Maastricht University

Department of Advanced Computing Sciences

Building and Mining Knowledge Graphs

Project Proposal – Hunting for Exoplanets

Author:
Aurélien Bertrand

Student number:
i6256590

February 24, 2024

1 Significance

The interest in space exploration has significantly grown over time. The historic Moon landing in 1969 marked a turning point that strengthened a dedicated commitment to improve our understanding of the universe. Nowadays, the space industry is partly focused on Mars, with missions like the Mars Exploration Rovers (2003) and the Perseverance rover (2021) exploring Martian soil. However, the broader scope has shifted towards exploring planetary bodies beyond our solar system.

Exoplanets refer to planets located outside our solar system. Over the last two decades, thousands of them have been identified, predominantly through NASA's Kepler Space Telescope (2010). Scientists prioritize hunting for a specific type of exoplanet: those resembling Earth in size, orbiting a star similar to our sun within the habitable zone. Different methodologies such as analyzing the age, atmosphere, and composition of exoplanets can be used to understand and identify planets similar to ours.

Despite the excitement about space exploration, ethical concerns emerge due to human pollution. It already poses a threat to our planet and its surroundings. This will likely extend to other planets and likely destroy the habitats of potential extraterrestrial species.

This project suggests a framework for building RDF graphs using planetary bodies' data. It also introduces queries designed to identify the main characteristics of exoplanets as well as the exoplanet that resembles the most Earth. I will compare my findings to the current knowledge and argue about the usability of the graph mining method compared to state-of-the-art methods.

2 Related work

State-of-the-art methods for detecting exoplanets involve various techniques such as such as radial velocity (2018), transit photometry (2018), direct imaging (2023), and relativistic beaming (2013), among others. Recently, Sekhar et al. (2023) introduced a novel approach to exoplanet detection using the You Only Look Once (YOLO) model, one of the most popular algorithms for object detection. Their method, when compared to existing algorithms, demonstrates significantly higher accuracy in predicting brightness dips in light curves and requires less training time.

Although extensive research has been conducted on exoplanet identification, a specific focus is on identifying Earth-like planets which may be habitable. To date, NASA reports over 5,000 identified exoplanets (2024), with around 200 potentially Earth-like (2024). Capistrant et al. (2024) present the discovery of HD 63433 d-a nearby (22 pc), an exoplanet with a similar size to Earth that orbits around a young Sun-like star (TOI-1726, HD 63433). At the time of writing, they reported that HD 63433 d is the confirmed nearest young Earth-sized planet.

Many of the existing studies employ specialized methods that may not be easily accessible to people with general knowledge. In response to this common limitation, this project suggests a simple approach, based on the use of planetary systems data (2020), combined with Kepler object of interest (2019) data, to offer a clear and comprehensible understanding for a broader audience.

3 Goal and specific objectives

The goal of this project is to investigate how to leverage knowledge graphs to accurately represent and use exoplanet data and infer rules to answer key research questions. I will focus on making the approach accessible to general knowledge people. The specific objectives of this project are:

- Implement a pipeline to access NASA's data through their API.

- Create an RDF schema for both datasets and link them.
- Implement queries to retrieve information from the graph to answer the research questions.

4 Methodology

I will start the project by setting up a mechanism to access NASA's API and establish calls to gather the necessary data. I justify this decision based on my lack of experience in this area, providing an opportunity for me to acquire new skills. Should this approach be challenging, I will resort to downloading files, albeit potentially resulting in a smaller dataset due to download size limitations.

Following this, I will formulate a vocabulary for constructing my RDF graph. I will explore optimal methods for integrating both datasets, considering options such as merging tables or creating shared nodes within the graph. To ensure interoperability, I will meticulously verify the use of appropriate resources. Subsequently, I will develop a pipeline to automatically and systematically convert the data into their RDF representations.

Finally, I will study the graph to find strategies for addressing my research questions. Then, I will write queries to get insights and answer as accurately as possible the research questions.

For future work, potential approaches may include incorporating updated datasets, integrating supplementary planetary data, or merging with other pertinent datasets.

5 Milestones and Deliverables

In this section, I detail the main checkpoints of the project which I will use to measure my progress throughout the work. I will set up a public and documented GitHub repository.

