

CS510 Project Proposal - Development Track

Question-answering system for game reviews

Team 24 - Prathik Srinivasan (ps94@illinois.edu)

[Functions and Users]

I am developing an interactive web-based intelligent agent that will allow users to view generated collective summaries of user reviews about PC games as well as use natural language to explore specific patterns within the reviews. The system will scrape review data from Steam, the most popular PC game distribution platform, and utilize this data to generate responses to user queries as well as display a general summary.

Features:

- Users will be able to search for a specific game distributed on steam by name or url
- For any game that is found, a generated summary will be provided based on user-generated reviews (for games with an extreme amount of user reviews, a subset will be used for generation)
- Users will be able to use natural language to query reviews with information about specific topics, which will then be retrieved and summarized

Users:

- Game developers seeking to understand general sentiments about their user base without needing to explore every single review
- Game developers seeking to do market research about common issues or successful implementations in other similar games
- Consumers seeking a generalized summary of common sentiment about specific games

[Significance]

There are some other existing implementations of a similar idea, especially with regards to summarization. However, the pain point that I specifically want to look to address is the ease of **finding specific information within Steam user reviews** rather than only generalized information.

I believe this is an aspect that is relevant for both users and developers who seek information from reviews. Users will want to know of specific issues that may have been faced by others with similar hardware, or want to seek out information from user reviews that appear to have similar tastes to themselves. Developers on the other hand, will want to understand the target market and explore if common pitfalls are present across a genre as well as exploring patterns within the feedback of various users.

Currently, Steam has some amount of filtering functionality present for reviews, but none that pertain to review content beyond language. Other existing tools also do not

offer this sort of question answering system and simply focus on summarization of general sentiment, which may include some common issues but is not particularly specific.

I aim to overcome this pain point by creating a system that will maintain the ability of using natural language search capabilities, but be able to both summarize the result of such queries while also retrieving specific relevant reviews to the query.

[Approach]

For the general flow of the application, the user will search for a specific Steam game by its URL. The app will display general info about this game such as overall score, number of reviews, estimated revenue, etc. The app will also scrape the steam API for a subset of reviews representative of the overall score, filtering to include generally 'helpful' reviews, and generate an overall summary.

From there, the user will be able to create queries about specific information using natural language. If no information about the query is found within the current subset of reviews, more will be requested by the API (until some threshold). The relevant reviews will then be summarized by the LLM, as well as returned in their original form

- The steam API will allow me to scrape reviews for any given game available on the platform
- Interactive web front-end will be built with React
- For generating summaries, processing natural language queries, and question answering, I plan to use a relatively smaller LLM like Llama 3 or something similar and LangChain

[Evaluation]

Usefulness:

- User Surveys: Feedback will be collected from both game consumers and game developers about tool's effectiveness
- Accuracy: Summaries will be validated by manual review of some example games, and the original text used will be evaluated to ensure hallucination is avoided

Technical Validation:

- Ensure robustness against malformed queries
- Ensure functionality on games with both an extremely large and extremely small amount of reviews

[Timeline]

Week 1: Setup of basic frontend, develop system of collecting info via steam API

Week 2: Setup LLM and query processing, develop summarization system

Week 3: Create query processing and review retrieval/summarization system

Week 4: Testing, evaluation, user feedback collection

[Task division]

Prathik Srinivasan - everything