

## Introduction

- The software development process is directed by humans and is subsequently affected by its human workers
- Angry or sad workers may affect the progress of a company's project
- Difficult for big companies to gauge how employees may feel

## Project Objectives

- Use sentiment analysis to extract the sentiment of developers' commit messages and comments, and compare it to the success of the build
- Create a visual representation, in the form of a calendar, from the data extracted

## Sentiment Analysis

- Computational process of determining whether the opinion expressed in a piece of text is positive, negative, or neutral<sup>1</sup>
- Used to determine how people feel about a certain topic
- Plays an important role in social media and other business initiatives
- Used to improve customer service and relations

<sup>1</sup>Oxford English Dictionary

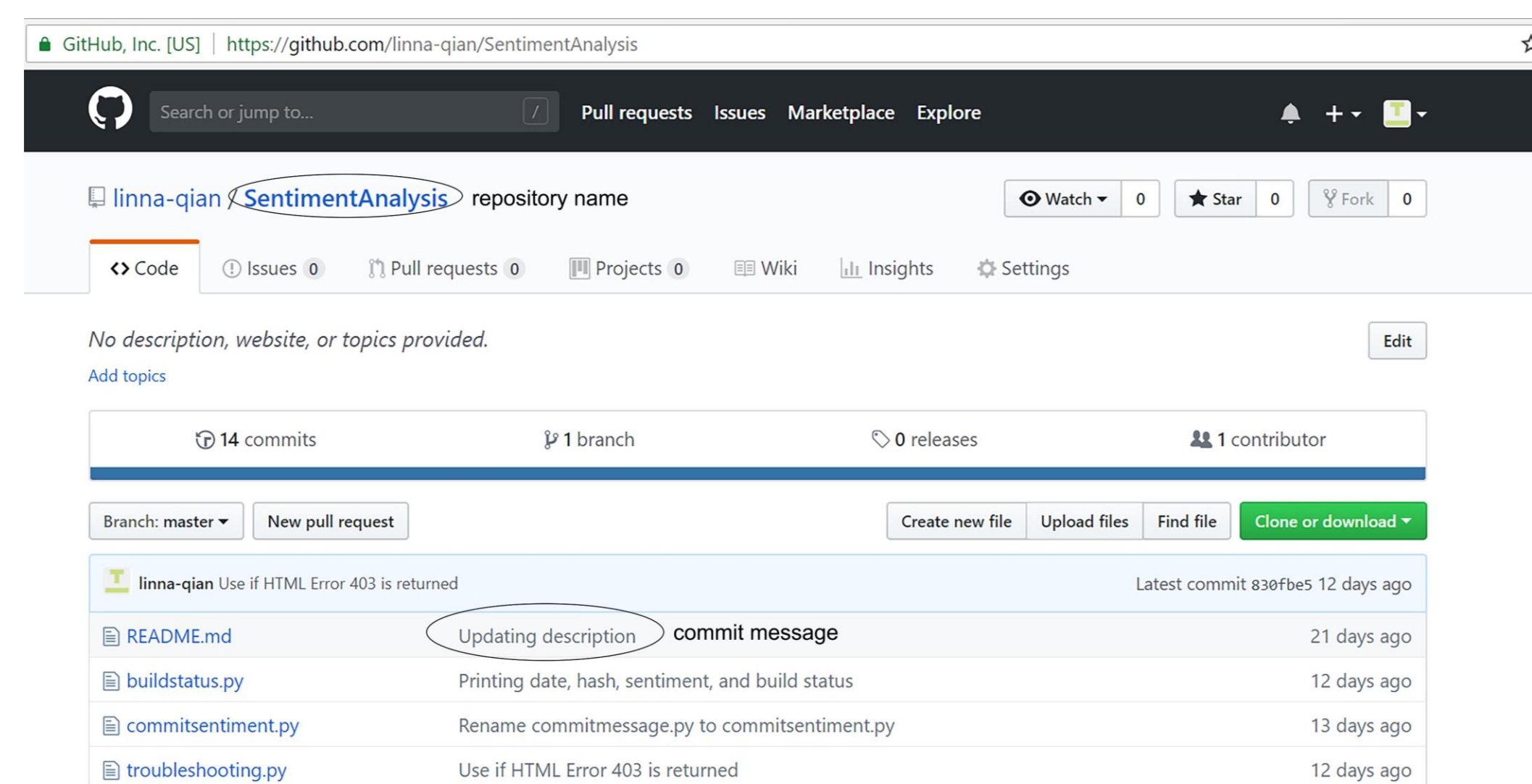


Figure 1. Example of a GitHub repository.

