# Event extraction from Tweets

***Abstract— The increasing use of Twitter as a platform for real-time information sharing has led to the need for effective event extraction from tweets. Extracting events from the noisy and unstructured nature of tweets presents challenges that need to be addressed. In this report, we propose a methodology for event extraction from tweets, which involves data collection, data pre-processing, event detection, and event classification. We outline the steps involved in detail and highlight the potential applications and impact of our proposed methodology in social media analytics.***

## I. INTRODUCTION

Twitter is a popular social media platform where users share their opinions, thoughts, and news in real-time. With over 330 million active monthly users, Twitter has become a valuable source of information for various applications such as news monitoring, disaster management, and marketing analysis. However, extracting relevant information from Twitter is challenging due to the high volume, velocity, and variety of data.

Event extraction from Twitter involves identifying and extracting events from tweets, along with their relevant attributes such as type, location, and time. This task is important for various applications such as tracking disasters, monitoring news, and analyzing social trends. However, event extraction from Twitter is challenging due to the informal nature of the language used in tweets, the high volume of data, and the presence of noise and irrelevant information.

Natural language processing (NLP) techniques such as named entity recognition, part-of-speech tagging, and sentiment analysis have been widely used for event extraction from Twitter.

## II. DATASET USED

The dataset used is a collection of tweets related to medical research studies involving the drug acalabrutinib. The dataset used contains textual data in the form of tweets. Each entry includes information such as the tweet author, tweet text, and links to relevant articles or resources.

## III. METHODOLOGY

There are several approaches to event extraction from tweets, including rule-based, machine learning, and hybrid methods.

Some of the key techniques used in event extraction from tweets include named entity recognition, semantic role labelling, and temporal analysis. Named entity recognition involves identifying entities such as people, organizations, and locations mentioned in tweets.

Steps involved:

- Raw data Analysis
- Conversion of Raw data to Dataframe
- Defining functions for future uses
- Translating Tweets to the English language
- Data Pre-processing (Cleaning Tweets)
- Extracting Entities from the tweets with their frequency
- Data Visualization

## IV. RESULTS

Visualizing the Entities and their count using graphs for better insights. I created a new data frame with the entities and their frequency count in descending order. Put

the file location and with help of the to_csv function, saved it in the directory.

## V. CONCLUSION

Data Cleansing and Entities Extraction were a crucial part of the code. Removing unnecessary elements makes tweets ready for further processes, and This part of the process consumes most of the time. After then, Entities Extraction is a key point in the code. Finding the right entities defines extracting the right phrases from the tweets on which the core meaning of tweets relies on.

## VI.REFERENCES


https://www.cse.iitd.ac.in/~mausam/papers/kdd12.pdf

https://aclanthology.org/E17-1076.pdf

https://paperswithcode.com/task/twitter-event-detection

https://link.springer.com/chapter/10.1007/978-3-642-31178-9_32