World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:12, No:06, 2018

# Knowledge Graph Development to Connect Earth Metadata and Standard English Queries

**Authors :** Gabriel Montague, Max Vilgalys, Catherine H. Crawford, Jorge Ortiz, Dava Newman

**Abstract :** There has never been so much publicly accessible atmospheric and environmental data. The possibilities of these data are exciting, but the sheer volume of available datasets represents a new challenge for researchers. The task of identifying and working with a new dataset has become more difficult with the amount and variety of available data. Datasets are often documented in ways that differ substantially from the common English used to describe the same topics. This presents a barrier not only for new scientists, but for researchers looking to find comparisons across multiple datasets or specialists from other disciplines hoping to collaborate. This paper proposes a method for addressing this obstacle: creating a knowledge graph to bridge the gap between everyday English language and the technical language surrounding these datasets. Knowledge graph generation is already a well-established field, although there are some unique challenges posed by working with Earth data. One is the sheer size of the databases – it would be infeasible to replicate or analyze all the data stored by an organization like The National Aeronautics and Space Administration (NASA) or the European Space Agency. Instead, this approach identifies topics from metadata available for datasets in NASA's Earthdata database, which can then be used to directly request and access the raw data from NASA. By starting with a single metadata standard, this paper establishes an approach that can be generalized to different databases, but leaves the challenge of metadata harmonization for future work. Topics generated from the metadata are then linked to topics from a collection of English queries through a variety of standard and custom natural language processing (NLP) methods. The results from this method are then compared to a baseline of elastic search applied to the metadata. This comparison shows the benefits of the proposed knowledge graph system over existing methods, particularly in interpreting natural language queries and interpreting topics in metadata. For the research community, this work introduces an application of NLP to the ecological and environmental sciences, expanding the possibilities of how machine learning can be applied in this discipline. But perhaps more importantly, it establishes the foundation for a platform that can enable common English to access knowledge that previously required considerable effort and experience. By making this public data accessible to the full public, this work has the potential to transform environmental understanding, engagement, and action.

**Keywords :** earth metadata, knowledge graphs, natural language processing, question-answer systems
**Conference Title :** ICMLMEES 2018 : International Conference on Machine Learning Methods in Ecological and Environmental Sciences
**Conference Location :** Copenhagen, Denmark
**Conference Dates :** June 11-12, 2018

Open Science Index, Environmental and Ecological Engineering Vol:12, No:06, 2018 publications.waset.org/abstracts/90830/pdf