INTRODUCTION

This project aims to develop an NLP-supported user interface for ELKON to minimize the brute force put into the electrical circuit design process. Because of the uniqueness of each ship, ELKON employees need to spend a significant amount of time while reading the design documents of each ship. In this project, we aimed to write a raw text processing algorithm that can automatically read out these design documents and reduce the labor force needed to complete this task. Thus, the design process will be faster, stable, and costless for the client.

OBJECTIVES

- We have conducted research about the ship electrical design process.
- We have created questionnaires to be completed by the related ELKON departments to establish the realistic expectations and use cases.
- Machine Learning algorithms by using Spacy library.
- User Interface is created in line with the company expectations.
- Database has been formed using Django to store identified labels with the aid of the NER model as well as their corresponding context (i.e. sentence) and the document.

MOTIVATION

Each SFI describes different constraints of the ship to be built. For each SFI, a different kind of label domain is needed to be defined. That is why for each SFI, different machine learning models have been implemented. For example, for SFI 6, Machinery Main Components, labels determined as system, name, power, rpm, and quantity.

To capture the words for classification, regular expression libraries of Spacy and Python are utilized. After defining the rules for each label, the training data is processed by the Matcher tool of Spacy according to the patterns defined so that the span of each sentence is labelled with the corresponding character number capturing the word and its label. For training, the data is split into batches with the size of the parameter compounding(4.0, 32.0, 1.001). With 60 iterations the model is trained with different batches of the training data is split into batches with the size of the parameter compounding(4.0, 32.0, 1.001). With 60 iterations the model is trained with different batches of the training data with the dropout rate of 0.5 to avoid overfitting.

To prevent the loss of pre-trained model information, the training data is pickled and loaded to the backend application, and the input taken from the user to update the model is appended to the training data for further upgrades of the model.

PROJECT DETAILS II

The main purpose was automating the design phase as much as possible by eliminating the stage of reading the design document. So, having a user-friendly interface was crucial. While deciding on little details, future user's feedback was very guiding. In conclusion to ongoing research, it is decided that having extendable tables would be more useful rather than having more than one table at a time.

A framework called Nuxt used for frontend, a framework that offers material design for frontend which looks professional. The systems that we build works this way; first we upload the design document from the frontend, and it sends it to the backend. In Django, we take this document, parse it and read it with early mentioned methods.

Parts loaded in the NLP code and took the output and added it to the database. Then, added labels and values from Django are sent to Nuxt and printed in the table.

CONCLUSION

As a result of our project, we managed to implement a machine learning algorithm to extract useful information from a marine vessel design document.

- Particular machine learning models for each essential section, SFI, in the document given.
- Practical user interface and database to store and demonstrate the information extracted from design documents.

There were some limitations on the project, due to the nature of machine learning itself. We cannot guarantee that our initial machine learning models will work accurately for every type of design document. The more data we have makes the system more accurate. However, these design documents are highly confidential and difficult to access. This created a realistic limit on the expected accuracy of the project results.

Our project is a research and development project, so both we and the company ELKON are expecting some other students or researchers to follow up with improvements. When some group of researchers start developing the project from where we left; they would want to investigate options to increase comprehensiveness and accuracy of the machine learning models. One other improvement could be to enhance model accuracy by adding more train data to each model. We believe that this improvement could immensely improve the accuracy of our machine learning models.

At the end, we anticipate that this machine learning software we have implemented will be the one true source of information taken from design documents of marine vessels.

REFERENCES