

# Analysis of Different Representations of Twitter Data

- Problem Definition & Motivation:
  - Analyzing query time and space complexity of different vectorized representations of twitter data.
- Related Work:
  - Keyword Extraction using Natural Language Processing (NLP)
  - COVID-19 Case Study
- Data Models & Algorithms:
  - CountVect - Basic keyword extraction.
  - TF-IDF - Measures frequency and relevance of keywords.
  - Brown Clustering - Clustering words with similar meaning.
  - Query: Select tweets containing user-passed keywords

# Experiments and Results

- 1) Test query runtimes of each representation with different query size
- 2) Test relationship between number of tweets in dataset and size of data representation
- 3) Test relationship between number of tweets in dataset and query time