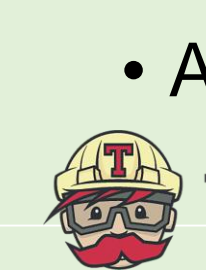
## Definitions and Examples

### GitHub

- A website where programmers can store their projects for others to see, use, and develop



### Travis CI



- A continuous integration platform that automatically builds and tests code changes

- A **GitHub repository** is where the contents of a project are stored (project file)

- Commit messages** are messages that programmers attach to the work they upload to their repository

### Continuous Integration

- Merging smaller code changes more frequently rather than one big merge at the end

### Travis CI Builds

- The GitHub repository will be copied and tested in a virtual environment. If any part of the build fails, the build status returned will be either errored or failed<sup>1</sup>

### JSON (JavaScript Object Notation)

- JSON is a formatting used in data transfer between remote servers and web applications

- Extract the sentiment of the commit messages with a Python library called **TextBlob**. The sentiment is returned as a value between -1 and +1

<sup>1</sup>Core Concepts for Beginners." *Travis CI Docs*, Travis CI, docs.travis-ci.com/user/for-beginners.

```
"state": "success",
"description": "The Travis CI build passed",
```

Figure 2. Example of the output from a Travis CI build.

```
"committer": {
  "name": "linna-qian",
  "email": "linnaqian776@gmail.com",
  "date": "2018-07-26T17:30:11Z"
},
```

Figure 3. Example of a JSON file.

```
"message": "Changing the colour of calendar cells by date",
```

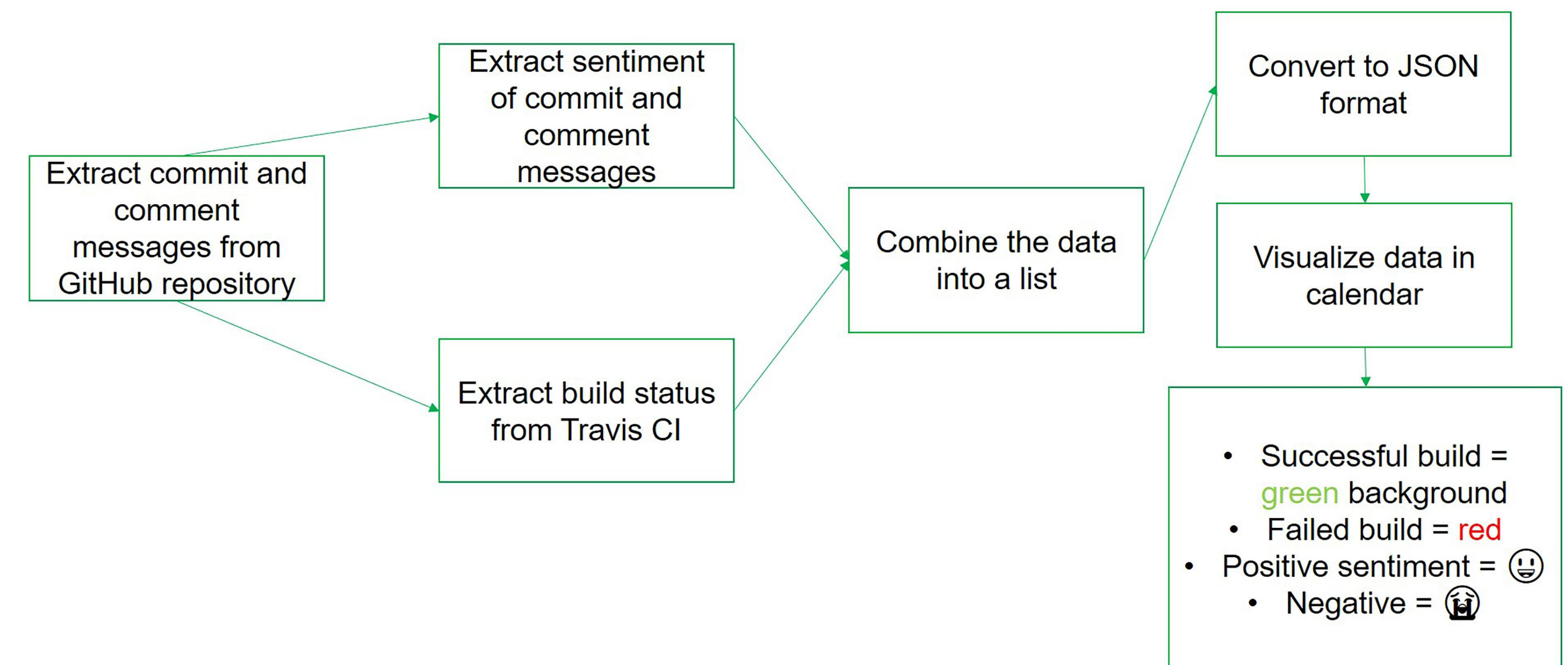
```
from textblob import TextBlob
example = TextBlob("Nice to meet you all!")
print(example.sentiment)
```

