End of Semester project 02

This project is builds upon the Mid Semester Project

1. Course details:

Course Code: MCN 7105

Course Name: Structure & Interpretation of Computer Programs

2. Project aims

The goal of this project is to enable you study an existing code base, identify abstractions (data, & procedure abstractions) and **build new abstractions for a given problem domain.** For this project, you will explore Data Science abstractions for Scheme - https://github.com/n3mo/data-science. The purpose of the abstractions is to allow developers to write data analytics and manipulation software in a language that does not natively provide the support. In this case the language is Scheme.

1.This project is mandatory and is to be done individually
2.The project will contribute 30% towards the end of Semester Examination.
3. **Deadline**: 9th November 2018 at 5:00PM (Fixed and cannot be extended).

3. Deliverables

- **Project report:** The report must not be longer than 5 pages and in PDF format. The report should describe/motivate any design decisions taken and include code snippets where applicable and a github link to your code base.
- **Submission:** The report should be submitted via Muele as a single .pdf file named as *Firstname-Lastname+-2.pdf* You are encouraged to try and submit before the deadline. **Email and late submissions will not be accepted!**

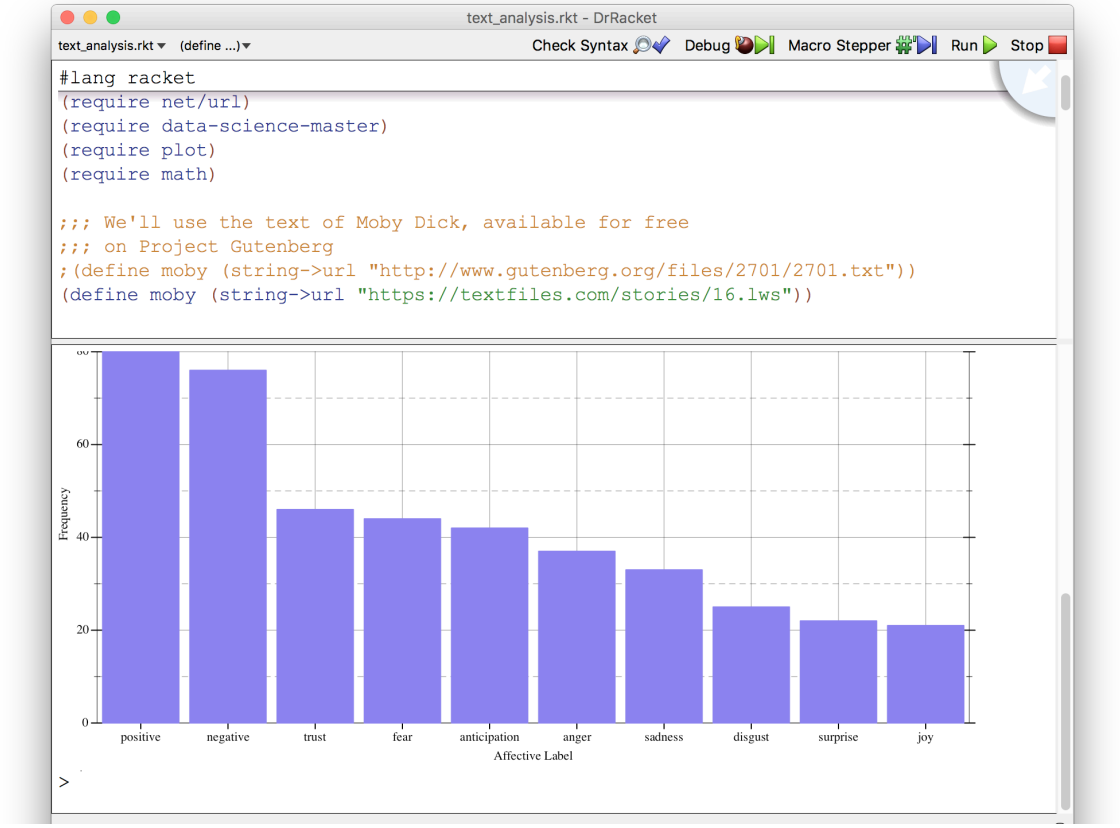
4. Project/Task

I. In the previous project, you studied the implementation of the following data science abstractions in the code base of https://github.com/n3mo/data-science . The data science abstractions offer support for sentiment analysis, a technique that is commonly used to quickly determine the mood, or emotional valence of a body of text. For example, using the system to analyze the mood of words for textual novels provided by *Stories and Fiction* http://textfiles.com/stories/ yields the following results in Examples I and II.

