

МАГИСТЕРСКАЯ ПРОГРАММА 09.04.01/07 Интеллектуальные системы анализа,
обработки и интерпретации больших данных

Вариант 13

Дисциплина: Языки программирования для работы с большими данными

Преподаватель	<hr/>	П.В. Степанов
	(Подпись, дата)	(И.О. Фамилия)

Москва, 2022

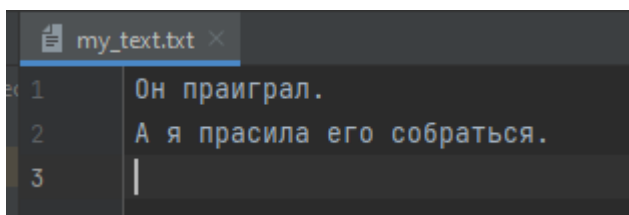
Цель лабораторной работы: получение первичных навыков работы со строками и регулярными выражениями в Java.

Ход работы:

Задание №1:

1. В тексте после буквы Р, если она не последняя в слове, ошибочно напечатана буква А вместо О. Внести исправления в текст.
2. В тексте слова заданной длины заменить указанной подстрокой, длина которой может не совпадать с длиной слова.

Входной файл:



```
1 Он праиграл.
2 А я прасила его собраться.
3 |
```

Листинг выполнения подзадачи 1:

```
package laba_7.task_1;
import java.io.File;
import java.io.FileNotFoundException;
import java.io.IOException;
import java.nio.charset.StandardCharsets;
import java.nio.file.Files;
import java.nio.file.Path;
import java.nio.file.Paths;
import java.nio.file.StandardOpenOption;
import java.util.Collections;
import java.util.Scanner;

//3. В тексте после буквы Р, если она не последняя в слове, ошибочно напечатана
буква А вместо О. Внести исправления в текст.
public class Textst {
    public static void main(String[] args) {

        File file_1 = new
File("C:\\Users\\metel\\IdeaProjects\\my_lab_7\\src\\laba_7\\task_1\\my_text.txt");
        Path file_2_path =
Paths.get("C:\\Users\\metel\\IdeaProjects\\my_lab_7\\src\\laba_7\\task_1\\my_text_cha
nged.txt");
        File file_2 = new
File("C:\\Users\\metel\\IdeaProjects\\my_lab_7\\src\\laba_7\\task_1\\my_text_changed.
txt.txt");

        if(file_2.delete()) {
            try {
                file_2.createNewFile();
            } catch (IOException e) {
                e.printStackTrace();
            }
        }
    }
}
```

```

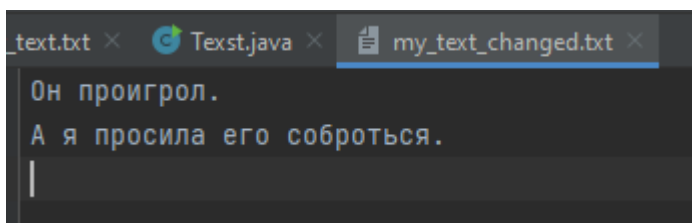
Scanner scanner = null;
try {
    scanner = new Scanner(file_1);
} catch (
    FileNotFoundException e) {
    e.printStackTrace();
}

while(scanner.hasNextLine()) {
    String line = scanner.nextLine();
    line = line.replaceAll("PA", "PO");
    line = line.replaceAll("pa", "po");
    line = line.replaceAll("Pa", "Po");
    line = line.replaceAll("pA", "pO");

    try {
        Files.write(file_2_path, Collections.singleton(line),
StandardCharsets.UTF_8, StandardOpenOption.APPEND);
    } catch (IOException e) {
        e.printStackTrace();
    }
}
}
}

```

Результат выполнения подзадачи 1:



