Trending Analysis on Twitter

Team Three:

- Kuan Hu
- Ningtong Bai

Use Cases

- Use Case 1: Generate top 10 daily trending topics without any input
- Use Case 2: For search key words that user inputs
  - ➢Simulated top popular topics associated with the searched key words
  - ➢return top mentioned adjective/descriptive words associated with the searched key words
- (Further Improvements) Customized topic recommendations to each user

General Methodologies

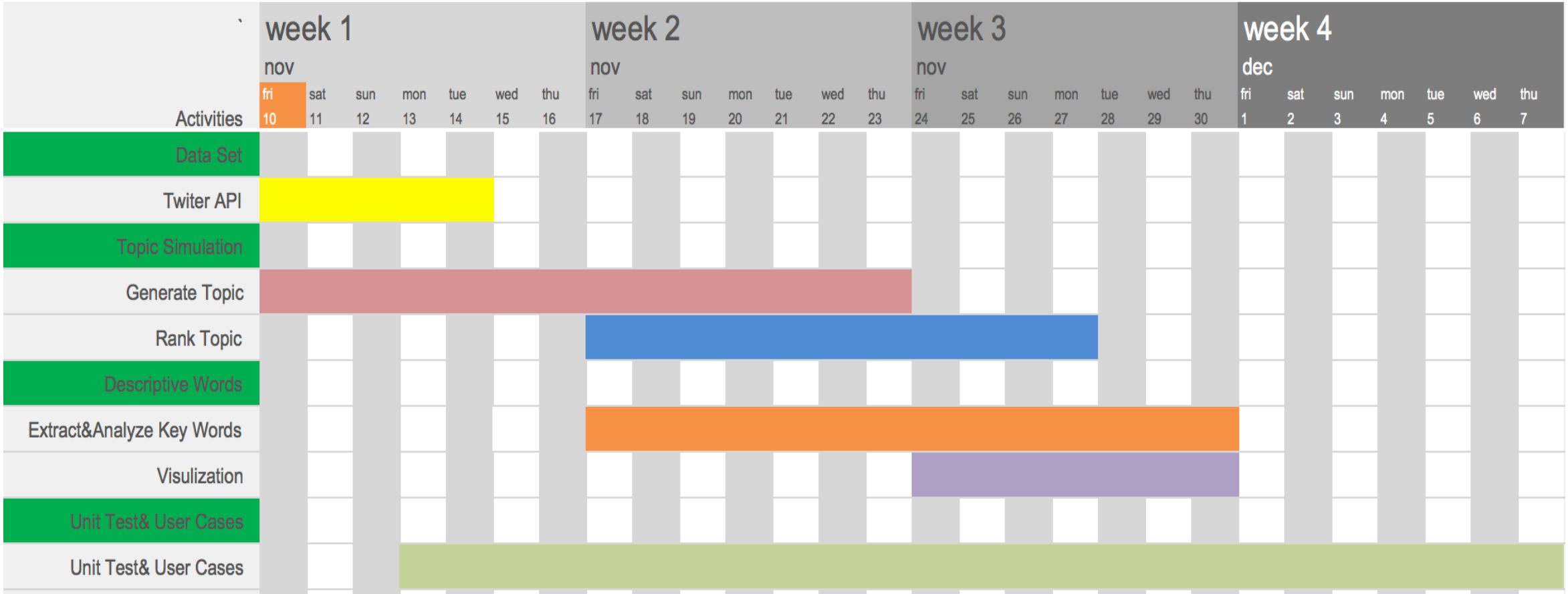
- Topics Generator
  - ✓ Extract key words from tweets
  - ✓ Generate topic based on key words and hashtag (supplementary)
- Rank Topics
  - ✓ Rank trending topics based on how many people like on the relevant tweets
  - ✓ Especially consider tweets from celebrities and influential medias
- Adjective Extractor
  - ✓ Analyze adjectives/descriptive words on relevant tweets
  - ✓ Visualize

Data Source – Twitter API

- Use search/stream API to collect tweets
- API Rate Limits
  - Limits are divided into 15 minutes interval
  - 180 calls every 15 minutes for GET requests
- Magnitude:
  - ❑ 100,000 tweets in the past 7 days

Weekly Milestones

Start Date: 11/10/17



Scala and Code Repository

- Code Repository: Github
  - ✓https://github.com/KuanHu/TopHitsAnalysisOnTwitter
- Scala will be used in :
  - Download data through Twitter API
  - Key words Filter
  - Natural Language Processing Library
  - Unit Tests

Acceptance Criteria

- Ensure topic generated reasonably from tweets
    - Acceptance Criteria: 60% accuracy
- Rank Top 10 trending topics
    - Compare results to hashtag popularity
    - Acceptance Criteria: 60% accuracy
- Adjective/Descriptive Words Extractor
    - Acceptance Criteria:
        - ✓ Characteristics of the words are correctly determined
        - ✓ At least five descriptive words would be extracted

Project Goals

- Provide straight-forward information about trending news/topics/things to users
- Simulate popular topics relevant to user's searched words
- Analyze the big data to generate desired outcome
- Learn how to use data analysis techniques and languages, such as Spark, NLP and Scala to solve real-world big data problems