Министерство образования Республики Беларусь

Учреждение образования

«Брестский государственный технический университет»

Кафедра ИИТ

Лабораторная работа № 4

По дисциплине ЕЯИИС

Тема: «Текстовые корпусы и методы идентификации языка текста»

Выполнил:

Студент 4 курса

Группы ИИ-16

Пилипук З.А.

Проверил:

Слинко Е.В.

Брест, 2021

**Цель работы:** закрепить знания в решении задач построения и работы с корпусами текстов.

**Код программы:**

#include <iostream>

#include <fstream>

#include <vector>

#include <string>

#include <algorithm>

struct LanguageHandler

{

void setupDictionary()

{

getData("data\_src/rus.txt", rus);

getData("data\_src/eng.txt", eng);

}

std::string getLanguage(const std::string &fname)

{

std::ifstream file(fname);

size\_t r = 0, e = 0;

while (!file.eof())

{

std::string temp;

file >> temp;

for (const auto &c : temp)

{

if (rus.find(c) != std::string::npos)

{

++r;

}

else if (eng.find(c) != std::string::npos)

{

++e;

}

}

}

file.close();

return e > r ? "English" : "Russian";

}

void getData(const std::string &fname, std::string &data)

{

std::ifstream file(fname);

std::getline(file, data);

file.close();

}

std::string rus;

std::string eng;

};

int main()

{

LanguageHandler lg;

lg.setupDictionary();

std::cout << lg.getLanguage("text.txt") << std::endl;

return 0;

}

**Тестирование:**

Файловая система содержит файлы:





**Текстовый файл:** ПРИВЕТ КАК ДЕЛА ЧТО ДЕЛАЕШЬ?

**Вывод:** закрепил знания в решении задач построения и работы с корпусами текстов.