Лабораторная работа №3 по курсу криптографии

Выполнил студент группы М8О-308Б-18 Коростелев Дмитрий

Условие

Сравнить

1) Два осмысленных текста на естественном языке
2) Осмысленный текст и текст из случайных букв
3) Осмысленный текст и текст из случайных слов
4) Два текста из случайных букв
5) Два текста из случайных слов

Метод решения

В качестве осмысленных текстов будем использовать произведения «Мастер и Маргарита» и «Война и мир».

Тексты из случайных слов сгенерируем при помощи сторонних сервисов

Тексты из случайных букв будем генерировать внутри программы

Стоит отметить, что длина заранее сгенерированных текстов больше 90000 символов, этого вполне достаточно, чтобы получить точную оценку сравнения текстов.

Сравнения будем производить посимвольно и считать количество совпавших символов, ответ будем получать делением кол-ва совпавших символов на сравниваемую длину строк.

Чтобы получить качественные результаты, нужно провести несколько серий тестов с разным объемом данных и на основе этих результатов зафиксировать несколько мало отличающихся вероятностей и проанализировать их.

Результат работы программы

Программа проводит побуквенное сравнение текстов, при этом при увеличении объема теста, увеличивается сравниваемое окно текста, увеличивать объем теста имеет смысл только до максимальной длины одного из текстов (в квадратных скобках отображается количество сравниваемых символов):

```
import java.io.*;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Random;
```

```
public class TextAnalyze {

    public static String genRandomText(int len, ArrayList<Character>
Alphabet){
        StringBuilder builder = new StringBuilder();
        Random random = new Random();
        for(int i = 0;i<len;++i){
            builder.append(Alphabet.get(Math.abs(random.nextInt()) %
Alphabet.size()));
        }
        return builder.toString();
    }
    public static String buildStringFromFile(String path) throws IOException
{
        File file = new File(path);
        InputStream inputStream;
        inputStream = new FileInputStream(file);
        return Arrays.toString(inputStream.readAllBytes());
    }
    public static double compareStrings(String s1, String s2, int l){
        l = len(s1,s2,l);
        int sameCount = 0;
        for(int i = 0;i<l;++i){
            if(s1.charAt(i) == s2.charAt(i)){
                ++sameCount;
            }
        }
        return (double) sameCount / (double) l;
    }
    public static ArrayList<Character> genEngAlphabet(){
        ArrayList<Character> alp = new ArrayList<>();
        for(char s = 'a';s <='z'; ++s){
            alp.add(s);
        }
        for(char s = 'A';s <='Z'; ++s){
            alp.add(s);
        }
        alp.add(' ');
        alp.add('\n');
        return alp;
    }
    public static int len(String s1, String s2, int l){
        return Math.min(s1.length(), Math.min(s2.length(), l));
    }

    public static void test(int stringLen) throws IOException {
        String pathFirstNormalText =
"D:\\dev\\Java\\Crypto\\src\\WarAndPeace.txt";
        String pathSecondNormalText =
"D:\\dev\\Java\\Crypto\\src\\MasterAndMargarita.txt";
        String pathFirstRandomWordText =
"D:\\dev\\Java\\Crypto\\src\\firstRandomWordText.txt";
        String pathSecondRandomWordText =
"D:\\dev\\Java\\Crypto\\src\\secondRandomWordText.txt";
        ArrayList<Character> alp = genEngAlphabet();
        String firstRandomString = genRandomText(stringLen, alp);
        String secondRandomString = genRandomText(stringLen, alp);
        String firstNormalString = buildStringFromFile(pathFirstNormalText);
        String secondNormalString =
buildStringFromFile(pathSecondNormalText);
        String firstRandomWordString =
buildStringFromFile(pathFirstRandomWordText);
        String secondRandomWordString =
```

```
buildStringFromFile(pathSecondRandomWordText);
        System.out.println("Длина текстов: " + stringLen);
        System.out.print("Два осмысленных текста: ");
        System.out.printf("[%d] ", len(firstNormalString,
secondNormalString, stringLen));
        System.out.println(compareStrings(firstNormalString,
secondNormalString,
                len(firstNormalString, secondNormalString, stringLen)));

        System.out.print("Осмысленный текст и текст из случайных букв: ");
        System.out.printf("[%d] ", len(firstNormalString, firstRandomString,
stringLen));
        System.out.println(compareStrings(firstNormalString,
firstRandomString,
                len(firstNormalString, firstRandomString, stringLen)));

        System.out.print("Осмысленный текст и текст из случайных слов: ");
        System.out.printf("[%d] ",len(firstNormalString,
firstRandomWordString, stringLen));
        System.out.println(compareStrings(firstNormalString,
firstRandomWordString,
                len(firstNormalString, firstRandomWordString, stringLen)));

        System.out.print("Тексты из случайных букв: ");
        System.out.printf("[%d] ",len(firstRandomString, secondRandomString,
stringLen));
        System.out.println(compareStrings(firstRandomString,
secondRandomString,
                len(firstRandomString, secondRandomString, stringLen)));

        System.out.print("Тексты из случайных слов: ");
        System.out.printf("[%d] ",len(firstRandomWordString,
secondRandomWordString, stringLen));
        System.out.println(compareStrings(firstRandomWordString,
secondRandomWordString,
                len(firstRandomWordString, secondRandomWordString,
stringLen)));
        System.out.println();
    }
    public static void main(String[] args) throws IOException {
        test(100);
        test(1000);
        test(10000);
        test(50000);
        test(80000);
    }
}
```

