

Search for the Exoplanets

Team ExoTerra

Thomas Dill, Thomas.dill@wsu.edu

Brian Haugen, brian.haugen@wsu.edu

Daniil Kernazhytski, daniil.kernazhytski@wsu.edu

1 Introduction

The project will scrape a Wikipedia page on exoplanets- those planets outside of our solar system. The user will be presented with a web interface with some information on how to search for exoplanets, and they will be allowed to enter in a search query. Relevant results will be displayed. A link to the relevant Wikipedia page for the exoplanet will be presented, and information regarding the parent star will be included. If there is an actual picture or artist's rendering of the exoplanet, that will be shown.

2 Database Summary

An SQL database will be constructed from web scrapping Wikipedia. The tuples included will be: <name, mass, period, Discovery Method, Year Discovered, distance, Host star mass, Host Star temp. (K) > - Example of a tuple from our database.

A secondary SQL will be created from Wikipedia for the parent stars associated with the exoplanet.

Whoosh will be used to search through the NASA Exoplanet web page for additional information (images, parent stars, additional information to be determined later).

The size of the database is over 30,000 tuples.

3 Indexing

The names of the planets will be searchable. The search results can be filtered by the mass, period, discovery year, distance away from earth as well as the information about the host star. There will be no indexing or stemming, since the information is mostly scientific and numerical in nature. Therefore, the results can be restricted exactly as specified. For example, all planets under 100 years light years away, discovered in the past 20 years.

Search for the Exoplanets

4 Features

4.1 Feature 1

An online web entry form, where the user can specify the details of his search, such as the mass of the planet, distance from earth, or distance from host star. This will be guided by information on the splash page.

4.2 Feature 2

If there is an image or an artist's rendering of the planet available, it will be displayed along with the results.

4.3 Feature 3

Information about the parent star, and Wikipedia links (if available) for the exoplanet will be displayed.

5 Member Responsibilities

Web Page

- HTML/CSS

Web Scraper

- Beautiful Soup for Wikipedia

- Whoosh NASA Exoplanet page

- Populates a database with planet and star info

Search Engine

- Search the database

- Returning the results

 - Results organized by search

 - Returned in a particular format

Daniil - setting up the SQL database from .CSV file (Wikipedia Exoplanets and possibly NASA Exoplanet data from Whoosh), integrating the database with web page

Brian - building the webpage search form and results page using HTML/CSS.

Thomas- Program Manager; developing web scraper; Presenting the project, documentation. Scrape Wikipedia. Develop a Whoosh search function through NASA Exoplanet website.