CosmOntology: Creating an Ontology of the Cosmos

Vasilis Efthymiou

FORTH-ICS, N. Plastira 100, GR 70013, Heraklion, Crete, Greece

Abstract

Today, we are gathering more and more new astrophysics knowledge, but we are using obsolete ways of representing and processing it. This position paper discusses potential ways of constructing an astronomical KG, semantically annotating and reasoning on such data, using neuro-symbolic methods.

1. Introduction

AI is already revolutionizing knowledge acquisition and management, allowing computers to understand and process resources that are otherwise consumable only by humans. A common requirement for such technologies to operate successfully is a machine-readable conceptual modelling of the domain of interest. Having such a model enables, in turn, reasoning on the data to infer new knowledge. Typically, domain-specific models (e.g., SNOMED in healthcare, GeneOntology in genomics) are stored in knowledge graphs (KGs), formatted as ontologies.

Astrophysics data are stored in various online catalogues[1] and tables within research papers. Those data sets are frequently overlapping and they may not even be in a machine-readable format, which means that human experts should manually read and cross-check which data refer to the same object, so they should be combined. On the other hand, existing ontologies often contain inconsistent information about common astronomical concepts. E.g., Wikidata defines an Active Galactic Nucleus as a subclass of Galaxy, while DBpedia defines it as a subclass of Settlement. The lack of a shared understanding of common concepts and their relationships, requiring hard human labor to detect and utilize the available data, decreases the value of data.

In 2010, Borne [1] suggested the use of ontologies as a meta-data layer for the information extracted from astronomical data. In parallel, the IVOA group was studying what expressivity and reasoners would fit for such an ontology, but stopped reporting progress in 2010 [2]. Sarro and Martinez [3] suggested some first steps for generating an ontology for astrophysics, concluding that it remains unclear how this modeling can be applied.

2. Managing Astrophysics Data with CosmOntology

In recent years, there have been a plethora of works that try to exploit the potential of tables available on the Web for a multitude of applications, ranging from knowledge base augmentation

DL4KG: Deep Learning for Knowledge Graphs Workshop at ISWC 2022
vefthym@ics.forth.gr (V. Efthymiou)

[1]E.g.: vizier.u-strasbg.fr/, simbad.u-strasbg.fr/, ned.ipac.caltech.edu/, cosmos.esa.int/web/esdc

to question answering, schema linking, and data integration [4, 5, 6]. The first steps towards constructing a rich KG (an ontology) of the cosmos can be the extraction [7, 8], semantic annotation [6], partial KG creation [9, 10, 11] and unification [12, 13] of knowledge from tables found on scientific publications (e.g., astro-ph within arXiv), as well as online catalogues.

In more detail, for the generation and curation of CosmOntology, a human-in-the-loop approach is probably more promising, in which a board of experts will handle requests to update the ontology. Such requests will be made both by algorithms and by humans. A set of many shallow ontologies can be generated initially from structured data, which will be later enriched by combining the many, shallow ontologies into one [14], through ontology matching [15] and a consensus-based ontology curation platform (e.g., based on [16]).

A recent work [17] that exploits textual information from AI publications with state-of-the-art NLP tools, has shown very promising results. A combination of such tools that process textual, tabular, and image data in the field of astrophysics would set new standards in mining knowledge available online and modeling it in a unified way.

Applications. A unified KG modeling of astronomical concepts will allow a number of AI tools and methods to become available. For example, *accessing astrophysics data* even with natural language interfaces [18], like query answering systems or even chatbots, can become seamless. *Graph data analysis tasks*, such as clustering, node classification, and link prediction have seen significant advances in the presence of ontologies (e.g., [13, 19]). Such tasks can be utilized for generating new data insights and visualizations, and boost new scientific discoveries. An ontology of astronomical concepts can constitute a global point of reference for astrophysicists, computer scientists and practitioners that work with astrophysics data. Furthermore, *pre-trained KG embeddings* can become publicly available to facilitate such tasks.

Some of the reasoning tasks that could be enabled after such a KG construction are: managing inconsistencies [20, 21] that may arise after matching and curation, enriching question answering based on inferred knowledge from the constructed ontology [15], KG summarization and modular reuse of ontologies [22, 23].

Provenance and *explainability* are also major issues for astronomers, determining the credibility of the underlying KG. Data and workflow provenance information can be used for explainability, answering questions like "WHY am I (not) seeing this result?" and "HOW was this data acquired?" (e.g., with which instruments, under which conditions), respectively.

