ASTRUN 3646 Rhessa Weber Langstaff, Ngan Bao, Alexandra Savino

Project Phase II

Main Project Topic: The Robot Astronomer

For our project, we have decided to build our own AI astronomer which determines the most searched objects in a given day and gets the coordinates of said objects through a DSS (Digitized Sky Survey) query. The project itself can be navigated through a simple and easily understandable website.

Background:

The application of Artificial Intelligence (AI) and automation in astronomy has brought about revolutionary changes. The advancements in these technologies have led to significant improvements in our ability to collect and analyze vast amounts of astronomical data, resulting in significant discoveries about the universe.

AI and automation have been primarily used in the collection of data in astronomy. With the development of advanced telescopes, vast amounts of data can now be captured about the universe. However, manual processing and analysis of this data can be time-consuming and labor-intensive. AI and automation have made the data collection process more efficient, enabling astronomers to gather more data in less time [**Spindler**]. Additionally, AI and automation have been used to identify and select objects of interest for further observation. This is particularly crucial in astronomy, where there are billions of stars and other celestial objects to observe. By using AI algorithms to analyze the data collected by telescopes, astronomers can quickly identify objects that are most likely to yield valuable scientific insights.

Being able to automate the process of selecting stellar objects to observe has become more desirable [**Ananthaswamy**]. The use of AI and automation has made this process more efficient by enabling astronomers to identify the objects they are interested in quickly. This approach allows them to concentrate their attention and resources on the most promising objects. Thus, the main objective of our project was to determine the most searched objects that general astronomy websites were discussing.

Program Structure:

- COMPONENTS: All Python scripts
- DATA: All data files
- IMAGES: Where all images used for websites and images downloaded from STScI database are stored
- PAGES: All HTML files
- SCRIPTS: All javascript files
- STYLES: All CSS files
- UNIVERSAL: All common header and footer across all pages

Data Collections and Methods:

1. Stage 1: Collect data from space news sites: `website_scraping.py`:
   - Websites: https://www.space.com/science-astronomy, https://www.sciencenews.org/topic/astronomy, https://www.nature.com/natastron/news-and-comment, https://phys.org/space-news/
   - Write article headlines and links to `headlines.txt` and `links.txt`
2. Stage 2: Process keywords `keyword_processing.py`
   - Use nltk for language and POS tagging
   - Find most popular topics `popular_topics()`:
     - Read `headlines.txt`
     - Get nouns with nltk
     - Write nouns in a dictionary and track count in headlines.
     - Exclude generic words, websites, org names
     - Get the top 10 words with the highest counts
     - Write these words in `popular_topics.txt` and their occurrences
3. Stage 3: Find object names from popular topics list: `back_search.py`
   - Use genism for summaries and object locating
     - Search all headlines for top words
     - Scrape articles from their corresponding links

- Create 40% sized summaries of each article to ease the process of scanning for specific objects to observe.
- Filter through summaries with a large list of object names accumulated from multiple databases
- Create list of final object names `list_objects.txt`

4. Stage 4: Search top topics in databases to find more information about the objects, the coordinates, and populate to HTML: `popular-object-to-html.py`:
   - Reverse search using keywords on websites
   - Get exact objects names
   - Get objects' coordinates from databases, can feed object name to $https://archive.stsci.edu/cgi-bin/dss\_form$, use Astropy
   - Determine if objects are observable, flag, and move to the next object if not observable
   - Use coordinates RA and Dec to get image from $https://archive.stsci.edu/cgi-bin/dss\_form$
   - Save information about objects: coordinates and images and update on HTML files
5. Stage 5: Create a user interface (website) that:
   - Displays images of the 3 objects that are mentioned the most frequently on the news within the day (catalog and originally taken), general information about each object (ie. each's location in the sky)
   - Includes a link to GitHub documentation
   - Includes an 'About Us' page, a page about the history of the stellarium telescope, a project 'README' page, and possibly additional informational pages

How the Program Runs:

Run `pip install -r requirements.txt` to install all the necessary packages.

This section will describe the steps of how to run the program. We decided that creating a `main.py` file that can be run each day and runs the files from each of the sections described above would be the best option to make this program as user-friendly as possible. This file calls functions from each back-end file, runs it, and moves on to the next step. We are currently working on creating the arguments needed for the command line option but the goal is that our website with the day's popular images would automatically open in the user's browser where they roam around on the interface.