Листинг выполнения подзадачи 2:

```

package laba_7.task_1;
//4. В тексте слова заданной длины заменить указанной подстрокой, длина которой
может не совпадать с длиной слова.

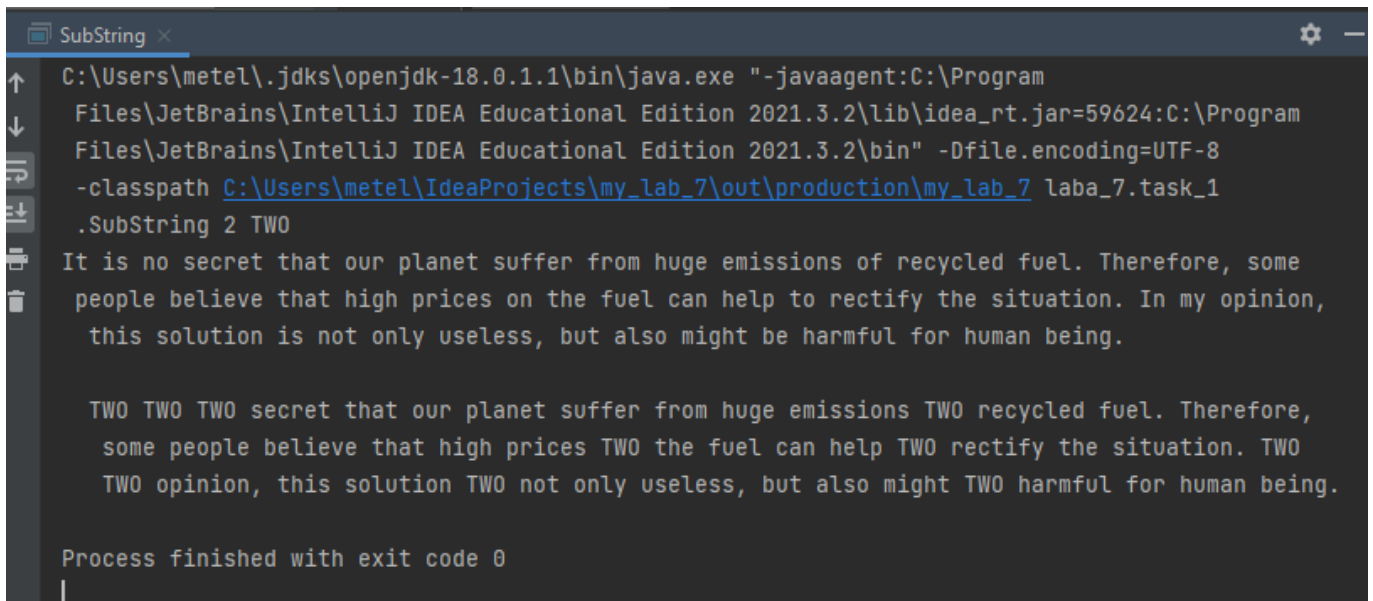
public class SubString {
    public static void main(String[] args) throws Exception{
        String text = "It is no secret that our planet suffer from huge emissions of
recycled fuel. Therefore, some people believe that high prices on the fuel can help
to rectify the situation. In my opinion, this solution is not only useless, but also
might be harmful for human being.";
        String textNew = " ";

        int kol = Integer.parseInt(args[0]);
        String substring = args[1];

        String [] all_words = text.split("\\s");
        for (int k=0; k<all_words.length; ++k){
            if (all_words[k].length() == kol){
                textNew += " " + substring;
            }else {
                textNew += " " + all_words[k];
            }
        }
        System.out.println(text + "\n");
        System.out.println(textNew);
    }
}

```

Результат выполнения подзадачи 2:



```
SubString x
C:\Users\metel\.jdk\openjdk-18.0.1.1\bin\java.exe "-javaagent:C:\Program
Files\JetBrains\IntelliJ IDEA Educational Edition 2021.3.2\lib\idea_rt.jar=59624:C:\Program
Files\JetBrains\IntelliJ IDEA Educational Edition 2021.3.2\bin" -Dfile.encoding=UTF-8
-classpath C:\Users\metel\IdeaProjects\my_lab_7\out\production\my_lab_7 laba_7.task_1
.SubString 2 TWO
It is no secret that our planet suffer from huge emissions of recycled fuel. Therefore, some
people believe that high prices on the fuel can help to rectify the situation. In my opinion,
this solution is not only useless, but also might be harmful for human being.

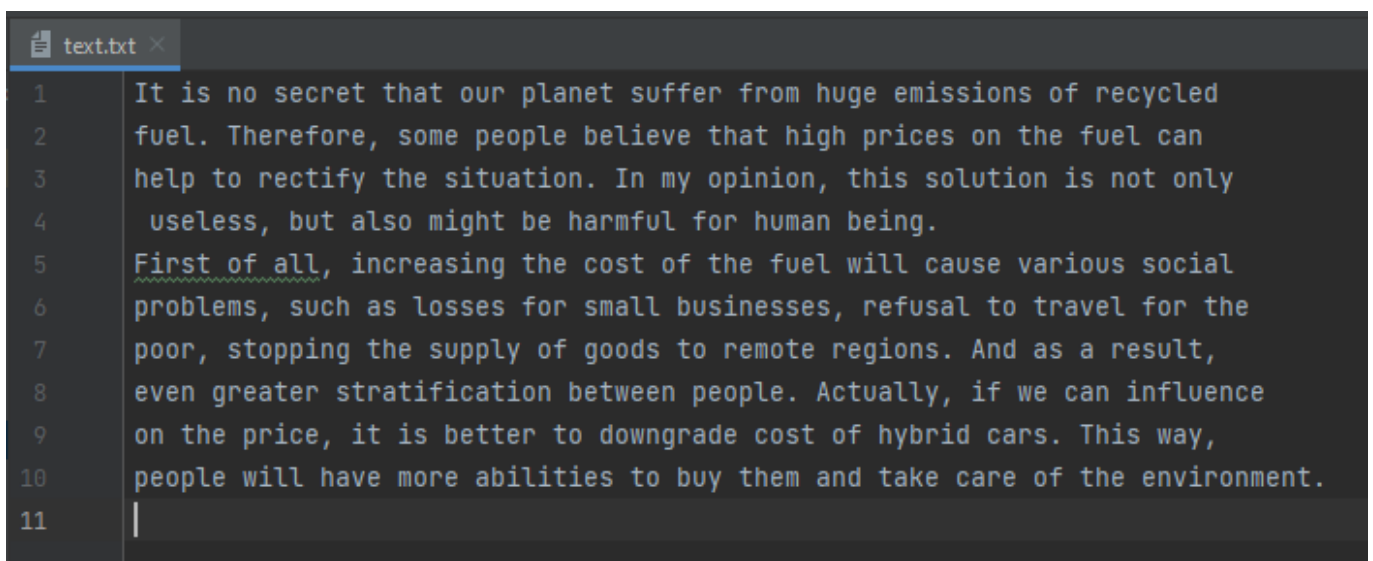
TWO TWO TWO secret that our planet suffer from huge emissions TWO recycled fuel. Therefore,
some people believe that high prices TWO the fuel can help TWO rectify the situation. TWO
TWO opinion, this solution TWO not only useless, but also might TWO harmful for human being.

Process finished with exit code 0
```