3. Concluding Remarks

This paper discusses some possible steps for the construction of a KG that captures our knowledge of the cosmos, mostly relying on tabular data found in astrophysics publications and catalogues. Such a KG will boost research in astrophysics, widen the public knowledge in astronomy, and pose new challenges that can further improve research in computer science. For example, the resources produced by this effort could be used to improve deep-learning tools that mine such data, or semantically annotate them (e.g., systems participating in SemTab). The processes followed in this endeavor should be thoroughly documented, with the ultimate goal to create a generalizable, open-access methodology that gets adopted by other disciplines.

References

[1] K. D. Borne, Astroinformatics: data-oriented astronomy research and education, Earth Sci. Informatics 3 (2010) 5–17.

[2] Ontology of astronomical object types version 1.3, 2010. URL: https://www.ivoa.net/documents/Notes/AstrObjectOntology/.

[3] L. M. Sarro, R. Martínez-Tomás, First steps towards an ontology for astrophysics, in: KES, 2003.

[4] D. Ritze, O. Lehmberg, Y. Oulabi, C. Bizer, Profiling the potential of web tables for augmenting cross-domain knowledge bases, in: WWW, 2016.

[5] O. Lehmberg, O. Hassanzadeh, Ontology augmentation through matching with web tables, in: OM@ISWC, 2018.

[6] V. Cutrona, J. Chen, V. Efthymiou, O. Hassanzadeh, E. Jiménez-Ruiz, J. Sequeda, K. Srinivas, N. Abdelmageed, M. Hulsebos, D. Oliveira, C. Pesquita, Results of semtab 2021, in: SemTab@ISWC, 2021.

[7] A. Gilani, S. R. Qasim, M. I. Malik, F. Shafait, Table detection using deep learning, in: ICDAR, 2017.

[8] N. X. R. Wang, D. Burdick, Y. Li, Tablelab: An interactive table extraction system with adaptive deep learning, in: IUI, 2021.

[9] C. Lei, F. Özcan, A. Quamar, A. R. Mittal, J. Sen, D. Saha, K. Sankaranarayanan, Ontology-based natural language query interfaces for data exploration, IEEE Data Eng. Bull. 41 (2018) 52–63.

[10] H. M. Zahera, S. Heindorf, S. Balke, J. Haupt, M. Voigt, C. Walter, F. Witter, A. N. Ngomo, Tab2onto: Unsupervised semantification with knowledge graph embeddings, in: ESWC, 2022.

[11] D. Chaves-Fraga, A. Dimou, Declarative description of knowledge graphs construction automation: Status & challenges, in: KGCW@ESWC, 2022.

[12] S. Mudgal, H. Li, T. Rekatsinas, A. Doan, Y. Park, G. Krishnan, R. Deep, E. Arcaute, V. Raghavendra, Deep learning for entity matching: A design space exploration, in: SIGMOD, 2018.

[13] J. Hao, C. Lei, V. Efthymiou, A. Quamar, F. Özcan, Y. Sun, W. Wang, MEDTO: medical data to ontology matching using hybrid graph neural networks, in: KDD, 2021.

[14] H. Alani, Position paper: ontology construction from online ontologies, in: WWW, 2006.

[15] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, M. Zakharyaschev, Ontology-based data access: A survey, in: IJCAI, 2018.

[16] T. Patkos, G. Flouris, A. Bikakis, Symmetric multi-aspect evaluation of comments - extended abstract, in: ECAI, 2016.

[17] D. Dessì, F. Osborne, D. R. Recupero, D. Buscaldi, E. Motta, H. Sack, AI-KG: an automatically generated knowledge graph of artificial intelligence, in: ISWC, 2020.

[18] A. Quamar, V. Efthymiou, C. Lei, F. Özcan, Natural language interfaces to data, Found. Trends Databases 11 (2022) 319–414.

[19] J. Chen, P. Hu, E. Jiménez-Ruiz, O. M. Holter, D. Antonyrajah, I. Horrocks, Owl2vec*: embedding of OWL ontologies, Mach. Learn. 110 (2021) 1813–1845.

[20] Z. Huang, F. van Harmelen, A. ten Teije, Reasoning with inconsistent ontologies, in: IJCAI,

2005.
[21] S. Schlobach, R. Cornet, Non-standard reasoning services for the debugging of description logic terminologies, in: IJCAI, 2003.
[22] B. C. Grau, I. Horrocks, Y. Kazakov, U. Sattler, Modular reuse of ontologies: Theory and practice, J. Artif. Intell. Res. 31 (2008) 273–318.
[23] A. Bonifati, S. Dumbrava, H. Kondylakis, Graph summarization, CoRR abs/2004.14794 (2020).