

Development of a Nuclear Intelligent Retrieval and Analysis System

Byeongmun Ahn^{a*}, Byeong-hyeok Ha^a, Byoungchan Han^a, Tongkyu Park^a, Sung-Kyun Zee^a, Young-Jin Oh^b,
Jaeseok Yoo^b

^aFNC Technology Co., Ltd., 13 Heungdeok 1-ro, Giheung-gu, Yongin-si, Gyeonggi-do, 16954, Korea

^bPower Technology Institute, KEPCO Engineering & Construction Co. Inc., Republic of Korea

*Corresponding author: dinsslug@fnctech.com

1. Introduction

The design of a nuclear power plant should make extensive use of technical documents and human resources. Many efforts are being made to increase the effectiveness of nuclear power plant design work and cut costs and time. The use of Natural Language Processing (NLP) technologies for the configuration management of nuclear power plant design requirements and the search features is in line with the technical improvement efforts.

However, different NLP technologies necessitate software-based knowledge, and each technology must be learned separately before use. Most design workers are expected to have difficulty using this technology right away, and it does not serve the purpose of supporting plant design tasks.

In this study, we want to enhance the current systems [1] that can efficiently use and manage NLP technologies and models developed for the nuclear power plant design, to suit the environment of domestic plant designers. This study aims to make it as easy for designers to use NLP technologies in plant design as possible.

2. Application for the System

The modules implementing NLP technology [1] in this study were built as an integrated system called the Nuclear Intelligent Retrieval and Analysis System (NIRAS).

The system was divided into two modules: the manager module and the user module. The manager module allows you to process Lucene indexing training, BERT modeling, Topic modeling, and manage produced models. The user module assists plant designers by offering a feature akin to a search engine in web services.

The manager module operates on Windows OS whereas the user module runs on Linux OS, which is better in terms of design and performance and hence the reason for the separation of the two modules.

Fig. 1 depicts the manager module. The NLP module can be trained using the menu; it is designed as a Windows application with a user-friendly interface.

The user module is shown in Fig. 2. The layout is similar to that of the manager module, but it differs in that it provides a search function through a trained model. Because the module was developed as a web

service, it should be easier to transplant it to other web services in the future.

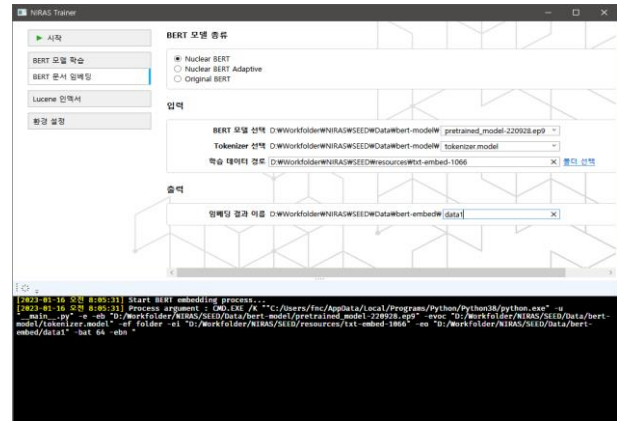


Fig. 1. The manager module of NIRAS.

3. Natural Language Processing Modules

NIRAS includes several NLP modules developed using Java or Python, and Fig. 2 shows the layout of NLP user modules.

3.1 Lucene Searcher

Lucene searcher is the search module developed using the Lucene engine [2], an information library (IR) that searches by keyword input similar to Google. It can be used to support plant design work and to compare performance with the BERT searcher.

The search is currently carried out using the Lucene engine's default settings. It is possible to have more improved search performance by applying additional extensions or plugins, such as Korean input.

3.2 N-BERT Searcher

Bidirectional Encoder Representations from Transformers (BERT) [3] is a Google artificial intelligence language model that performs well in understanding human language.

In a previous study, a new BERT training method and model were developed by combining the BERT training process with a database of nuclear technology related to document, the model was dubbed Nuclear BERT (N-BERT). For the production of N-BERT, NIRAS provides a user interface for plant designer to easily use the training process using learning data.