1. **Data Collection Complete (Week 3)**: The procedure to access NASA's API is implemented. The data is collected and stored. A thorough analysis of the data is also performed, enabling me to better understand the data and the present columns.
2. **RDF Schema Complete (Week 4)**: A vocabulary is clearly and neatly created for the data. The RDF schema is also completed. Lastly, the RDF graph is created for both datasets. A link between the two is generated.
3. **Research Questions Answers (Week 5)**: Queries are written to explore the data. Other queries (or potentially machine learning methods) are used to specifically answer the research questions.
4. **Documentation Complete (Week 6)**: The code is neatly documented in a Jupyter notebook and a technical report is written. These are the deliverables.

6 Anticipated results

By the end of the project, I aim to create a knowledge graph representation specifically for exoplanet data. The main outputs are expected to provide new insights into exoplanetary systems, facilitated by a user-friendly and interpretable method. This initiative addresses the challenge of accessibility for non-experts, potentially leading to broader participation in exoplanet research and fostering a collaborative and inclusive approach to advancing our understanding of these celestial bodies.

References

Borucki, W., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W., Devore, E., Dunham, E., Dupree, A., Gautier, T., Geary, J., Gilliland, R., Gould, A., Howell, S., Jenkins, J., Kondo, Y., and Prsa, A. (2010). Kepler planet-detection mission: Introduction and first results. *Science (New York, N.Y.)*, 327:977–80.

Capistrant, B. K., Soares-Furtado, M., Vanderburg, A., Jankowski, A., Mann, A. W., Ross, G., Srdoc, G., Hinkel, N. R., Becker, J., Magliano, C., Limbach, M. A., Stephan, A. P., Nine, A. C., Tofflemire, B. M., Kraus, A. L., Giacalone, S., Winn, J. N., Bieryla, A., Bouma, L. G., Ciardi, D. R., Collins, K. A., Covone, G., de Beurs, Z. L., Huang, C. X., Jenkins, J. M., Kreidberg, L., Latham, D. W., Quinn, S. N., Seager, S., Shporer, A., Twicken, J. D., Wohler, B., Vanderspek, R. K., Yarza, R., and Ziegler, C. (2024). Tess hunt for young and maturing exoplanets (thyme). xi. an earth-sized planet orbiting a nearby, solar-like host in the 400 myr ursa major moving group. *The Astronomical Journal*, 167(2):54.

Crisp, J., Adler, M., Matijevic, J., Squyres, S., Arvidson, R., and Kass, D. (2003). Mars exploration rover mission. *J. Geophys. Res.*, 108.

Currie, T., Biller, B., Lagrange, A.-M., Marois, C., Guyon, O., Nielsen, E., Bonnefoy, M., and Rosa, R. D. (2023). Direct imaging and spectroscopy of extrasolar planets.

Deeg, H. J. and Alonso, R. (2018). *Transit Photometry as an Exoplanet Discovery Method*, page 633–657. Springer International Publishing.

Faigler, S., Tal-Or, L., Mazeh, T., Latham, D. W., and Buchhave, L. A. (2013). Beer analysis ofkeplerandcorotlight curves. i. discovery of kepler-76b: A hot jupiter with evidence for superrotation. *The Astrophysical Journal*, 771(1):26.

Green, J. (2021). Perseverance rover and its search for life on mars. *Communications of the Byurakan Astrophysical Observatory*, pages 464–469.

NASA (2024). Exoplanet Exploration: Planets Beyond our Solar System. https://exoplanets.nasa.gov/. Accessed: 22 February 2024.

NASA Exoplanet Science Institute (2019). Kepler objects of interest cumulative table. https://catcopy.ipac.caltech.edu/dois/doi.php?id=10.26133/NEA4. Version: 2024-02-24 15:45.

NASA Exoplanet Science Institute (2020). Planetary systems table. https://catcopy.ipac.caltech.edu/dois/doi.php?id=10.26133/NEA12. Version: 2024-02-24 15:45.

Pat Brennan (2024). Cosmic Milestone: NASA Confirms 5,000 Exoplanets. https://exoplanets.nasa.gov/news/1702/cosmic-milestone-nasa-confirms-5000-exoplanets/. Accessed: 24 February 2024.

Sekhar, M., Tejas, C., Kanna, V., and Kaur, A. (2023). Finding exoplanets using object detection. *Astrophysics and Space Science*, 368.

Wright, J. T. (2018). *Radial Velocities as an Exoplanet Discovery Method*, page 619–631. Springer International Publishing.