**Research and Development Project**

**Project title:** Informational Tool for Alignment of Old Texts to Their Modern Representation

**Problem statement:**

The information within old books remain the most truthful exposures of the past, but they cannot always be easily read. Aligning old texts to their modern representation means translating them into a contemporary language by replacing the phrases that are out of use and archaisms with modern expressions.

Any person speaking a contemporary language should easily interpret the resulted text.

**Specifications:**

This project aims at creating an informational tool with all functionalities, which will serve at automation of the alignment of old text to the modern text.

Some formal definitions are provided bellow:

* old text – set of sentences written at least 150 years ago.
* modern text – set of sentences written up to 50 years ago.
* text structure – set consisting of word order, punctuation and vocabulary.

Aligning old texts in this respect, is the transformation of the old text structure into the modern text structure.

The first step in studying this issue is to establish differences between old text and modern one and this will lead to developing unsupervised (automatic) methods, which will find patterns of similarities and differences among texts.

The other steps include usage of neural machine translation technics and parallel corpora:

* to perform learning from the parallel corpus with the previously aligned texts
* taking into account the differences and similarities between them.

In the final step we will provide a model for machine translation between old and modern texts.

**The results:**

The proposed tool will resolve an interesting and challenging task from Natural Language Processing and some examples of its applications could be:

1. Translation of old documents and books into refreshed ones.
2. Study of old low resourced languages.
3. Research the development of the semantic and syntactic structure of the sentences.