Задание №2:

1. В тексте найти и напечатать n символов (и их количество), встречающихся наиболее часто.
2. Найти, каких букв, гласных или согласных, больше в каждом предложении текста.

Входной файл подзадачи 1:



```
text.txt x
1 It is no secret that our planet suffer from huge emissions of recycled
2 fuel. Therefore, some people believe that high prices on the fuel can
3 help to rectify the situation. In my opinion, this solution is not only
4 useless, but also might be harmful for human being.
5 First of all, increasing the cost of the fuel will cause various social
6 problems, such as losses for small businesses, refusal to travel for the
7 poor, stopping the supply of goods to remote regions. And as a result,
8 even greater stratification between people. Actually, if we can influence
9 on the price, it is better to downgrade cost of hybrid cars. This way,
10 people will have more abilities to buy them and take care of the environment.
11 |
```

Листинг выполнения подзадачи 1:

```
package laba_7.task_2;
//В тексте найти и напечатать n символов (и их количество), встречающихся наиболее
часто.
import java.io.File;
import java.io.FileNotFoundException;
import java.util.*;

public class Symbols {
    public static void main(String[] args) {
        File inp_file = new
File("C:\\Users\\metel\\IdeaProjects\\my_lab_7\\src\\laba_7\\task_2\\text.txt");

        Scanner scanner = null;
        try {
            scanner = new Scanner(inp_file);
        } catch (FileNotFoundException e) {
            e.printStackTrace();
        }

        HashMap<Character, Integer> dictionary = new HashMap<>();

        while(scanner.hasNextLine()) {
            String line = scanner.nextLine();
            line = line.replace(" ", " ");
            line = line.replace("\n", "");
            line = line.toLowerCase(Locale.ROOT);
            char[] chars = line.toCharArray();

            for (char ch: chars) {
                if('a' <= ch && 'я' >= ch) {
                    if (dictionary.containsKey(ch)) {
                        dictionary.put(ch, dictionary.get(ch) + 1);
                    } else {
                        dictionary.put(ch, 1);
                    }
                }
            }
        }

        Comparator<Map.Entry<Character, Integer>> valueComparator = new
Comparator<Map.Entry<Character,Integer>>() {
            @Override
            public int compare(Map.Entry<Character, Integer> e1, Map.Entry<Character,
Integer> e2) {
                int v1 = e1.getValue();
                int v2 = e2.getValue();
                return (v2 - v1);
            }
        };

        Set<Map.Entry<Character, Integer>> entries = dictionary.entrySet();
        List<Map.Entry<Character, Integer>> listOfEntries = new
ArrayList<Map.Entry<Character, Integer>>(entries);
        Collections.sort(listOfEntries, valueComparator);

        System.out.println("Input number of letters: ");
        Scanner in = new Scanner(System.in);
        int n = in.nextInt();

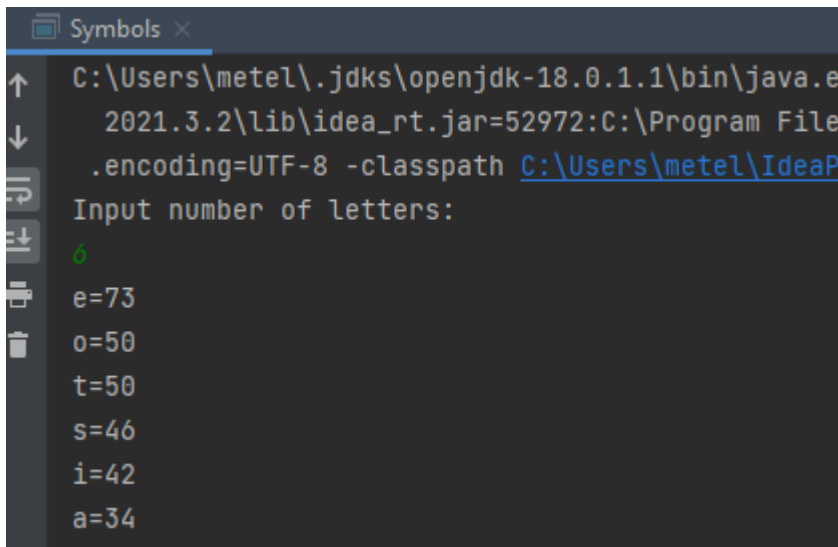
        for (int i = 0; i < n; i++) {
            System.out.println(listOfEntries.get(i));
        }
    }
}
```

```

    }
}
}