The user interface starts with the title page (`title-page.html`) which, when clicked, transitions into the homepage or most popular object page, `object-1-page.html`. Here the user can choose to click between the three most popular objects of the day using the center buttons or surf through our information pages using the drop across the menu in the top right corner which gives the user access to the About Us page, (`about-us.html`), the Final Paper page (`final-paper-page.html`) and References page (`ref-site-page.html`).

GitHub button in the left bottom corner of our website that brings the user straight to the public GitHub repository and documentation.

The User Interface:

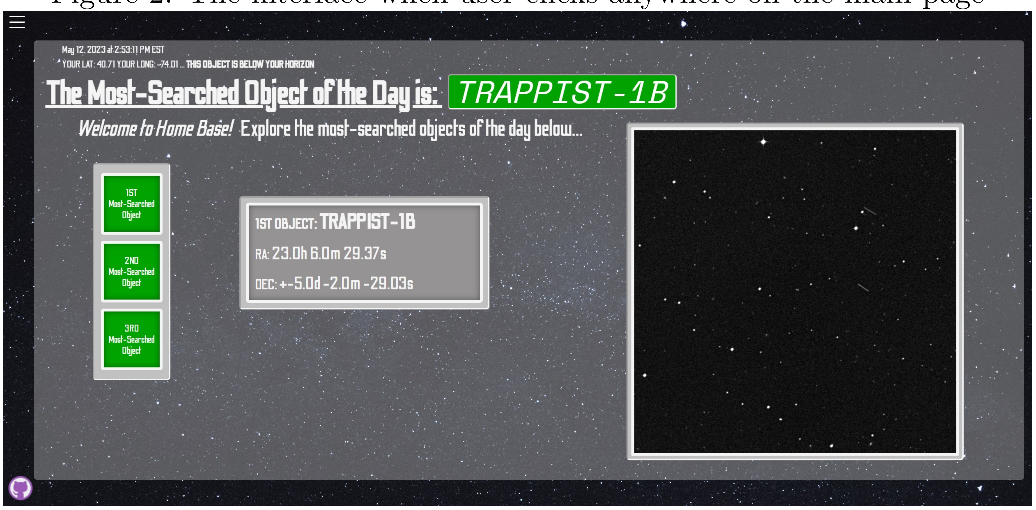
The user interface (website) of our project was designed to be easy to navigate and informative. It features seven pages, including a title page, three pages for each of the most-searched objects, an 'About Us' page, a page that contains this final paper, and a references page. Except for the title page, the website has a hamburger menu on every page, which makes it easy to navigate to any page from anywhere on the website. There is also a link to our Github repository on every page aside from the title page, which contains more technical details about our program.

When a user first lands on the title page, the program automatically takes them to the first most-searched object page upon clicking anywhere on the screen. The first object page displays the user's local time, their longitude and latitude, and information about whether or not the object is currently visible in their sky, calculated using astropy. This is followed by the name of the object, an easily navigable button table that can bring the user to each other object page, the name of the object again, its Right Ascension and Declination, and an image of it. The same format is repeated for the other two object pages, featuring respective information.

The reference page is continually updated with a list of all the online articles that were read and searched through to inform the AI Astronomer's decision about what the 'most-searched' objects of any given day were. This page provides users with a comprehensive list of resources used to develop our project, allowing them to explore further on their own. Figure 1,2,3,4 show samples of the interface.

