Instituto Tecnológico Autónomo de México

Engineering Department

A syntactic approach to text analysis

Fuentes de datos

January 25, 2022

Abstract

In this project, you will implement a fully integrated solution for performing comparisons between texts based on lexical and syntactical analysis of their constituents.

Contents

1 Implementation

1.1 Setting environment

Please implement the following functionality.

1. [**10 points coding**]

    a) Ask the user to provide a valid path for the environment. `set_env.sh`

    b) Check if the entered path is empty. `set_env.sh`

2. [**15 points coding**]

    a) Complete and validate implementation of `get_bookshelves_list.sh`

    b) Complete implementation of `get_bookshelves_list_sed`

This will execute all the setup steps for generating a working environment for the project. In addition, it will also get a list containing all the bookshelves within the Project Guttenberg website.

1.2 Extracting booklists

Please implement the following functionality

1. [**10 points coding**]

    a) Get bookshelf_url on `get_bookshelves.sh`

2. [**10 points coding**]

    a) Desribe with words what each sed script is doing at `get_bookshelves_titles.sh`

These functions enable the script to download the book collections on each bookshelf.

1.3 Extracting book main text

1. [**25 points coding**]

    a) Please complete the implementation of `get_books.sh`

2 Presentation

2.1 Peer review session

1. [**30 points**]

    a) Please prepare a brief report to be presented in class; detailing your main findings, debugging process, test cases and conclusions.