```

Результат выполнения подзадачи 1:



```

C:\Users\metel\.jdk\openjdk-18.0.1.1\bin\java.exe
2021.3.2\lib\idea_rt.jar=52972:C:\Program File
.encoding=UTF-8 -classpath C:\Users\metel\IdeaP
Input number of letters:
6
e=73
o=50
t=50
s=46
i=42
a=34

```

Листинг выполнения подзадачи 2:

```

package laba_7.task_2;
import java.util.regex.Matcher;
import java.util.regex.Pattern;
//4. Найти, каких букв, гласных или согласных, больше в каждом предложении текста.
public class Glasn_sogl {
    public static void main(String[] args) {
        String text = "It is no secret that our planet suffer from huge emissions of
recycled fuel. Therefore, some people believe that high prices on the fuel can help
to rectify the situation. In my opinion, this solution is not only useless, but also
might be harmful for human being.";

        Pattern vowel = Pattern.compile("[aeiouy]");
        Pattern consonant = Pattern.compile("ab cd fgh jkl mnp qr st vwxz");

        for (String sent : text.split("[.!?]")) {
            int vowelCount = 0;
            int consonantCount = 0;

            Matcher vMatcher = vowel.matcher(sent);
            Matcher cMatcher = consonant.matcher(sent);

            while (vMatcher.find()) {
                vowelCount++;
            }

            while (cMatcher.find()) {
                consonantCount++;
            }

            System.out.print("In sentance '" + sent + "' ");
            if (vowelCount > consonantCount) System.out.println("Glasnue");
            else if (vowelCount < consonantCount) System.out.println("Soglasnue ");
            else System.out.println("Equel glasn and soglasn");
        }
    }
}

```

```
}  
}
```

Результат выполнения подзадачи 2:

```
Glasn_sogl x  
C:\Users\metel\.jdk\openjdk-18.0.1.1\bin\java.exe "-javaagent:C:\Program Files\JetBrains\IntelliJ IDEA Educational Edition  
2021.3.2\lib\idea_rt.jar=53034:C:\Program Files\JetBrains\IntelliJ IDEA Educational Edition 2021.3.2\bin" -Dfile  
.encoding=UTF-8 -classpath C:\Users\metel\IdeaProjects\my_lab_7\out\production\my_lab_7 laba_7.task_2.Glasn_sogl  
In sentence 'It is no secret that our planet suffer from huge emissions of recycled fuel' Glasnue  
In sentence ' Therefore, some people believe that high prices on the fuel can help to rectify the situation' Glasnue  
In sentence ' In my opinion, this solution is not only useless, but also might be harmful for human being' Glasnue  
  
Process finished with exit code 0
```

Задание №3:

1. Найти такое слово в первом предложении, которого нет ни в одном из остальных предложений.
2. Во всех вопросительных предложениях текста найти и напечатать без повторений слова заданной длины.

