

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ  
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**Факультет прикладной математики и информатики**

**Кафедра информационных систем управления**

Аннотация к дипломной работе  
**Разработка система машинного перевода на основе интерлингвы**

**Гаврилов Ярослав Антонович**

Научный руководитель - ст. преподаватель кафедры информационных систем  
управления Рубашко Н. К.

**Минск, 2024**

## РЕФЕРАТ

Дипломная работа, 41 стр., 6 иллюстр., 8 источников.

**Ключевые слова:** ИНТЕРЛИНГВА, МАШИННЫЙ ПЕРЕВОД, НЕЙРОННЫЕ СЕТИ, КОДИРОВЩИК, ДЕКОДИРОВЩИК.

**Объект исследования:** использование интерлингвы при разработке системы машинного перевода.

**Предмет исследования:** кодирование в контекстный вектор исходного естественного языка и декодирование из контекстного вектора в целевой естественный язык.

**Цель работы:** разработка модели машинного перевода, имеющую в основе контекстный вектор как нейронную интерлингву, и создание интерфейса этой модели.

**Методы исследования:** программирование на языке Python с фреймворком Pytorch, обучение моделей, оценка метриками BLEU и METEOR, наблюдения.

**Результат:** обучены несколько моделей с разными экспериментами в сферах токенизации, гиперпараметров. Архитектура модели и ее обучающий цикл был реализован на языке Python с фреймворком Pytorch. Разработан интерфейс, реализованный на языке Python с фреймворком PyQt6, демонстрирующий переводы модели с исходного языка на целевой. Приведены выводы по результатам экспериментов и дана характеристика модели получившей наивысшее количество баллов в системе оценок BLEU и METEOR.

**Области применения:** архитектура модели с нейронной интерлингвой может быть применена в области машинного перевода в качестве основы для программного переводчика для коммерческих или других целей. В общем случае система кодировщик-декодировщик используется для широкого спектра задач обработки естественных языков, включая языковой перевод, субтитры к изображениям, разговорные модели и обобщение текста.

# ABSTRACT

Graduate work, 41 pages, 6 pictures, 8 sources.

**Keywords:** INTERLINGUA, MACHINE TRANSLATION, NEURAL NETWORKS, ENCODER, DECODER.

**Object of research:** the use of Interlingua in the development of a machine translation system.

**Subject of research:** encoding into a context vector of the source natural language and decoding from the context vector into the target natural language.

**Objective:** development of a machine translation model based on a context vector as a neural interlingua, and creation of an interface for this model.

**Research methods:** programming in Python with the Pytorch framework, training models, observations, evaluation with BLEU and METEOR metrics.

**Result:** several models were trained with different experiments in the areas of tokenization, hyperparameters and samples. The architecture of the model and its training cycle were implemented in Python with the Pytorch framework. An interface was developed, implemented in Python with the PyQt6 framework, demonstrating translations of the model from the source language to the target language. Conclusions based on the results of the experiments are presented and characteristics of the model that received the highest score in the BLEU and METEOR rating system are given.

**Scope:** the architecture of the neural interlingual model can be applied in the field of machine translation as the basis for a software translator for commercial or other purposes. In general, the Encoder-Decoder system is used for a wide range of natural language processing tasks, including language translation, image captioning, conversational models, and text summarization.