



# Twitter Keyword Search

Junyou Chi & Hao Zuo



# Objective

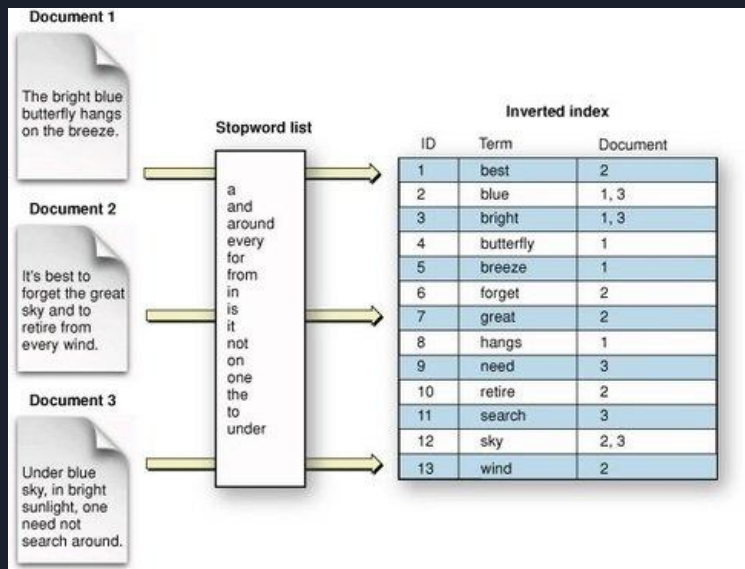
1. Create a keyword search engine that accepts a text file containing a number of tweets as input.
2. Provides a simple user interface for querying the list of tweets against keywords.



# Procedures

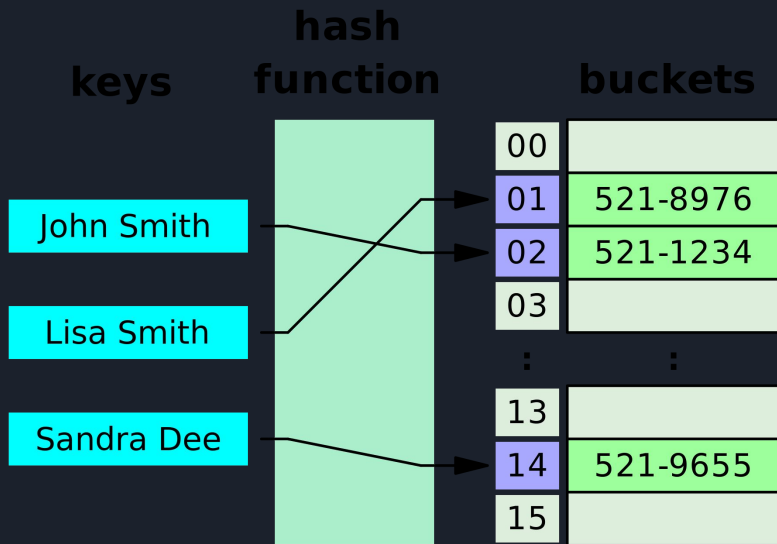
1. Handling the raw data
  - a. Get tweeters from internet(tweepy with python maybe)
  - b. Twitter segmentation if needed
  - c. Word segmentation
2. Build database
  - a. Using inverted index to store twitter feeds
  - b. Using hash map to store key words for inverted index
3. Build searching function to find and sort twitter feeds
4. Build GUI

# About Inverted Index



An inverted index (also referred to as a postings file or inverted file) is a database index storing a mapping from content, such as words or numbers, to its locations in a table, or in a document or a set of documents (named in contrast to a forward index, which maps from documents to content). The purpose of an inverted index is to allow fast full-text searches, at a cost of increased processing when a document is added to the database.

# About Hash Map



In computing, a hash table (hash map) is a data structure that implements an associative array abstract data type, a structure that can map keys to values. A hash table uses a hash function to compute an *index*, also called a *hash code*, into an array of *buckets* or *slots*, from which the desired value can be found. During lookup, the key is hashed and the resulting hash indicates where the corresponding value is stored.