

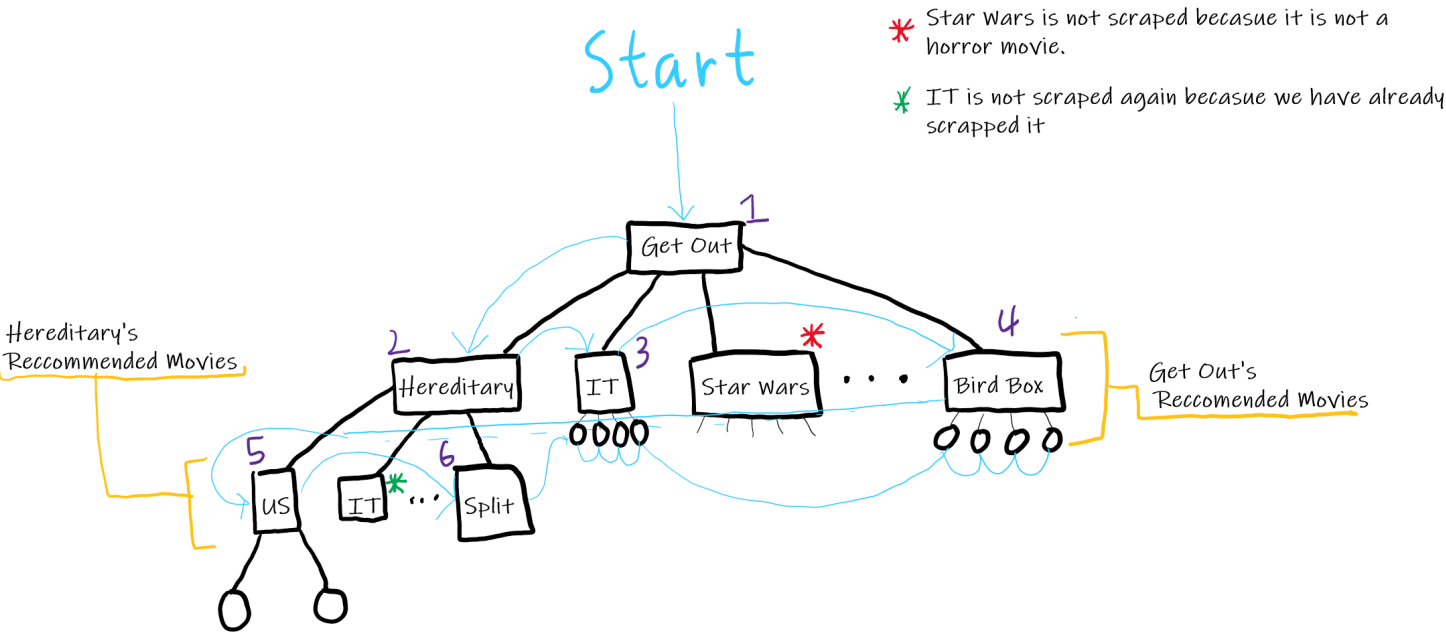
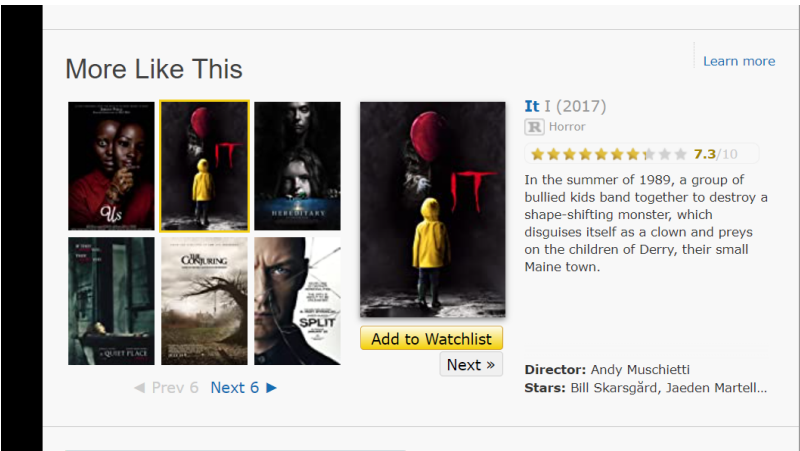
# IMBD Web Scrape Description



Each horror movie has a corresponding webpage on IMDB.com, with information like the runtime, genres, and reviews. The python webscraping library selenium was used to gather this information from each movie's webpage.

Each page also had a list of recommended movies. This was used to determine the next movies to collect information on.

This is done using a breadth-first graph traversal, treating each movie as a node in a graph, and the recommended movies as edges of those nodes.



The above diagram demonstrates this process.

First, "Get Out" is scraped, meaning the information about it is gathered.

Then, all of the movies in "Get Out"'s recommended list are scraped, skipping non-horror movies, and movies that have already been scraped.

Then, this process is repeated starting with the first of "Get Out"'s recommended movies. This process continues recursively until it is manually terminated.

Note, to ensure that this process would capture a variety of movies from across time periods, it was initialized with three movies: Get Out (2017), Alien (1979), & Psycho (1960).