Листинг выполнения подзадачи 1:

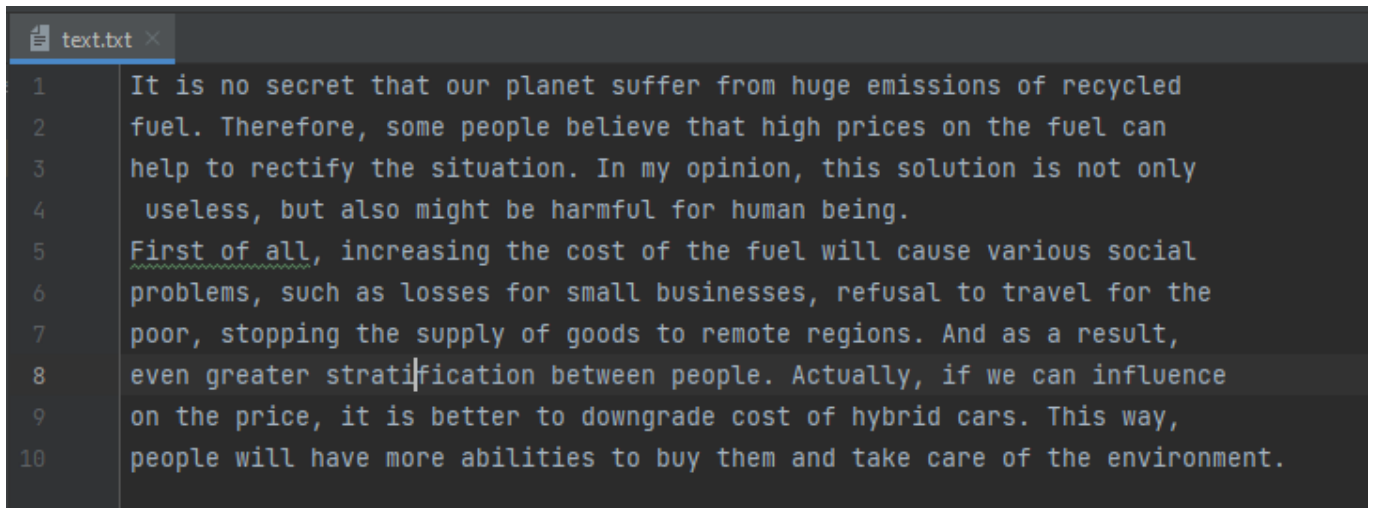
```
package laba_7.task_3;  
import java.io.File;  
import java.io.FileNotFoundException;  
import java.util.*;  
//3. Найти такое слово в первом предложении, которого нет ни в одном из остальных  
предложений.  
public class Popular_word {  
    public static void main(String[] args) {  
        File inp_file = new  
File("C:\\Users\\metel\\IdeaProjects\\my_lab_7\\src\\laba_7\\task_3\\text.txt");  
  
        Scanner scanner = null;  
        try {  
            scanner = new Scanner(inp_file);  
        } catch (FileNotFoundException e) {  
            e.printStackTrace();  
        }  
  
        String text = scanner.nextLine();  
        text = text.toLowerCase(Locale.ROOT);  
        String[] sentences = text.split("[.?!]");  
        String[] words = sentences[0].split("[ <>—:]");  
  
        boolean foundFlag;  
  
        for (String word : words) {  
            foundFlag = true;  
            for (int i = 1; i < sentences.length; i++) {  
                if (sentences[i].contains(word)) {  
                    foundFlag = false;  
                }  
            }  
        }  
    }  
}
```

```

    }
    if(foundFlag){
        System.out.print("Word in first sentence: ");
        System.out.println(word);
        break;
    }
}
}
}

```

Входной текст:

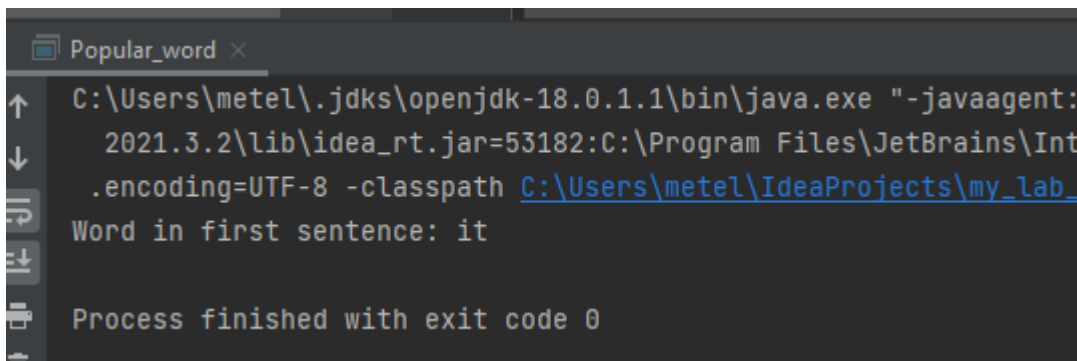


```

1  It is no secret that our planet suffer from huge emissions of recycled
2  fuel. Therefore, some people believe that high prices on the fuel can
3  help to rectify the situation. In my opinion, this solution is not only
4  useless, but also might be harmful for human being.
5  First of all, increasing the cost of the fuel will cause various social
6  problems, such as losses for small businesses, refusal to travel for the
7  poor, stopping the supply of goods to remote regions. And as a result,
8  even greater stratification between people. Actually, if we can influence
9  on the price, it is better to downgrade cost of hybrid cars. This way,
10 people will have more abilities to buy them and take care of the environment.

