**Introduction, Problem definition**

Summarization - is a complex and time–consuming process, whose complexity increases with the length of the text. Financial documents are large volumes of texts that require serious capacities for summarization. There is an idea: to take only more useful parts of the text – paragraphs, to reduce the time of summarization. It’ll be useful to create a simple and fast tool for selecting the best part of the text to summarize for reduce time and capacity.

**The purpose of the research work**

The research in the field of text summarization has mainly focused on investigations of different sentence extraction techniques, with a particular emphasis on news articles.

Financial documents have a number of features, including the dependence of the informative value of words on their position in the document. In this article, the PLM method was used to process the results of summaries of BERT excerpts. this post-processing of BERT results allows to take into account the location of words in relation to each other.

**Literature review**

Generalization systems should create a brief and cursory summary that conveys key information in the input data. These generalization tools identify the most important sentences in the input data, which can be either a separate document or a group of related documents, and combine them into summary information. [1]

Extractive summarisation is a widely researched area, and several techniques are worth mentioning. A survey by [survey of text summarization techniques categorizes the extractive summarisation system based on whether they are topic representation based or indicator representation based. A topic representation-based system can vary from as simple representation as tf-idf or word frequency, to topic signatures or latent semantic indexing. Other alternative text representations are term distributions or Latent semantic indexing. Some recent attempts, like, have used word embeddings-based representation. As opposed to these, indicator representation approaches do not rely on extracting or interpreting topics. They instead represent a document in a way that direct ranking of sentences becomes possible. [2]

Positional language model (PLM) - The key idea is to define a language model for each position of a document (thus the name positional language model), and score a document based on the scores of its PLMs. An important advantage of introducing a language model for each position is that it can allow us to model the “best-matching position” in a document with probabilistic models, thus supporting “soft” passage retrieval naturally. The PLM at a position of a document would be estimated based on the propagated word counts from the words at all other positions in the document. Specifically, we let each word at each position of a document to propagate the evidence of its occurrence to all other positions in the document so that positions close to the word would get more share of the evidence than those far away. This way, each position would receive propagated counts of words from all the words in the document with most propagated counts coming from words near the position. We can then estimate a language model for the position based on the propagated counts reaching the position. Overall, the PLM is shown to be able to achieve “soft” passage retrieval and capture proximity heuristic effectively in a unified probabilistic framework. [3]

**Description of the method**

To study this problem, an experimental method was chosen, which allowed us to test the effectiveness of the proposed approach on real data. To do this, a database of financial documents was collected, containing information about various companies and their financial indicators for a certain period of time.

the texts were analyzed from them it follows that most of the information used in summarization is in the first 10% of financial documents

**Data description**

Within the framework of this study, data from financial reports of companies were used, which contained information about UK annual reports. These data were collected from Financial Narrative Summarization (https://aclanthology.org/2021.fnp-1.22.pdf) and consisted of long texts with an average volume of about 6 thousand words.

To test the operation of the PLM model, we used the texts of reports and summary written manually by experts. Each text was processed by a model, and the results were compared with the results obtained by experts manually.

**Results of work and evaluation of reliability and measurement errors**

The PLM method has been tested only on financial statements and may have limitations when applied to other types of documents. For example, PLM may not work as well on documents containing a large amount of information about products and services, because PLM does not take into account the context and cannot take into account semantic connections between words. In addition, PLM cannot process documents without pre-processing.

**Comparison of this work with earlier works known in the literature.**

**Future work**

**Conclusions and possible applications of the results obtained**