CSCE 470 Task 1 Report

Gamer's Master Management Application

Background

This project aims to improve the current Steam's game indexing that will shine a fair light to all game titles. By utilizing data crawling technique, search engine and trend knowledges, we hope to bring about the changes to the current indexing system and help smaller game companies reaching their consuming mass.

Motivation

It goes without saying that the current gaming industry is massive and on the rise. Totaling over $120 billion globally, the industry is dominated by a handful of huge gaming companies. This brings about an extremely competitive game development environment, and cutthroat marketing. Industry giants like EA, Ubisoft, and Activision are notorious for high budget marketing campaigns with no consideration of releasing a complete or bugtested game. These games are high in price and made to look more polished visually yet offer little to no actual gameplay content. However, due to the extensive marketing campaigns, these incomplete products still sell in high enough volume to corner the market and limit the success of smaller, lesser known game companies.

Previous Work

In our previous class, CSCE 315 Programming Studio, the first project was to pull data from Yelp and then create an application that allows the user to view the database organized by the team from the Yelp data. Task 1 is reminiscent of the part of the project where we were required get the Yelp data and organize it into the database.

Objective

Our objective is the development of a service that allows the user to search games based on their personal preferences (genre, developer, tags, reviews, Metacritic score, etc) without the influence of advertisements preference pushing products at the top of the list. More options within refining the search include the toggling of repeated products (Game of the Year Edition, Bundles), DLC content from common searches, recommendations off of similar preferences, filtering developers, and other features that the team may include if deemed feasible.

Data Crawling Methodology

Task 1: Data Crawling

We are using Scrapy, a Python framework for scraping and web crawling, to extract relevant data from Steam. This traverses Steam and pulls from each game all relevant information we may need in regards to consumer's preference. This includes the game's title, description, review aggregation, Metacritic score, genre, tags, system requirements, developer, publisher, early access status, price, dlc list, and rating.

Results

Following twenty-six hours of crawling Steam, we were able to acquire more than the requested 100,000 documents required for this task; this totalled to over 50 MB of data consisting of game description, tags, information on the product, and related statistics.

Point of Contacts

- Han Hong: hongichhan@tamu.edu
- Juan Duran: juan_duran26@tamu.edu
- Anh Nguyen: nducanh2311@tamu.edu
- Github Repository: https://github.com/harryluffy/Data-Crawler