The abstractions have also been used for analysis Twitter data see http://www.nicholasvanhorn.com/posts/trump-tweets.html *For this project you are required to build a set of abstractions that can be used by developers to build software systems that analyze the moods of Tweets for a given country (e.g., Uganda) over a period of 12 months.* You can assume that locality information provided by Twitter API is accurate. You may draw inspirations or build upon some of the abstractions already provided in http://www.nicholasvanhorn.com/posts/trump-tweets.html and the data science

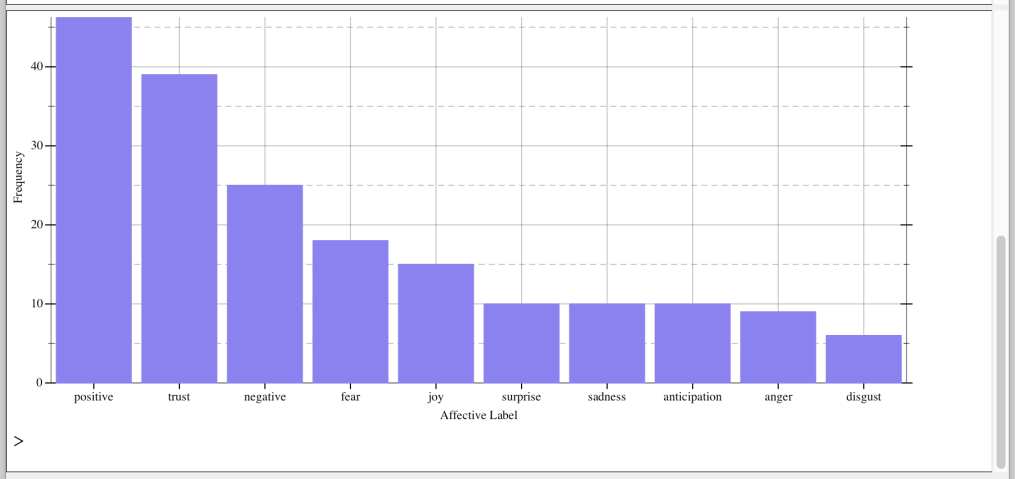
layer. Provide description of the new abstractions developed (including motivation and any design decisions) and drawings showing the resulting levels of the new abstractions introduced. The report should also include sample outputs of a sample system built using your abstractions. Your codebase should be posted on github and a link provided in the report.

Example I: The Reality of Our Situation by Kortron
https://textfiles.com/stories/reality.txt



Example II: Two Guys in a Garage, by M. Pshota
https://textfiles.com/stories/16.lws

```
;;; We'll use the text of Moby Dick, available for free
;;; on Project Gutenberg
;(define moby (string->url "http://www.gutenberg.org/files/2701/2701.txt"))
(define moby (string->url "https://textfiles.com/stories/reality.txt"))
```



Preparing the Setup. The project was designed using Racket V6.11. What follows is a short step-by-step description on how to setup Racket and the data science abstractions for your platform and to run the examples above.

1. If you don't have Racket installed, Download and install Racket from https://racket-lang.org/download/
2. Download the data science abstractions from https://github.com/n3mo/data-science
3. Extract the data science abstractions zip folder into the Racket v6.11/collects folder. This will add a new folder named "data-science-master"
4. The data science abstractions may require other packages. To install any additional package download and unzip the file into /collects/ folder
    a. csv-reading https://pkgs.racket-lang.org/package/csv-reading
    b. mcfly https://pkgs.racket-lang.org/package/mcfly
    c. overeasy https://pkgs.racket-lang.org/package/overeasy
5. Any other Racket packages that may be missing can be installed from https://pkgs.racket-lang.org manually as above or using the *raco* automated tool
6. Download the example from the Muele course page text_analysis.rkt to test the installation and the data science abstractions. More examples available at https://github.com/n3mo/data-science