FERC eLibrary Tool – New Techniques for Knowledge Extraction

FERC's eLibrary (https://elibrary.ferc.gov/eLibrary/search) is an on-line records information system that contains documents and information for the four industries FERC regulates: electric, hydropower, natural gas, and oil. The eLibrary contains millions of individual documents, including:

- electronic versions of documents issued by FERC from 1989-present;
- documents received and issued by FERC:
  - a description/index of documents from 1981-present;
  - microfilm and aperture cards of documents for 1981-1995;
  - scanned images of paper documents from 1995-present; and
  - native files electronically submitted from November 2000-present.

The eLibrary is a treasure trove of information for all four industries and is especially useful for the hydropower community because it provides historic and current information about environmental studies and license conditions required for individual hydropower projects. However, eLibrary users, including FERC staff, express frustration with the difficulty of finding and accessing relevant documents and information within eLibrary. Since the early 1990s, there has been a consistent call from users in the hydropower community for FERC to improve the process of finding and accessing documents and information in eLibrary. FERC updated its eLibrary interface in 2020, but the update did not address many of the problems users experience when searching for documents in the system.

With funding from DOE's Water Power Technologies Office, PNNL is developing a software solution to improve hydropower users' ability to locate and access documents and information contained within the existing on-line eLibrary. To begin developing the tool, the PNNL team has acquired a substantial collection of hydropower documents from the eLibrary system. The team explored various approaches, both manual and automated, to retrieve these documents. The team then leveraged various text analytic technologies to extract additional concepts and metadata from the documents that can be used to more effectively model and organize the information for search and retrieval within the existing eLibrary. Using the results of this analysis, the team has developed a pilot application based on informed design to demonstrate the improved visual analytic search and retrieval capabilities. The pilot offers a visual interface for users to interactively navigate and explore the information space to improve the user's experience in identifying and accessing relevant documents of interest. When completed, the publicly-available tool will make it easier for users to find and access documents and information within FERC's existing eLibrary system without altering the system itself.

For more on DOE's Water Power Technologies Office Hydropower Program, please see: https://www.energy.gov/eere/water/hydropower-program

For more information about the project, please contact Bo Saulsbury, james.saulsbury@pnnl.gov