CS242 Online Group 3 Project Proposal

Hovanes Keseyan

Janakiram Kuppa

Jorge Mercado

Group 3 has decided to create a database index of data from a fictional world. We will attempt to do this for the world created in legendary author J.R.R. Tolkien’s works: The Lord of the Rings, The Hobbit, and The Silmarillion. If for some reason (size, accessibility, etc), we are not able to use this world, we will adapt our process to collect data from another world/worlds.

We will design a web crawler to crawl websites that contain data about characters, their details, locations in the world, and other information gathered from plot points and stories told in these works. We will begin by crawling a large and popular online Wiki located at <http://lotr.wikia.com>, known as “The One Wiki to Rule Them All”. This wiki is populated by fans with data from all of the literature behind this world, including not only the original books but other canonical media such as the popular and successful movie series from the 2000s. Each character has a wiki page which describes any known information about their timeline, city/town of origin, actors that have portrayed them in media, relationships to other characters, and important plot points among the immense library of information that is available.

We will index this data using Lucene first, and later using Hadoop. We will compared the performance of these alternatives.

We have ideas for interesting and creative ways to visualize the data queried from this database. One of which is to create a map of the fictional world and plot locations where the characters can be found at different points of the story, including the paths they traveled based on what information we can gather. We can also visualize character relationships by linking them in a graphical interface.