MAKERERE UNIVERSITY
SCHOOL OF COMPUTING & IT
MCN 7105: STRUCTURE AND INTERPRETATION OF COMPUTER PROGRAMS

END OF SEMESTER PROJECT
ACADEMIC YEAR 2021/2022

Instructions

Deadline: 15th May 2022, 11:00PM.
Deliverables: Source code file and a report (max. 5 pages, PDF) as a single zip file.
Submission: This is an individual assignment to be submitted via Muele. No email submissions please!
File Naming: Firstname-Lastname-project2.zip

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. 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.

Assignment

1. Study the implementation of the following data science abstractions in the code base of `https://github.com/n3mo/data-science` and describe their different levels of data and/or procedure abstraction layers. Provide a diagram showing the different abstractions levels per abstraction. A step-by-step guide is provided at the end of this document to setup and run the data science abstractions.

   (a) linear-model
   (b) list->sentiment
   (c) read-csv
   (d) qq-plot*
   (e) hist

   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.

Example I: The Reality of Our Situation by Kortron `http://textfiles.com/stories/reality.txt`
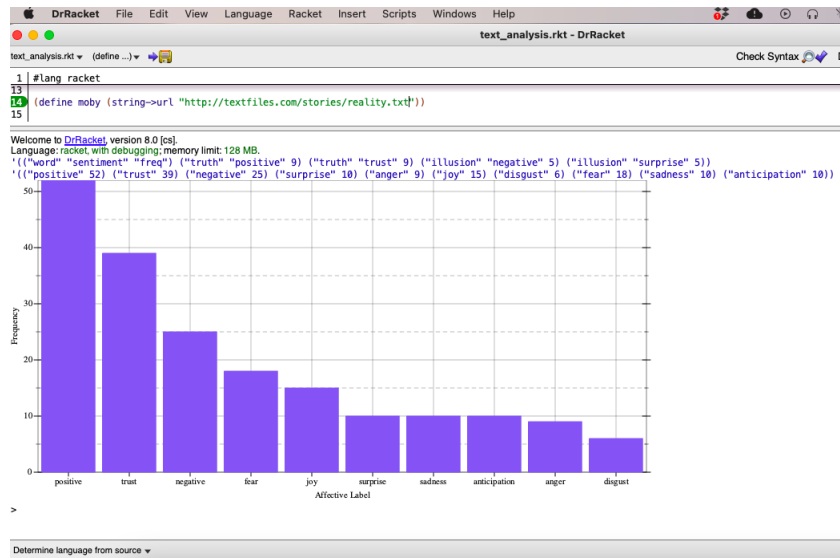
Example II: Two Guys in a Garage, by M. Pshota `http://textfiles.com/stories/16.lws`
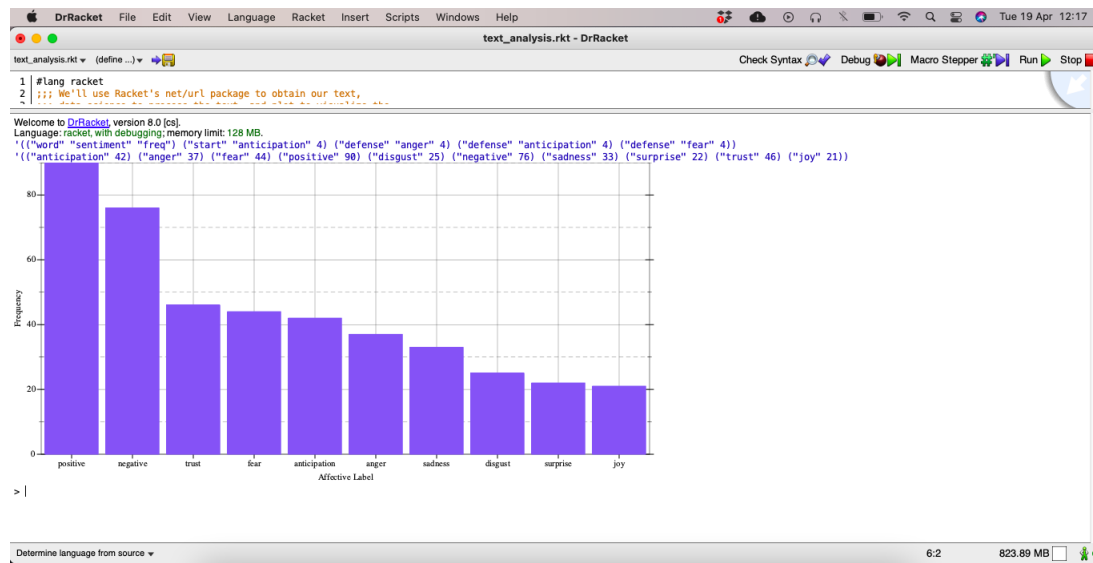
2. The abstractions have also been used for analysis Twitter data see `http://www.nicholasvanhorn.com/posts/trump-tweets.html`. For this project you are required to build a set of abstractions that can be used by developers to build software systems that analyze the moods of tweets for a given country (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 code base should be posted on git hub and a link provided in the report.

Preparing the Setup

The project was designed using Racket V8.0. 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. Download the data science abstractions from `https://github.com/n3mo/data-science`

2. Extract the data science abstractions zip folder into the Racket v8.0/collects folder. This will add a new folder named "data-science-master"

3. 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`

4. 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

5. Download the example from MUELE **text-analysis.rkt** to test the installation and the data science abstractions. More examples available at `https://github.com/n3mo/data-science`

THE END