

Web Scraping and Text Analysis of Game Reviews on Rock Paper Shotgun

Overview

This repository contains the code and resources for a school NLP project. The project focuses on leveraging web scraping & topic modelling techniques to extract and analyse game reviews from the chosen website - Rock Paper Shotgun

(<https://www.rockpapershotgun.com/reviews>).

The code and comments for building the dataset & conducting text analysis are available in the submitted zip file and on GitHub

(<https://git.arts.ac.uk/23001934/rps-game-reviews-web-scrape>).

Dataset

rps_links_output_file.json - JSON file contains all the titles and links on

RPS game_reviews_rps_in_recent_years.json

game_reviews_rps_in_recent_years.csv

game_reivews_rps_in_recent_years (folder)

The dataset can be reviewed in JSON, CSV, and TXT formats, all available in the submitted zip file and on GitHub

(<https://git.arts.ac.uk/23001934/rps-game-reviews-web-scrape>). Please note that the dataset only covers the reviews in the recent three and half years (published from 01/06/2020 to 01/12/2023 on RPS), with the raw dataset (contains missing data and incorrect data) available in the 'test_output' file for comparison purposes.

For more information on the dataset, please refer to the datasheet

(`rps_datasheet.md`)

Web Scraping & Dataset Analysis

1a_web_scrape_test_code.ipynb 1b_web_scrape_rps_links.ipynb

1c_web_scrape_rps_contents.ipynb

1d_reviews_dataset_formatting.ipynb 2a_reviews_data_analysis.ipynb

2b_reviews_topic_modelling_LSA.ipynb

2c_reviews_topic_modelling_LDA.ipynb

The above files which can be found in the submitted zip file and on GitHub (<https://git.arts.ac.uk/23001934/rps-game-reviews-web-scrape>) include all the code for building the dataset and conducting text analysis. The notebooks are named sequentially, with comments and references available in each file.

Project Report

nlp_project_final_report.pdf can be found in the submitted zip file.

Disclaimer

This project is created purely for learning and educational purposes. The content, information, and materials presented herein are intended for academic use only.

LLM Disclaimer & References

ChatGPT 3.5 has been used in this project for grammar checks, translation, helping in code debugging, as well as for searching certain Python codes which have not been taught in class. Comments were added to each notebook where ChatGPT was used, including links to the reference code obtained online.