```

Результат выполнения подзадачи 1:



```

C:\Users\metel\.jdk\openjdk-18.0.1.1\bin\java.exe "-javaagent:
2021.3.2\lib\idea_rt.jar=53182:C:\Program Files\JetBrains\Int
.encoding=UTF-8 -classpath C:\Users\metel\IdeaProjects\my_lab_
Word in first sentence: it
Process finished with exit code 0

```

Листинг выполнения подзадачи 2:

```

package laba_7.task_3;
import java.util.ArrayList;
import java.util.regex.Matcher;
import java.util.regex.Pattern;
//4. Во всех вопросительных предложениях текста найти и напечатать без повторов
слова заданной длины.

public class Question_sen {
    public static void main(String[] args) {
        String text = "It is no secret that our planet suffer from huge emissions of
recycled\n" +
            "fuel. Therefore, some people believe that high prices on the fuel
can\n" +
            "help to rectify the situation. In my opinion, this solution is not
only\n" +

```



```

        " useless, but also might be harmful for human being being?\n" +
        "First of all, increasing the cost of the fuel will cause various
social\n" +
        "problems, such as losses for small businesses, refusal to travel for
the\n" +
        "poor, stopping the supply of goods to remote regions. And as a
result,\n" +
        "even greater stratification between people. Actually, if we can
influence\n" +
        "on the price, it is better to downgrade cost of hybrid cars. This
way,\n" +
        "people will have more abilities to buy them and take care of the
environment.";
    String patternStr = "[^.?!*][?]*";

    Pattern p = Pattern.compile(patternStr);
    Matcher m = p.matcher(text);

    ArrayList<String> words = new ArrayList<String>();

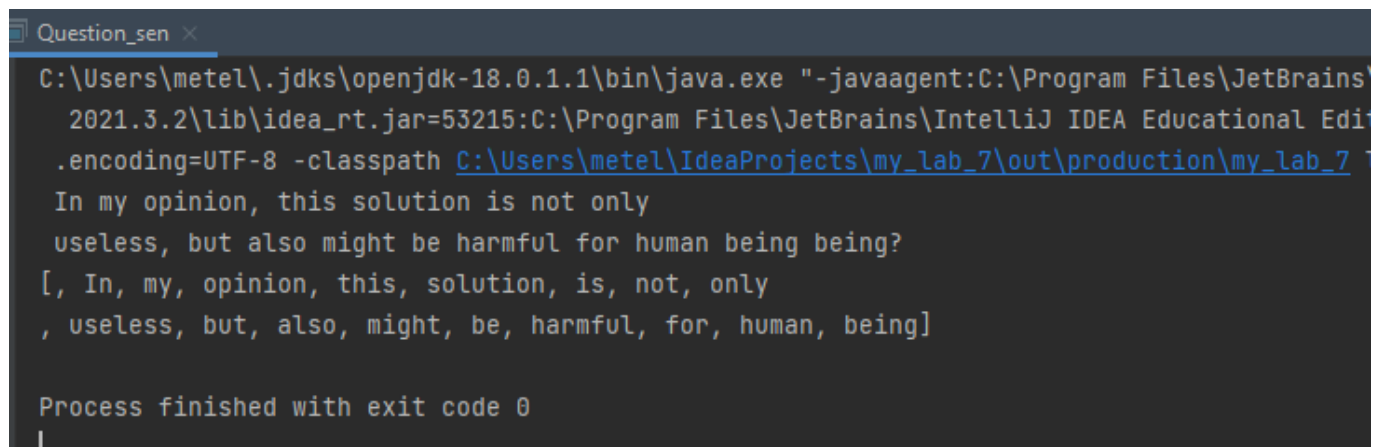
    while (m.find()) {
        String sent = m.group();
        System.out.println(sent);

        for (String word : sent.split(" ")) {
            word = word.replaceAll("[.?,]", "");
            if (!words.contains(word)) words.add(word);
        }
    }

    System.out.println(words);
}

