Contribution ID: **94** Type: **Oral**

Enhancing Fusion Research with TokSearch: Updates and Integration into the Fusion Data Platform

Wednesday, 17 July 2024 11:20 (20 minutes)

The rapidly evolving field of fusion research demands sophisticated data processing and management tools to handle the immense volumes of experimental data generated by tokamak operations. In response to these needs, TokSearch, a fusion data processing framework developed by General Atomics and recently open-sourced, has undergone significant upgrades to better serve the fusion research community and is currently being integrated with the Fusion Data Platform (FDP). The FDP, currently under development by a team led by General Atomics, is a comprehensive, open-access infrastructure designed to streamline the storage, processing, and sharing of fusion energy research data, facilitating collaboration and accelerating scientific discovery in the field. We present the latest advancements in TokSearch, focusing on its integration within the Fusion Data Platform project, and the substantial improvements in data processing capabilities it offers.

Key to these advancements is the introduction of a new, more versatile pipeline abstraction that supports arbitrary data processing tasks. This enhancement is underpinned by an updated application programming interface (API) that incorporates a flexible plugin framework, enabling seamless extension and customization of data processing functions. This new architecture facilitates the integration of diverse data sources and processing methods, significantly broadening the applicability of TokSearch in fusion research workflows.

Benchmarking results highlight the efficiency gains achieved through these updates, demonstrating orders of magnitude improvements in the retrieval and processing of DIII-D experimental data. These improvements are attributed to the optimized use of parallel file systems, which enable rapid access and manipulation of large datasets, along with integration of multiple distributed computing frameworks, providing flexible deployment options in high performance computing environments.

The integration of TokSearch into the FDP exemplifies how the project will provide powerful, scalable tools for data-driven research in fusion energy. Details of the FDP project will be provided, with a particular focus on how TokSearch's capabilities are leveraged to support large-scale data processing tasks.

Work supported by General Atomics' Internal Research and Development Funding and in part by the U.S. Department of Energy, Office of Science, Office of Fusion Energy Sciences, using the DIII-D National Fusion Facility, a DOE Office of Science user facility, under Award No. DE-FC02-04ER54698, along with Office of Fusion Energy Sciences Award No. DE-SC0024426.

Speaker's Affiliation

General Atomics, San Diego, CA, USA

Member State or IGO

United States of America

Primary author: SAMMULI, Brian (General Atomics)

Co-authors: Dr OLOFSSON, Erik (General Atomics); Mr OROZCO, David (General Atomics); Mr CLARK, C. Mitchell (General Atomics)

Presenter: SAMMULI, Brian (General Atomics)

Session Classification: Data Storage and Retrieval, Distribution and Visulaization

Track Classification: Data Storage and Retrieval, Distribution and Visulaization