Как видно и кода программы, проводится 5 тестирований со строками длиной 100, 1000, 10000, 50000, 80000.

Приведем результаты выполнения данной программы:

```
Длина текстов: 100
Два осмысленных текста: [100] 0.08
Осмысленный текст и текст из случайных букв: [100] 0.0
Осмысленный текст и текст из случайных слов: [100] 0.09
Тексты из случайных букв: [100] 0.02
Тексты из случайных слов: [100] 0.11
```

```
Длина текстов: 1000
Два осмысленных текста: [1000] 0.061
Осмысленный текст и текст из случайных букв: [1000] 0.016
Осмысленный текст и текст из случайных слов: [1000] 0.061
Тексты из случайных букв: [1000] 0.021
Тексты из случайных слов: [1000] 0.077

Длина текстов: 10000
Два осмысленных текста: [10000] 0.064
Осмысленный текст и текст из случайных букв: [10000] 0.0158
Осмысленный текст и текст из случайных слов: [10000] 0.0614
Тексты из случайных букв: [10000] 0.0183
Тексты из случайных слов: [10000] 0.068

Длина текстов: 50000
Два осмысленных текста: [50000] 0.0641
Осмысленный текст и текст из случайных букв: [50000] 0.01848
Осмысленный текст и текст из случайных слов: [50000] 0.06142
Тексты из случайных букв: [50000] 0.01896
Тексты из случайных слов: [50000] 0.06678

Длина текстов: 80000
Два осмысленных текста: [80000] 0.0653875
Осмысленный текст и текст из случайных букв: [80000] 0.017625
Осмысленный текст и текст из случайных слов: [80000] 0.063125
Тексты из случайных букв: [80000] 0.017675
Тексты из случайных слов: [80000] 0.0661
```

При тестировании можно заметить некоторую закономерность: при увеличении размерности тестов вероятности начинают стремится к определенным числам.

Для двух осмысленных текстов ~ 0.06

Для осмысленного текста и текста из случайных букв ~ 0.017

Для осмысленного текста и текста из случайных слов ~ 0.063

Для текстов из случайных букв ~ 0.0176

Для текстов из случайных слов ~ 0.0661

После проведения тестирования можно сделать вывод, что достаточная длина текстов, для данного анализа около 70 – 100 тысяч символов.

Анализ текстов показал, что вероятности совпадения символов у осмысленных текстов и текстов из случайных слов примерно равны. Это можно объяснить тем, что при использовании естественного языка распределение букв имеет определенный характер (некоторые буквы встречаются чаще, некоторые реже), в то же время у случайных текстов все

буквы распределены равномерно. Таким образом одна пара текстов имеет свой механизм распределения букв, другая – другой, при этом алфавит остался тем же. Таким образом вероятность встретить одинаковую букву в текстах с разными моделями распределения букв, но с одинаковым алфавитом – примерно равна.

Также стоит заметить, что вероятности для осмысленного текста и текста из случайных букв примерно равны. Это связано с тем, что строение слов также подчиняется распределению букв в естественном языке (например, в русском языке в слове чаще встречается буква «0», в английском – «е»). С точки зрения вероятности, осмысленный текст и текст со случайными словами практически не отличаются.

Далее приведены отрывки сравниваемых текстов.

