Trending Analysis on Twitter

Team Three:

- Kuan Hu

- Ningtong Bai

Use Case

- Use Case 1: Generate top 10 daily trending topics with empty input

- Use Case 2: For every Search Key word 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

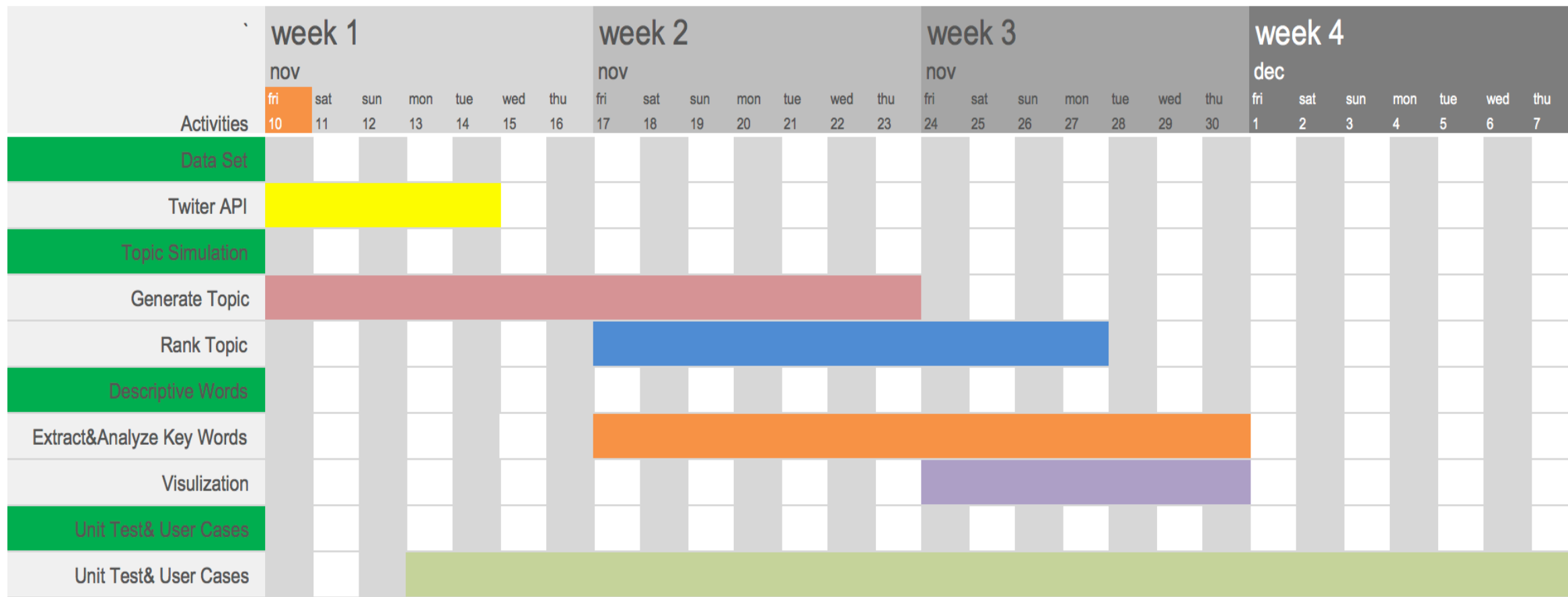- (Further Improvements) Customized topic recommendations to each user

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

Activities	week 1 nov							week 2 nov							week 3 nov							week 4 dec						
	fri 10	sat 11	sun 12	mon 13	tue 14	wed 15	thu 16	fri 17	sat 18	sun 19	mon 20	tue 21	wed 22	thu 23	fri 24	sat 25	sun 26	mon 27	tue 28	wed 29	thu 30	fri 1	sat 2	sun 3	mon 4	tue 5	wed 6	thu 7
Data Set																												
Twiter API																												
Topic Simulation																												
Generate Topic																												
Rank Topic																												
Descriptive Words																												
Extract&Analyze Key Words																												
Visulization																												
Unit Test& User Cases																												
Unit Test& User Cases																												

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

  o Acceptance Criteria: 60% accuracy

- Rank Top 10 trending topics

  o Compare results to hashtag popularity

  o Acceptance Criteria: 60% accuracy

- Adjective/Descriptive Words Extractor

  o 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