



Methodology to Build Scalable Knowledge Graphs for Earth Science

Rahul Ramachandran (1), Manil Maskey (1), Patrick Gatlin (1), Jia Zhang (2), Xiaoyi Duan (2), Jeffrey Miller (3), Kaylin Bugbee (3), and Sundar Christopher (3)

(1) NASA/MSFC, Huntsville, United States (rahul.ramachandran@nasa.gov), (2) CMU Silicon Vally, United States, (3) University of Alabama in Huntsville, United States

Estimates indicate that the world's information will grow by 800% in the next five years. In any given field, a single researcher or a team of researchers cannot keep up with this rate of knowledge expansion without the help of cognitive systems. Cognitive computing, defined as the use of information technology to augment human cognition, can help tackle large systemic problems. Knowledge graphs, one of the foundational components of cognitive systems, link key entities in a specific domain with other entities via relationships. Researchers could mine these graphs to make probabilistic recommendations and to infer new knowledge.

However, there is a dearth of tools to generate scalable Knowledge graphs using existing corpus of scientific literature for Earth science research. Our project is currently developing an end-to-end automated methodology for incrementally constructing Knowledge graphs for Earth Science.

Two key components make our methodology unique and innovative. First, the Semantic Entity Recognition for Earth Science uses external resources (including metadata catalogs and controlled vocabulary) as references to guide entity extraction and recognition (i.e. labeling) from unstructured text, in order to build a large training set to seed the subsequent auto-learning component. This process is both objective and scalable as it requires minimum manual interactions. Second, the Deep Learning Powered Auto-learning goes beyond heuristics-based entity recognition and leverages state-of-the-art machine learning to incrementally learn and refine the rules and patterns through iterations. Results from several SER experiments will be presented as well as lessons learned.