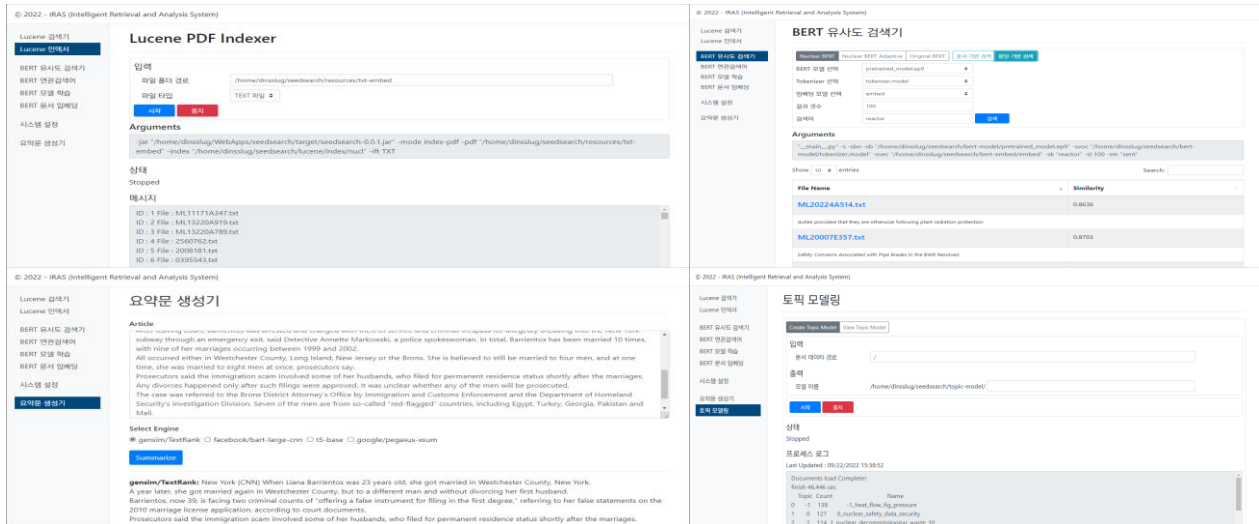


Fig. 2. The web service layout of four user modules for Lucene searcher, N-BERT searcher, Summary Sentence Generator, Topic Modeling.

3.3 Summary Sentence Generator

The summary sentence generator is the NLP module that summarizes long paragraphs in short sentences. NIRAS provides several summary sentence generators built with libraries such as TextRank, BART, T5, and PEGASUS.

3.4 Automatic Document Classification Module

The automatic document classification module is a technological implementation that clusters documents into specific subject features through an unsupervised learning classification method called Topic Modeling. Nuclear design documents can be classified using topic models generated through topic modeling. In this study, the module built with the BERTopic technique was integrated into NIRAS.

Fig. 3 shows the results of categorizing 1066 nuclear power plant design documents from the U.S. Nuclear Regulatory Commission’s website into six clusters.

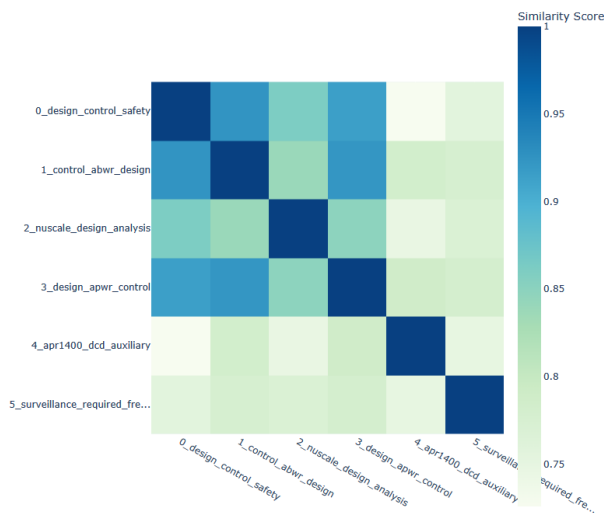


Fig. 3. Similarity matrix by topic group.

4. Conclusions

The approach developed in this study should make it easier for anyone involved in nuclear power plant design to apply NLP technology more effectively. By completing modeling processes more simply, which take a lot of time to train, it will be possible to assist in finding specific needs of nuclear specialized NLP technology in addition to search.

The system was designed to facilitate other NLP modules in the future, the time spent configuring the module usage environment should be greatly reduced.

Acknowledgement

This research was supported by KEPSCO Engineering & Construction Company, Inc. (KEPCO E&C) (No. TRS29).

REFERENCES

- [1] T. K. Part et al., Development of Intelligent Retrieval and Analysis System for NPP Design by Employing Natural Language Processing, Korean Society of Pressure Vessels and Piping, 2022.
- [2] A. Białecki, R. Muir, and G. Ingersoll, Apache Lucene 4, Proceedings of the SIGIR 2012 Workshop on Open Source Information Retrieval, 2012.
- [3] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805, 2018.