup>1,3</sup>Department of Computer Science and Engineering,

<sup>2</sup>Department of Computer and Information Sciences, Annamalai University,  
Annamalainagar, Tamilnadu, India

**Abstract** : The massive amount of user generated content about the real world events is accumulated by social media at every minute. Twitter has gained tremendous popularity with more than 500 million tweets per day for the past few years. The twitter data can be monitored through the Twitter Streaming API. This paper proposed a new feature descriptor method on clustering techniques to improve the performance of the image retrieval system from Twitter dataset. The real time dataset is downloaded from Twitter Streaming API using Python script. From these dataset, the tweets and images are extracted and stored in the separate database. The tweets are analyzed using Naïve Bayes algorithm for tweet categorization. Here these tweets are categorized into different category and the top most categories are detected for further process. Then toppest category tweets are filtered separately and the toppest keyword is detected from these filtered tweets. Based on the keyword detection, the images are retrieved using the proposed Seg\_SIFT feature extraction method and subspace clustering algorithms. These algorithms are compared and analyzed based on the elapsed time of image retrieval based on the performance measures such as precision, recall and accuracy. The detailed experimental results show that the elapsed time of k-subspace is lesser than the seq-k-subspace clustering algorithm while retrieving the images from the database. Seq-k-subspace clustering algorithm performs better than the k-subspace clustering algorithm. This proposed work enables us to discover and understand the tweets and images easily.

**Keywords** : micro blogging; Twitter; Seg\_SIFT; k-subspace clustering; seq-k- subspace; social media.

**M. Vadivukarassi et al //International Journal of ChemTech Research, 2018,11(08): 325-346.**

DOI= <http://dx.doi.org/10.20902/IJCTR.2018.110841>

\*\*\*\*\*