```

Результат выполнения подзадачи 2:



```

C:\Users\metel\.jdk\openjdk-18.0.1.1\bin\java.exe "-javaagent:C:\Program Files\JetBrains\
2021.3.2\lib\idea_rt.jar=53215:C:\Program Files\JetBrains\IntelliJ IDEA Educational Edi
.encoding=UTF-8 -classpath C:\Users\metel\IdeaProjects\my_lab_7\out\production\my_lab_7
In my opinion, this solution is not only
useless, but also might be harmful for human being being?
[, In, my, opinion, this, solution, is, not, only
, useless, but, also, might, be, harmful, for, human, being]

Process finished with exit code 0

```

Задание №4:

1. Все слова текста рассортировать в порядке убывания их длин, при этом все слова одинаковой длины рассортировать в порядке возрастания в них количества гласных букв.
2. В тексте исключить подстроку максимальной длины, начинающуюся и заканчивающуюся заданными символами.

Листинг выполнения подзадачи 1:

```
package laba_7.task_4;
import java.io.File;
import java.io.FileNotFoundException;
import java.util.*;

//3. Все слова текста рассортировать в порядке убывания их длин, при этом
// все слова одинаковой длины рассортировать в порядке возрастания в них количества
// гласных букв.
public class Vov_Sort {
    public static void main(String[] args) {

        File inp_file = new
File("C:\\Users\\metel\\IdeaProjects\\my_lab_7\\src\\laba_7\\task_4\\text.txt");

        Scanner scanner = null;
        try {
            scanner = new Scanner(inp_file);
        } catch (FileNotFoundException e) {
            e.printStackTrace();
        }

        String glasniye = "УуЕеЫыАаОоЭэЁёЯяИиЮю";

        String text = scanner.nextLine();
        text = text.toLowerCase(Locale.ROOT);
        text = text.replace(" ", " ");
        ArrayList<String> words = new ArrayList<>(Arrays.asList(text.split("[.?!,<«»-
:;]"))));

        Comparator<String> valueComparator = new Comparator<String>() {
            @Override
            public int compare(String e1, String e2) {
                if (e1.length() > e2.length()) {
                    return -1;
                } else if (e1.length() < e2.length()) {
                    return 1;
                } else {
                    int count1 = 0;
                    int count2 = 0;
                    for (int i = 0; i < e1.length(); i++) {
                        if (glasniye.contains(e1.substring(i, i + 1))) {
                            count1++;
                        }
                    }
                    for (int i = 0; i < e2.length(); i++) {
                        if (glasniye.contains(e2.substring(i, i + 1))) {
                            count2++;
                        }
                    }
                }
            }
        };
    }
}
```

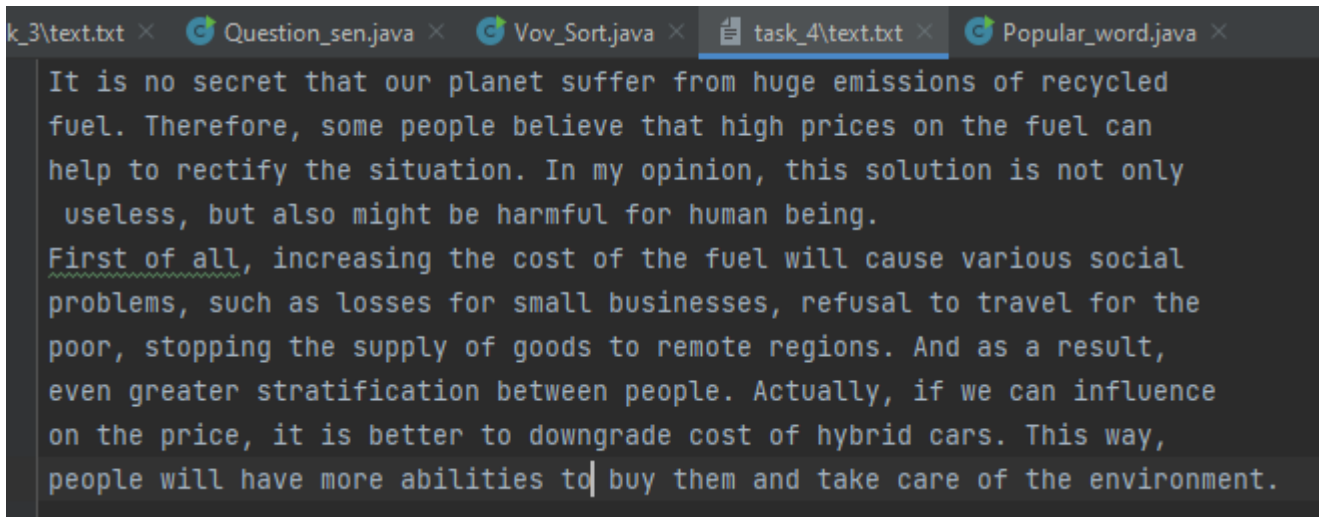
```

    }
    }
    return (count1 - count2);
}
};

words.removeAll(Collections.singleton(""));
Collections.sort(words, valueComparator);
System.out.println(words);
}
}