Figure 1: The interface after main.py finishes running, shows the main page

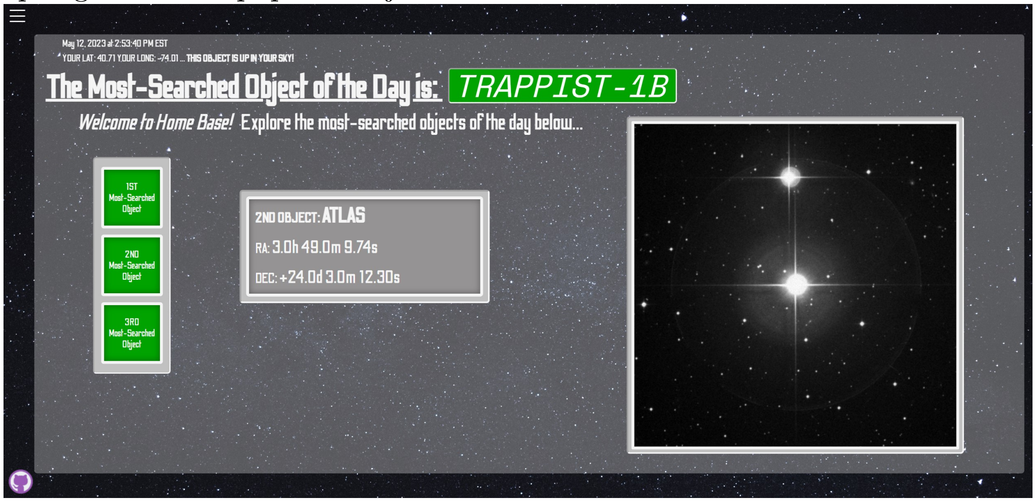
Figure 2: The interface when user clicks anywhere on the main page



Project Goals:

The goal of this project is to automate the process of retrieving popular stellar objects from space news websites, analyzing them, and quickly determining the most popular ones to observe. We argue that because the popular objects in question are popular amongst verifiable space news sources, they themselves are promising sources astronomically. The project aims to achieve this goal by utilizing advanced techniques in AI and automation.

Figure 3: The interface when user clicks the second button on the left-hand side, prompting 2nd most popular object



The first objective of the project is to develop a web scraping tool that can crawl popular space news websites and retrieve articles related to stellar objects. The tool will then use natural language processing techniques to identify and extract the names of the stellar objects mentioned in the articles.

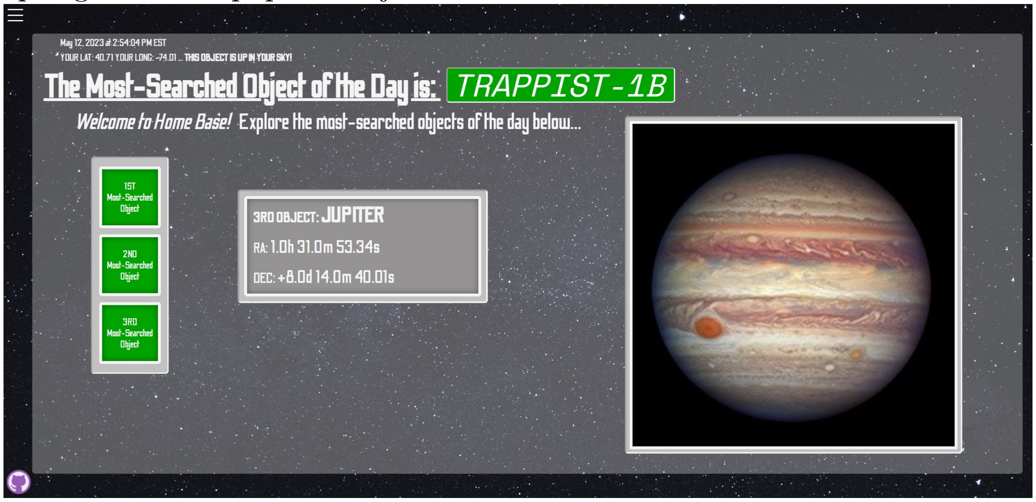
The second objective is to analyze the data collected by the web scraping tool to determine the most popular stellar objects among the articles retrieved. This will be achieved by using natural language processing to tag keywords mentioned in headlines to process the data and identify most mentioned objects in the articles.

The third objective is to search for the popular stellar objects identified in the second objective within the Space Telescope Science Institute (STScI) database using regular expression names. The tool will retrieve the International Celestial Reference System (ICRS) coordinates of the objects, their images, and related articles/research papers from the STScI database.

The final objective is to present the data retrieved in a user-friendly interface that allows astronomers to quickly identify the most popular and promising stellar objects to observe. The interface will provide users with a dashboard that displays the ICRS coordinates, images, and related articles/research papers of the popular stellar objects.

Overall, the project aims to streamline the process of identifying and observing

Figure 4: The interface when user clicks the third button on the left-hand side, prompting 3rd most popular object



popular stellar objects by automating the data retrieval and analysis process. By achieving these objectives, the project will enable astronomers to make more efficient use of their time and resources and increase the likelihood of making groundbreaking discoveries in the field of astronomy.