

LSKnowledge: Nexus for transformative scientific discoveries and enhanced information retrieval in NASA Life Sciences Portal

Hamed Valizadegan^{1,2}, Michael von Pohle^{1,2}, Adwait N. Sahasrabhojane^{1,2}, Janani S. Iyer^{1,2}, Sandeep D. Shetye², Dan Berrios³, Adrienne T. Hoyt^{3,4}, Truong Le³

¹Universities Space Research Association, ²NASA Ames Research Center, ³NASA Johnson Space Center, and ⁴KBR (University of Houston)

We stand at the brink of an extraordinary transformation in the field of AI, driven by the convergence of generative AI and semantic technologies (e.g., knowledge graphs). This fusion holds immense potential and could redefine the future of scientific exploration, particularly in the realm of life sciences research. In this context, we shed light on the pivotal roles that Large Language Models (LLMs) and semantic technologies will play in advancing research, unearthing and comprehending life sciences information through innovative approaches, and empowering researchers to extract insights from NASA's extensive Life Sciences Data Archive.

Within the NASA Life Sciences Portal (NLSP), the integration of LLMs and semantic technologies unlocks several advanced capabilities. First and foremost, it equips scientists with sophisticated tools to manage the ever-expanding wealth of scientific literature and data. Furthermore, it facilitates the creation of knowledge graphs that visually represent intricate relationships among biological entities, enabling comprehensive systems-level analysis. Additionally, the fusion of generative AI (including LLMs) and semantic technology can significantly benefit NASA's life sciences research by enhancing information retrieval and hypothesis generation. These tools enhance natural language understanding, facilitating knowledge discovery within NLSP.

The overarching vision is to establish a cohesive knowledge ecosystem within NLSP, harnessing the power of LLMs and semantic technologies to synthesize and cross-reference data from diverse missions, disciplines, and research domains. This holistic approach ultimately deepens our understanding of how space environments impact life sciences data.

To advance this initiative, we have launched LSKnowledge, aimed at enhancing the information retrieval capabilities of NLSP. In the short term, our primary goal is to develop a robust semantic search system. This system will empower HRP (Human Research Program) researchers to navigate NLSP data repositories more efficiently and precisely, catalyzing the process of hypothesis formation and scientific breakthroughs. To achieve this, we have employed pre-trained LLMs as part of a semantic search tool that can rank and highlight the most relevant records for user queries. To assess the tool's performance, we have curated a set of approximately 200 queries from subject matter experts (SMEs) and manually ranked the top records retrieved by both the current search system and the new semantic search, using SME judgments as the gold standard for relevancy. Herein, we present the results of our comparative analysis and illustrate how these findings have informed the fine-tuning of the system for enhanced performance. In the long term, our objectives include 1) retrieving publicly available information and integrating it with NLSP data to provide more precise answers to user queries, and 2) incorporating non-textual information from the NLSP database into our approach.

In conclusion, the fusion of LLMs and semantic technologies within NLSP represents a pioneering stride towards reshaping the landscape of scientific discovery. This synergy not only equips researchers with powerful tools to navigate the burgeoning sea of information but also facilitates a deeper understanding of complex biological relationships, all while accelerating hypothesis generation and knowledge discovery. Through our initiative, LSKnowledge, we are committed to continually refining and expanding these capabilities, with the aim of not only enhancing information retrieval but also integrating diverse data sources to provide more precise insights. In the grand vision, NLSP strives to become the cornerstone of a comprehensive knowledge ecosystem, unraveling the enigmatic intricacies of life sciences phenomena in the context of space environments.