

Title:

Automated development of an astrophysical semantic catalogue in PARSEC

Authors:

Sifis Oikonomakis, FORTH-ICS (sifisoik@ics.forth.gr)

Vasilis Efthymiou, FORTH-ICS (vefthym@ics.forth.gr)

Giorgos Flouris, FORTH-ICS (fgeo@ics.forth.gr)

Elias Kyritsis, FORTH-IA (ekyritsis@physics.uoc.gr)

Ioanna Leonidaki, FORTH-IA (ioanna@physics.uoc.gr)

Michael Mavrommatis, FORTH-IA (ph5604@edu.physics.uoc.gr)

Dimitrios Rompogiannakis, FORTH-IA (ph4502@edu.physics.uoc.gr)

Andreas Zezas, FORTH-IA (azezas@physics.uoc.gr)

Abstract:

PARSEC is a research project that aims to develop and apply a machine-assisted methodology for constructing and curating semantically enriched astrophysical catalogues in the form of knowledge graphs (KGs).

The PARSEC semantic catalogue is used to model and store knowledge recorded in astrophysical catalogues via ontologies represented using RDF/S and RDF-star. The former allows the representation of knowledge as triples, e.g., (GalaxyX has\_distance 500), that can be visualised as graphs, thereby forming KGs. RDF-star allows representing metadata about triples; this is extremely useful to record spatio-temporal or provenance information, e.g., the triple (GalaxyX has\_distance 500) was obtained from paper Y.

The main challenges associated with the development of the ontology are related to the richness and diversity of knowledge found in astrophysics papers, the sheer size of the data, and the characteristics of astrophysical data pertaining to the need for homogenization, to the uncertainty and errors in measurements, to missing data and magic numbers, etc.

Another important component of PARSEC is the automated population of the developed KG, using sub-symbolic AI methods that mine this information from tables found in publications and astrophysical catalogues.

The developed KG will be analysed by astrophysicists to generate new knowledge and scientific insights. The automation of KG population, in tandem with the richness of the information that is stored in the knowledge graph, will arguably speed up astrophysical research. Two use cases have been envisioned, one related to galaxies in distances up to 500 Mpc, and one related to SNRs in our galaxy.