ASTRUN 3646 Rhessa Weber Langstaff, Ngan Bao, Alexandra Savino

Project Phase II

Main project topic

The Robot Astronomer — We are building our own AI astronomer which determines the most searched objects in a given day, gets the coordinates of said object, AND decides what to observe beyond that. (The project will include an easily navigational and understandable website that consists of all relevant information.)

Background

The field of astronomy has been revolutionized by advances in Artificial Intelligence (AI) and automation. These technologies have greatly improved our ability to collect and analyze vast amounts of astronomical data, enabling us to make significant discoveries about the universe.

One of the primary applications of AI and automation in astronomy is the collection of data. With the development of advanced telescopes and sensors, astronomers can now capture vast amounts of data about the universe. However, processing and analyzing this data manually can be a time-consuming and labor-intensive process. AI and automation have therefore been used to streamline the data collection process, allowing astronomers to gather more data in less time. In addition to collecting data, AI and automation have also been used to identify and select objects of interest for further observation. This is particularly important 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 and sensors, astronomers can quickly identify objects that are most likely to yield valuable scientific insights.

Auto-selecting stellar objects to observe has become an important aspect of modern astronomy. AI and automation have made this process more efficient by enabling astronomers to quickly identify objects they are interested in, allowing them to focus their attention and resources on the most promising objects.

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
     - 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 brief information about objects, coordinates and images and update information on HTML files
5. Stage 5: Create a user interface (website) that:
   - Displays images of the most-searched objects (catalog and originally taken), general information about each object (ie. each's location in the sky), and information about how our original images will have been taken (ie. the time that an image was taken, etc)
   - Referential links to searched papers/articles
   - 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`).

Documentation can be found on each page using the little GitHub button in the left bottom corner that brings the user straight to the public GitHub repository.

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.

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 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.