Problem Solving Hackathon: Automatic Extraction of Acronyms : Rev 1.0

C.P. Ravikumar

December 8, 2021

The goal of the hackathon is to give NCG an opportunity to carry out a mini project and experience all the phases of a project, such as Design Documentation, Design Review, Project Development, Testing, User Documentation and Project Review.

1 Problem Description

We use a lot of TLA (Three Letter Acronyms) in TI. The intention of the project is to identify TLA from a given document. The input document may be in PDF or an HTML format and may be specified through a URL. Your program should maintain an acronymn database where acronyms of 2 or more characters are stored. You may assume that all the characters of an acronym are in capital English alphabets. The program must maintain an acronym database. New acronyms located in the given document must be added to the database. If you find an acronym being used without a definition, your program should print an error. Any new acronyms found in the document must be incrementally added to the acryonym database.

You may develop your program in Unix or in Windows. There is no restriction on the language of implementation.

2 Rules for the Hackathon

1. You must work in groups of 3. We will permit a group of 2 under exceptional circumstances.
2. You may reuse existing code, but give due acknowledgement in the "References" section of your User Document.
3. Your system should be demonstrated and should be usable by other TI-ers
4. Submit a design document by Friday, 5.00pm

5. Submit your code and a user document in PDF format by 14th Dec, 5.00pm. Demonstrate your software on 14th Dec. Each team will get around 10 to 15 minutes to showcase their demo.

6. The following points will be considered in evaluating the project submissions.

    - Timely submissions
    - Quality of code and documentation
    - How well were the specifications met
    - Demo

3 Groups

- Sundar, Krish, Goury
- Krishna Teja, Likhit, Shravan
- Aaron, Shilpa, Meenakshi
- Guru, Pavan, Krishnakant
- Mohit, Pallav, Sudhanshu
- Ram Prakash, Ian Sharma
- Maruthi, Aravindakshan, Ajinkya

4 Deliverables

- Design Document (this document will be briefly reviewed and comments will be sent back to you, which you may include in your revised version)
- List of test cases
- Final Report, User Documentation and Demo.

5 Format for Design Document

Your design document must have a suitable title, list of authors, date, revision number, and document classification information. You must bring out the problem you are solving with any assumptions you have made. You may want to create one or more test documents that test for various test cases. Here are some examples. Please look at several example documents before you finalize the test cases.

- The document may have TLA (expansion) or expansion (TLA)

- There may or may not be white spaces before and after the parentheses
- The expansion may be in title case or may be in lower or upper case

Please feel free to reuse code available from public sources. Provide acknowledgement in the "References" section.

6 Format for Report

Your final report must have a title, list of authors, date, revision number, and Document classification information. The report may be organized into sections such as "Problem Description," "Implementation Details," "Testing,", "Known Bugs," "Results and Conclusions" and "References."

In the "Results and Conclusions," *please write down one or two things that you learnt in the process of this project.*

Also develop a brief user document which can be a manual page or a help document which explains how the system can be installed and used.

References

[1] TI's TLA Database. URL: https://infolink.sc.ti.com/news_rooms/p/acronyms

[2] A. Khemeri. PDF Processing with Python. https://towardsdatascience.com/pdf-preprocessing-with-python-19829752af9f. 2019.