

CPC353/CPT346 Assignment 1:
Special Topic Study on Natural Language Processing
2021/2022

Released Date: 7 November 2021 (Sun)
Last Submission Date: 5 December 2021 (Sun)
Submission method: Softcopy in Word and Pdf (eLearn)

This project is a 2-in-a group project.

1. TOPIC SELECTION

Step 1: Choose the domain/genre

- Health Domain (e.g., Covid/Medical Reports/Health website etc.)
- Financial Domain (e.g., Financial News/Property Forum/Financial Report etc.)
- Tourism Domain (e.g., Tourism website/Tourism spot review/Tourism blog etc.)
- Scientific Literatures Domain
- Literature (Fiction/non-Fiction) Domain
- Etc.

Step 2: Perform the following discussion/analysis

- Domain Motivation and Sample data (Discussions)
- Text Exploration (Analysis – choose 2)
 - Concordances
 - Stop Words
 - Part of Speech
 - N-gram
 - Grammar (Phrase/Structure)
 - Semantics
- Applications, Visualization and Communication (Discussions)

Step 3: Form the topic

Exploring [domain] using [analysis technique]

Example topic:

Exploring Literatures for Children using Concordances and Part of Speech Methods

Step 4: Submit your assignment title at the Excel Sheet

<https://docs.google.com/spreadsheets/d/1b4bHlqHucGBVnOx4AOSD16IPgQgyJ-c2WspFw2RZsoQ/edit?usp=sharing>

2. PERFORM LITERATURE REVIEW & TEXT ANALYSIS

You are required to conduct literatures studies & analysis of your selected topic. The scope of the literatures/dataset/analysis should be relevant to topic selected.

Requirements:

In this assignment, you will need to

- Collect small sample datasets related to the selected domain. Minimum 20 texts.
- Identify reference materials related to the selected methods. References materials include journal articles, conference articles, books, URL. Main references should come from citable references, e.g. from Scopus, Web of Science. Additional references can come from website or other type of documents. Estimated 5-10 references are needed.
- Conduct review on selected references.
- Perform text exploration on the sample data. Exploration can be done using small sample investigation or concept level discussion.
- Report the outcome of exploration. The outcomes cover overview of the topic, discussion about the sample data, summary of your findings on related methods, and results of text exploration. You may support your outcome using diagrams, tables etc.

3. SUBMISSION

You will need to submit the report in softcopy format. Minimum 2 pages (max 4 pages) (excluding cover). The report needs to follow manuscript (IEEE format)

Template: http://www.ieee.org/conferences_events/conferences/publishing/templates.html

Guideline:

The report needs to include minimally these sections. (These are guidelines only; you may modify the outlines accordingly to improve your report)

- Title and Abstract
- Overview (state the NLP problem in this domain)
- Scenario (this part contains the dataset description)
- Related Methods (this part contains the related works on the selected methods)
- Text Exploration (this part contains your own analysis using the selected methods on your sample data)
- Conclusion (state how the findings can be used)
- References

4. EVALUATION CRITERIA

1. Understanding of selected topic & scenario.
 2. Selection of the literatures (timeliness, quality)
 3. Conciseness of Summary for related works.
 4. Breadth and accuracy of discussion for text exploration.
 5. References.
 6. Overall writing, structure, flow, and language.
- (Refer to rubric document for the criteria)