Figure 4. TextBlob input

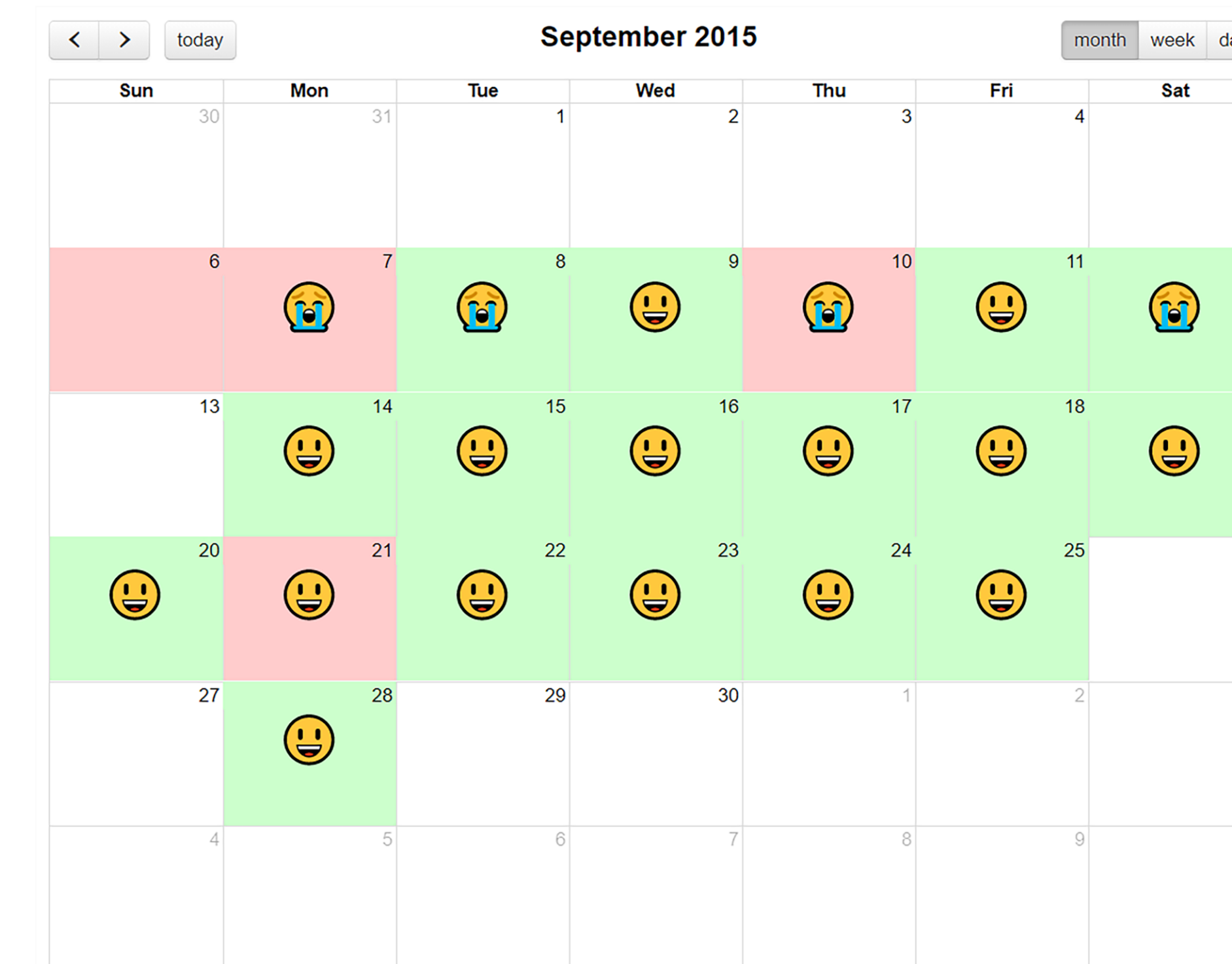
Figure 5. TextBlob output

Sentiment (polarity=0.75, subjectivity=1.0)

## Methodology Overview



## Result



This is the model calendar for the GitHub repository pingcap/tidb during September 2015. This was one of 15 repositories analyzed.

## Future Studies For Sentiment Analysis

- Improving the accuracy of sentiment analysis software (like TextBlob)
- Analyzing, not only text, but pictures and videos
- Increasing the complexity of text able to be analyzed (eg. correctly interpreting sarcasm and hyperboles)<sup>1</sup>

<sup>1</sup>Future of Sentiment Analysis and Problems Faced." *DatascienceCMU*, 19 Apr. 2014, datasciencecmu.wordpress.com/2014/04/18/future-of-sentiment-analysis-and-problems-faced/.

## Future Work

- Preprocess the commit messages to remove text like URLs and code snippets
- Adjust sentiment values based on the type of text being analyzed (commit and comment messages for this project)

## Acknowledgements

I would like to thank Dr. Sarah Nadi and Moein Owhadi-Kareshk for welcoming me into the SMR lab and giving me the opportunity to learn about sentiment analysis and program development.

I would also like to thank the Ross and Verna Tate foundation for providing me this opportunity through the High School Internship Program (HIP).