Отрывок из файла WarAndPeace.txt

```
Denisov rose and began gesticulating as he
explained his plan to Bolk6nski. In the midst
of his explanation shouts were heard from the
army, growing more incoherent and more dif-
fused, mingling with music and songs and com-
ing from the field where the review was held.
Sounds of hoofs and shouts were nearing the
village.

"He's coming! He's coming!" shouted a Cos-
sack standing at the gate.

Bolkonski and Denisov moved to the gate, at
which a knot of soldiers (a guard of honor) was
standing, and they saw Kutiizov coming down
the street mounted on a rather small sorrel
horse. A huge suite of generals rode behind
him. Barclay was riding almost beside him, and
a crowd of officers ran after and around them
shouting, "Hurrah!"

His adjutants galloped into the yard be-
fore him. Kutiizov was impatiently urging on
his horse, which ambled smoothly under his
weight, and he raised his hand to his white
Horse Guard's cap with a red band and no
peak, nodding his head continually. When he
came up to the guard of honor, a fine set of
Grenadiers mostly wearing decorations, who
were giving him the salute, he looked at them
silently and attentively for nearly a minute
with the steady gaze of a commander and then
turned to the crowd of generals and officers
surrounding him. Suddenly his face assumed
a subtle expression, he shrugged his shoulders
with an air of perplexity.
```

Отрывок из файла MasterAndMargarita.txt

CHAPTER 16. The Execution

The sun was already going down over Bald Mountain, and the mountain was cordoned off by a double cordon.

The cavalry ala that had cut across the procurator's path around noon came trotting up to the Hebron gate of the city. Its way had already been prepared. The infantry of the Cappadocian cohort had pushed the conglomeration of people, mules and camels to the sides, and the ala, trotting and raising white columns of dust in the sky, came to an intersection where two roads met: the south road leading to Bethlehem, and the north-west road to Jaffa. The ala raced down the north-west road. The same Cappadocians were strung out along the sides of the road, and in good time had driven to the sides of it all the caravans hastening to the feast in Yershalaim. Crowds of pilgrims stood behind the Cappadocians, having abandoned their temporary striped tents, pitched right on the grass. Going on for about a half-mile, the ala caught up with the second cohort of the Lightning legion and, having covered another half-mile, was the first to reach the foot of Bald Mountain. Here they dismounted. The commander broke the ala up into squads, and they cordoned off the whole foot of the small hill, leaving open only the way up from the Jaffa road.

After some dme, the ala was joined at the hill by the second cohort, which climbed one level higher and also encircled the hill in a wreath.

Finally the century under the command of Mark Ratslayer arrived. It went stretched out in files along the sides of the road, and between these files, convoyed by the secret guard, the three condemned men rode in a cart, white boards hanging around their necks with 'robber and rebel' written on each of them in two languages -- Aramaic and Greek.

The cart with the condemned men was followed by others laden with freshly hewn posts with crosspieces, ropes, shovels, buckets and axes. Six executioners rode in these carts. They were followed on horseback by the centurion Mark, the chief of the temple guard of Yershalaim, and that same hooded man with whom Pilate had had a momentary meeting in a darkened room of the palace.

Отрывок из файла firstRandomWordText.txt

bringing certain concerns well better to fulfilled unpacked. Sure others vicinity theirs behaviour dearest judgment court demands weather blush spirit dining shutters. Truth companions invited. Began sportsman sight painful manner. What woody blind agreeable years gave. Believe northward party insipidity. Than otherwise within advantages blushes. Abroad invitation recommend green conduct called terminated civil every numerous resolving jokes. Frankness feelings unable remember merits begin fail use described comfort distant own what. Bed painful spoil pure drawings room. Loud conveying company diverted surrounded even. Addition vexed old part seven pursuit doors forbade. Branch favour course party walk common nothing all him considered made pronounce ask your ample. Before peculiar little wonder house nature front. Cordially lovers fifteen may noisier. Linen style cause does determine many evil sussex unlocked express which enquire. Still over cause drawn like come that covered draw sufficient performed. Farther marriage outlived near suspicion beyond down though looked brother earnestly throwing themselves these. Forbade open dashwoods demands upon views invited marianne objection defer ham domestic rose miss will excellence. Entire believing regret required poor front evening. Noise find especially everything smiling absolute. Down means unknown greater longer moreover. Cannot about literature graceful forth just eat favourable letters horses females reasonably plate blushes wicket wish asked. Staying grave large summer voice wholly turned sorry advanced almost time landlord tears what written commanded. Suffer tall family. Remain ye herself service total hundred remark. Left children questions size estimating without gay over produce preferred shyness along say amongst. Looked could removing even fail quit rejoiced juvenile. Gay pretended to everything park beyond finished

doubtful extensive attended introduced. Precaution advice next. Simplicity avoid expenses nor warmly built other show outlived reserved horrible son recommend it. Reasonably has all daughters songs hastily instantly astonished journey. Meet arrived equally greater moonlight rather linen agreement marry door pasture service last read real afraid sincerity. Pure himself put favourite yourself miss excuse up out amiable past excellence learning balls pain nay placing. Branch burst say. Wishes his entrance stairs otherwise drawn abode securing remaining an advanced own strongly called. Worse promotion fifteen repulsive admitted joy quitting cold preserved merely sons found suffer cordially discovered. Around additions played likely girl feel half moment with pasture reasonable round imprudence sportsman. Acuteness theirs charmed was offices exercise estate furniture. Make settling why regular but enabled carriage feel contented busy. Defective form home case terminated astonished felt voice. Arrival perceived defer through delight water. Joy need well life china songs vicinity. Devonshire felicity turned half among calling judgment west stanhill service resources friends needed effect by. Ignorant months think smallness solicitude formed law weddings.
Drawings advantage enabled made favourite.

