TACTICIAN: AI-based applications for knowledge extraction from ESA's missions' scientific publications

Omiros Giannakis[1], Iason Demiros[2], Vassilios Antonopoulos[2], Guido De Marchi[3], Christophe Arviset[4], Athanasios Daglis[1], George Vasalos[1], Zoe Boutsi[1], Jan Tauber[3], Marcos Lopez-Caniego[4], Mark Kidger[4], Arnaud Masson[4], Philippe Escoubet[3], George Balasis[1]

[1]Institute For Astronomy, Astrophysics, Space Applications And Remote Sensing / National Observatory Of Athens, Penteli, Greece, [2]11Tensors, Athens, Greece, [3]European Space Research and Technology Centre, European Space Agency, the Netherlands, [4]European Space Astronomy Center, European Space Agency, Spain

Scientific publications in space science contain valuable and extensive information regarding the links and relationships between the data interpreted by the authors and the associated observational elements (e.g., instruments or experiments names, observing times, etc.). In this reality of scientific information overload, researchers are often overwhelmed by an enormous and continuously growing number of articles to access in their daily activities. The exploration of recent advances concerning specific topics, methods and techniques, the review and evaluation of research proposals and in general any action that requires a cautious and comprehensive assessment of scientific literature has turned into an extremely complex and time-consuming task.

The availability of Natural Language Processing (NLP) tools able to extract information from scientific unstructured textual contents and to turn it into extremely organized and interconnected knowledge, is fundamental in the framework of the use of scientific information. Exploitation of the knowledge that exists in the scientific publications, necessitates state-of-the-art NLP. The semantic interpretation of the scientific texts can support the development of a varied set of applications such as information retrieval from the texts, linking to existing knowledge repositories, topic classification, semi-automatic assessment of publications and research proposals, tracking of scientific and technological advances, scientific intelligence-assisted reporting, review writing, and question answering.

The main objectives of TACTICIAN are to introduce Artificial Intelligence (AI) techniques to the textual analysis of the publications of all ESA Space Science missions, to monitor and evaluate the scientific productivity of the science missions, and to integrate the scientific publications' metadata into the ESA Space Science Archive. Through TACTICIAN, we extract lexical, syntactic, and semantic information from the scientific publications by applying NLP and Machine Learning (ML) algorithms and techniques. Utilizing the wealth of publications, we have created valuable scientific language resources, such as labeled datasets and word embeddings, which were used to train Deep Learning models that assist us in most of the language understanding tasks. In the context of TACTICIAN, we have devised methodologies and developed algorithms that can assign scientific publications to the Mars Express, Herschel, and Cluster ESA science missions and identify selected named entities and observations in these scientific publications.

These methodologies can be applied to any other mission. The combination of NLP and ML constitutes a general basis, which has proved that it can assist in establishing links between the missions' observations and the scientific publications and to classify them in categories, with high accuracy.

This work has received funding from the European Space Agency under the "ArTificiAl intelligenCe To lInk publiCations wIth observAtioNs (TACTICIAN)" activity under ESA Contract No 4000128429/19/ES/JD.