Tyler Crabtree

Web Scraper Paper

CS483

Berserk Scraper

The Web Scraper scrape.py covers and collects data about a Japanese story named *Berserk*, originally written by Kentaro Miura. *Berserk* is comprised up different mediums with varying critical reception. Using Wikipedia's API, the goal of this scraper is to gather information on *Berserk* while separating the different mediums into their own categories: manga, original series, video game, modern series, and films. From there, the scraper will parse the critical reception and pair it with the respective medium.

The program forms two tables within a database called *Berserk*. These two tables are named, *BerserkTable* and *Reception. BerserkTable* has the tuple (Title, Summary, Image, Categories) while *Reception* only has the tuple (Title, Reception). These tables were broken up, so the information could be clearly depicted in the terminal. Here is a video depicting the functionality of the program (both of these tables are presented): https://www.youtube.com/watch?v=XznrqAo-73k&feature=youtu.be. Example tables are also attached within this project (within the .tar).

The program will also print the information that it is scraping from Wikipedia straight to the terminal (in a more verbose format). The terminal has a total of 10 commands. These commands consist of having the user either scrape individual mediums (for example, only information pertaining to the manga), or a user can choose to grab all information pertaining to berserk (all mediums). A user also has the option to delete either *BerserkTable* or *Reception*. Also, the user can enter their own type of

media to look for (for example a user could search “Pokemon”). The database eventually be expanded and elaborated to contain a more general genre of anime. All commands will be initially prompted to the user when the program is started. After the program is ran (without the delete table commands) the *Berserk* database should be full of useful information.

Sincerely,

Tyler Crabtree