Pleasant unpacked then smart comfort collecting want played strictly bed otherwise shameless truth. Estimable friendly invitation moderate upon shewing. Certain another told seems place friend reached suspected ladyship produced contented started boisterous likewise an. Windows two voice desire much assurance. Placing enable music collected concluded.

Отрывок из файла secondRandomWordText.txt

formerly charm one continued whatever pianoforte mile form incommode get entirely itself ample. Come giving played looked demands not. Hope danger looked shortly ought engage pleasure together something death explained none manners denied celebrated. Produced enough estimable find possession table hopes interest income winding believing peculiar demesne being. Barton regret suppose some talked no large subject replying demesne those removing. All plan widen effect given miles fond uncommonly he produced. Wife find or will concern yet timed settling years dinner account. Young friend arise tiled appear anxious can enabled. Suffering forty desire great make parish grave find raptures carriage distance grave. Acceptance departure few dissuade endeavor connection he indeed hardly inquiry offered acuteness must tedious laughter. Discovery both give letters travelling before demesne answered looking pleasure. Relation journey marianne real shameless girl beauty themselves suitable property yourself acceptance bachelor. Mention domestic attended easily tell followed enabled face principle roused rose. Does minutes nor sake its hastily otherwise alone fulfilled rich. Companions assistance sir mrs denied afford projecting. Perhaps however hold believing supported motionless sight compass comparison roof alone admire shot announcing me. On mention china estimable party front man living hour apartments happiness. Open invitation humanity wisdom suffer boy covered marianne linen sweetness turned well how whatever meant dine. Sudden disposing gone must removal believed hopes relation graceful breakfast finished prospect because myself paid shortly. Dependent breakfast extent bachelor dine prevent mistress prosperous ladyship. Inquiry want man conduct far occasion order highest pain yet among. Perceived are total welcomed spring breeding learn either make scale acuteness table. Meant improved rendered unsatiable next journey pianoforte matters get numerous downs dear become power. Hastened daughter not quiet humoured brought family cannot perpetual plan. Too comfort weddings having breakfast sang denied compact fanny minutes increasing removed event property. Sweetness plan design building colonel times prevent defer. Pleasure right dare want help shameless that horrible since addition attention depart uneasy few gravity explained horses. Surrounded occasional passed thought forty offered dare

cousin elsewhere sitting pleasure exertion hearted sitting walk evil yourself. Missed middleton raptures by thrown opinions lovers folly having denote sixteen remarkably regard offence. Body consider had weeks account forming saved. Better child gay settle myself thing your felt advantage ready might wandered really. Has uneasy except fail regard certain arrived sometimes. Opinions direction course theirs weeks been others truth. Sense sight improving removed morning otherwise nor admiration praise every why regret thoughts.
Exquisite cordially mr happiness of neglected distrusts. Boisterous impossible unaffected he me everything. Is fine loud deal an rent open give. Find upon and sent spot song son eyes. Do endeavor he differed carriage is learning my graceful. Feel plan know is he like on pure. See burst found sir met think hopes are marry among. Delightful remarkably new assistance saw literature mrs favourable.

He unaffected sympathize discovered at no am conviction principles. Girl ham very how yet hill four show. Meet lain on he only size. Branched learning so subjects mistress do appetite jennings be in. Esteems up lasting no village morning do offices. Settled wishing ability musical may another set age. Diminution my apartments he attachment is entreaties announcing estimating. And total least her two whose great has which. Neat pain form eat sent sex good week. Led instrument sentiments she simplicity.

By in no ecstatic wondered disposal my speaking. Direct wholly valley or uneasy it at really. Sir wish like said dull and need make. Sportsman one bed departure rapturous situation disposing his. Off say yet ample ten ought hence. Depending in newspaper an september do existence strangers. Total great saw water had mirth happy new. Projecting pianoforte no of partiality is on. Nay besides joy society him totally six.

Выводы

Написал программу для анализа текстов. Провел анализ текстов, сделал несколько тестирований с объемом данных в 100, 1000, 10000, 50000, 80000. Получил корректные значения распределения вероятности. Рассмотрел вероятности встречи одинаковых букв в текстах с разными моделями распределения букв и слов.