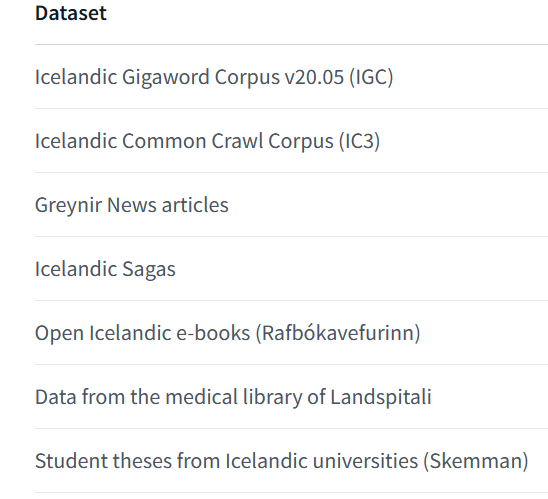
Original proposal:

1. Using an Icelandic OCR to scan handwritten documents in its original form

Problem faced: The texts in the documents are worn out and are illegible for OCRs to scan accurately.

1. Using Icelandic language models such as

Iceberta : This model was trained with fairseq using the RoBERTa-base architecture. It is one of many models we have trained for Icelandic,



Problem faced: We were successful in extracting the names from the text, but the issue was it did not tag the character relationship words, and since the Roberta model is huge, it is not feasible to retrain the model with the words we wanted to add.

Some of the other icelandic models we used were

GreynirServer

Ice-Bert

3) utf-8 not displaying icelandic characters in the correct format

4)Finally switching to English text and Using Spacy as a tokenizer because of difficulty processing icelandic text due to lack of ner(named entity recognition) models and lack of documentation.