PARSEC: Propelling Astrophysics Research through Semantically Enriched Catalogues

Giorgos Flouris[1*], Andreas Zezas[2,3*#], Vassilis Efthymiou[1,4], Iosif Oikonomakis[1], Katerina Papantoniou[1], Ariadne Pratikaki[1,] Elias Kyritsis[2,3,5], Ioanna Leonidaki[2,3], Michael Mavrommatis[2,3], Dimitrios Rompogiannakis[2,3]

[1] Institute of Computer Science, FORTH

[2] Institute of Astrophysics, FORTH

[3] Department of Physics, University of Crete

[4] Department of Informatics and Telematics Harokopio University of Athens

[5] Max Planck Institute for Extraterrestrial Physics

Presenting author: Andreas Zezas, azezas@physics.uoc.gr
* Corresponding author: Andreas Zezas, azezas@physics.uoc.gr Giorgos Flouris, fgeo@ics.forth.gr

ABSTRACT

In the era of major Astronomical surveys and multi-messenger Astrophysics the curation of multi-wavelength catalogues becomes a critical but increasingly challenging process. This is due to the continuing production of new data and measurements (either of direct or derived properties), and the diversity of these results which does not allow their automated ingestion in databases. As a step toward the automated curation of astronomical catalogues we embarked in the project "Propelling Astrophysics Research through Semantically Enriched Catalogues" (PARSEC). This is a FORTH "Theodore Papazoglou" Synergy grant between the Institute of Computer Science and the Institute of Astrophysics.

The backbone of this project is an ontology of astrophysics concepts pertaining to measurements and derived parameters characterizing astronomical objects of different types. This ontology is combined with an automated catalogue-mining and a catalogue-annotation procedure that combine a decision-making tree and the use of Large Language Models to produce a semantically annotated catalogue. This procedure also involves: the identification of data units and their homogenization in a common unit system, the automated detection of "magic numbers" (used as flags), and a structured assessment procedure which as a last step has the manual review of issues by an "experts committee". This committee also proposes developments in the data ingestion pipeline. All the above result in a seamless workflow that automatically ingests catalogs in a database, automatically identifies issues, proposes corrective actions, and only when such actions cannot be identified automatically its requires intervention by experts.

In the development phase of this project, we focused on applications on two different types of astronomical objects: (a) Supernova Remnants, and (b) Galaxies, allowing us to define a broad array of functionalities. The use case of Supernova Remnants requires the definition of hierarchical objects consisting of a parent object containing multiple subregions each with its own characteristics and measurements. The use case of Galaxies requires the definition of multiple and diverse (and often interconnected) concepts and the ability for easy implementation of new concepts to accommodate the fast-evolving landscape of Extragalactic Astrophysics. The combination of the two use cases places the foundation for the extension of the PARSEC system to a wide array of Galactic and extragalactic systems with multi-region measurements (e.g. star-clusters, spatially-resolved galaxies, clusters of galaxies etc). The catalogue is combined with a user-friendly front-end that allows the user to perform complex queries.