



Contribution ID: 94

Type: Oral

The NucScholar project: an AI/ML framework for compiling and searching the nuclear science literature

Literature search engines have become an integral and indispensable part of academic research. Search engines like google scholar rely on powerful AI/ML tools to return results that match the user's intended meaning as closely as possible, but they are better suited for general-purpose queries. As a result, they tend to overwhelm the user searching for domain-specific information with irrelevant hits. For this reason, the Nuclear Science References (NSR) database [1], hosted and maintained by the National Nuclear Data Center at Brookhaven National Laboratory (BNL), has become a standard search engine in the field. The goal of the NucScholar project is to build on the capabilities of NSR by automating onerous archival tasks needed to construct a database of nuclear science journal articles, and by augmenting the search/retrieval interface through the use of natural language processing (NLP) tools.

In this talk, I will present the current status of the project. I will focus in particular on the automated archiving functionality and on the NLP capability to perform both semantic and question-answering searches, and their implementation using deep learning models.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Berkeley National Laboratory under Contract DE-AC02-05CH11231. This work was supported in part by the U.S. Department of Energy, National Nuclear Security Administration, Office of Defense Nuclear Nonproliferation Research and Development (DNN R&D) through the Nuclear Science and Security Consortium under Award Numbers DE-NA0003180 and DE-NA0003996.

[1] B. Pritychenko et al., NIM A 640, 213 (2011).

Primary authors: YOUNES, Walid (Lawrence Berkeley National Lab); Dr GOLDBLUM, Bethany (LBL); Dr MANFREDI, Juan (AFIT); SHI, Laura (UC Berkeley); LI, Jonathan (UC Berkeley); CHIN, Char Juin (UC Berkeley)

Presenter: YOUNES, Walid (Lawrence Berkeley National Lab)

Session Classification: Poster Session

Track Classification: Poster Presentations