Project Proposal

Wordle Game Strategy

CS 410, Fall 2022

Question 1

- What are the names and NetIDs of all your team members?
  - Adam Florzak (aflor3)
- Who is the captain?
  - Adam Florzak

Question 2

- What is your free topic?
  - Wordle game strategy
- What is the task?
  - Develop an analysis tool for optimal Wordle game strategy using techniques learned in this class.
- Why is it important or interesting?
  - Wordle is an incredibly popular online word game that saw explosive user growth when The New York Times purchased the rights to it in January of 2022. Link (https://www.nytimes.com/2022/01/31/business/media/new-york-times-wordle.html)
  - The Times CEO even mentioned Wordle when presenting the company's quarterly results, stating, "Wordle brought an unprecedented tens of millions of new users to The Times, many of whom stayed to play other games which drove our best quarter ever for net subscriber additions to Games." Link (https://s23.q4cdn.com/152113917/files/doc_news/2022/05/NYT-Press-Release-3.27.2022-Final-O2ACvs2.pdf)
  - The popular animated math YouTube channel 3Blue1Brown posted a video entitled "Solving Wordle using information theory" and it quickly became one of the channel's most popular videos with over 9.2 million views to date. Link (https://www.youtube.com/watch?v=v68zYyaEmEA)
- What is your planned approach?
  - Analyze the `past_words.txt` dataset to determine if there are any significant patterns in terms of POS tags and/or TF-IDF scores among solutions as compared to the `allowed_words.txt` dataset.
  - Create a method to encode a 5-letter word as a 5×26 matrix and use matrix operations to efficiently validate guesses and conduct pattern matching with the vocabulary.
  - Given a current game state, return a ranking of the most likely possible solutions.
- What tools, systems or datasets are involved?
  - `allowed_words.txt` contains 12,972 5-letter words that are legal guesses
  - `possible_words.txt` contains 2,315 5-letter words that are possible solutions
  - `past_words.txt` contains 509 5-letter words that are past solutions
- What is the expected outcome?
  - In January 2022, the website WordTips used the Twitter API to extract 142,669 tweets containing Wordle scores, and found that the worldwide average number of guesses is 3.919. Link (https://word.tips/wordle-wizards/)
  - The expected outcome for this project is an analysis tool for optimal Wordle game strategy that performs significantly better than the average user score.

- How are you going to evaluate your work?
    - Test the tool against the `past_words.txt` dataset and compare the average score using the tool to the average user score of 3.919.

Question 3

- Which programming language do you plan to use?
    - Python

Question 4

Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

Task	**Estimated Time Cost**
POS Tagging and Analysis	2-4 hours
TF-IDF Scoring and Analysis	2-4 hours
Replicate Core Game Functionality	4-6 hours
Design the Matrix Operations based Game Strategy Tool	3-4 hours
Implement the Game Strategy Tool	9-12 hours
Testing	2-3 hours
Evaluation	2-3 hours
Total	**24-36 hours**