```

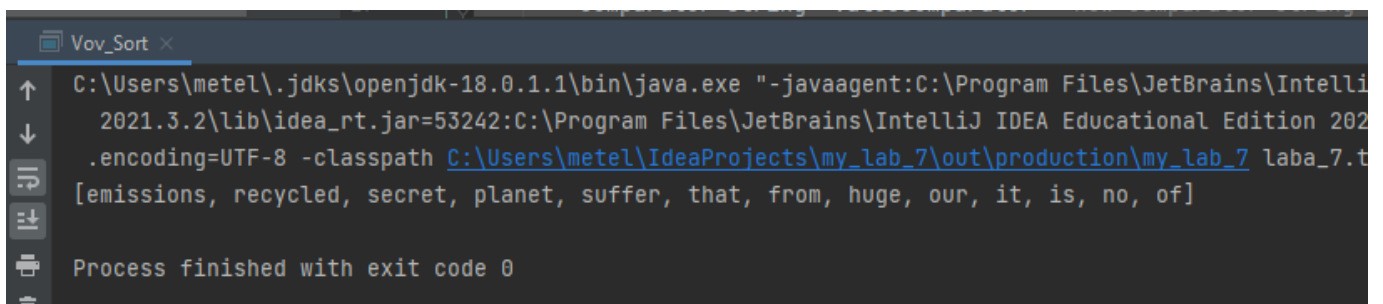
Входной текст:



It is no secret that our planet suffer from huge emissions of recycled fuel. Therefore, some people believe that high prices on the fuel can help to rectify the situation. In my opinion, this solution is not only useless, but also might be harmful for human being.

First of all, increasing the cost of the fuel will cause various social problems, such as losses for small businesses, refusal to travel for the poor, stopping the supply of goods to remote regions. And as a result, even greater stratification between people. Actually, if we can influence on the price, it is better to downgrade cost of hybrid cars. This way, people will have more abilities to buy them and take care of the environment.

Результат выполнения подзадачи 1:



```

C:\Users\metel\.jdk\openjdk-18.0.1.1\bin\java.exe "-javaagent:C:\Program Files\JetBrains\IntelliJ IDEA Educational Edition 2021.3.2\lib\idea_rt.jar=53242:C:\Program Files\JetBrains\IntelliJ IDEA Educational Edition 2021.3.2\bin" -classpath C:\Users\metel\IdeaProjects\my_lab_7\out\production\my_lab_7 laba_7.task_4.Minus_SubString
[emissions, recycled, secret, planet, suffer, that, from, huge, our, it, is, no, of]
Process finished with exit code 0

```

Листинг выполнения подзадачи 2:

```

package laba_7.task_4;
import java.util.regex.Matcher;
import java.util.regex.Pattern;
//4. В тексте исключить подстроку максимальной длины, начинающуюся и
заканчивающуюся заданными символами.

public class Minus_SubString {
    public static void main(String[] args) {
        String text = "aaaaazzzzzzaaazzaammaaaazzmzzzzmzaaaaa";

        Matcher matcher = Pattern.compile(args[0] + ".*" + args[1]).matcher(text);
        matcher.find();

        System.out.println("Text before: " + text);
    }
}

```

```
text = text.replace(matcher.group(), "");  
System.out.println("Text after: " + text);  
}  
}
```

Результат выполнения подзадачи 2:

```
C:\Users\metel\.jdk\openjdk-18.0.1.1\bin\java.exe "-javaagent:C:\Program  
2021.3.2\lib\idea_rt.jar=53256:C:\Program Files\JetBrains\IntelliJ ID  
.encoding=UTF-8 -classpath C:\Users\metel\IdeaProjects\my_lab_7\out\pr  
Text before: aaaaazzzzzzaaazzaammaaaazmzzzzmzaaaaa  
Text after: aaaaa  
  
Process finished with exit code 0
```

Программное решение представлено в репозитории распределённой системы управления версиями Git:

https://github.com/matvilen/BigDataLanguages/tree/main/my_lab_7/src/laba_7

Вывод: при выполнении лабораторной работы были получены навыки работы со строками и регулярными выражениями в Java.