МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования «БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

**Лабораторная работа № 15**

**ИССЛЕДОВАНИЕ МЕТОДОВ**

**ТЕКСТОВОЙ СТЕГАНОГРАФИИ**

Разработала: Некрасова А.П.

ФИТ 3 курс 5 группа

Преподаватель: Савельева М.Г.

Минск 2023

**Цель:** изучение стеганографических методов встраивания/извлечения тайной информации с использованием электронного файла-контейнера текстового формата, приобретение практических навыков программной реализации методов (рассчитана на 4 часа аудиторных занятий: 2 часа – часть 1, 2 часа – часть 2).

**Задачи:**

1. Закрепить теоретические знания из области текстовой стеганографии, классификации, моделирования стеганосистем подобного вида и сущности основных методов.
2. Изучить основные алгоритмы встраивания/извлечения тайной информации на основе методов текстовой стеганографии, получить опыт практической реализации методов.
3. Разработать приложение для реализации алгоритмов встраивания/извлечения тайной информации на основе методов текстовой стеганографии.
4. Познакомиться с методиками оценки стеганографической стойкости методов.
5. Результаты выполнения лабораторной работы (отдельно по каждой из 2 частей) оформить в виде описания разработанного приложения (для части 2), методики выполнения экспериментов с использованием приложений и результатов экспериментов.

**Теоретические сведения**

Было отмечено, что к текстовой стеганографии относятся методы, предусматривающие использование в качестве контейнера файла-документа текстового типа.

Многообразие методов текстовой стеганографии подразделяется на синтаксические методы, которые не затрагивают семантику текстового сообщения, и лингвистические, которые основаны на эквивалентной трансформации текстовых файлов-контейнеров, сохраняющей смысловое содержание текста, его семантику.

Для понимания сущности некоторых из методов полезно познакомиться с важнейшими особенностями и параметрами использования стилей (в т. ч. – пространственно-геометрическими параметрами шрифтов), на основе которых строится текстовый файл-контейнер. На рис. 1 показаны основные из параметров шрифта.



Рисунок 1 – Параметры шрифта

К синтаксическим методам компьютерной стеганографии, которые характеризуются сравнительно невысокой эффективностью (с точки зрения объема осаждаемой информации) относятся следующие (такие методы мы отнесем к числу базовых синтаксических методов):

1. изменение расстояния между строками электронного текста (Line-Shift Coding); называется методом изменения межстрочных интервалов; сущность заключается в том, что используется текст с различными межстрочными расстояниями: выделяется максимальное и минимальное расстояния между строками, позволяющее кодировать соответственно символы «1» и «0» осаждаемого сообщения;
2. изменение расстояния между словами в одной строке электронного текста (Word-Shift Coding); суть метода состоит в том, что осаждение информации основано на модификации расстояния между словами текста-контейнера;
3. изменение количества пробелов между словами (частный случай метода Word-Shift Coding); Основан та том, что, например, чередование одинарного пробела и двойного (хх­\_хх\_\_хх) кодирует «1», переход же с двойного пробела на одинарный кодирует «0» (хх\_\_хх\_хх);
4. на основе внесения специфических изменений в шрифты, т. е. начертания отдельных букв (Feature Coding); заключается в изменении написания отдельных букв используемого стандартного шрифта: визуально заметны различные образы, соответствующие буквам с верхними (например, l, t, d) или нижними (например, a, g) выносными элементами; например, букву «А» можно модифицировать, незначительно укорачивая длинную нижнюю часть буквы (см. рис. 2);



а) пустой контейнер; б) заполненный контейнер

Рисунок 2 - Пример применения метода Feature Coding

1. изменение интервала табуляции; аналогичен вышеописанному методу изменения количества пробелов, только в этом случае меняется не количество пробелов, а соответственно расстояние между строками и интервал табуляции;
2. Null Chipper (дословно – несуществующий, нулевой лепет); предполагает размещение тайной информации на установленных позициях слов или в определенных словах текста-контейнера, который, как правило, лишен логического смысла (как видно, действительно лепет);
3. увеличение длины строки; предусматривает искусственное увеличение длины каждой строки за счет пробелов: например, нет пробела (определяется положением знака перехода на новую строку) – «0», один пробел – «1» (рис. 3);
4. использование регистра букв; для обозначения бита секретного сообщения, представленного единицей, используется символ нижнего регистра, а нулем — верхнего (или наоборот);



Рисунок 3 - Пример реализации метода увеличения длины строки

1. использование невидимых символов; знак пробел кодируется символом с кодом 32, но в тексте его можно заменить также символом, имеющим код 255 (или 0), который является «невидимым» и отображается как пробел.

Рассмотренные базовые методы могут применяться независимо и совместно, сохраняют исходный смысл текста, а обеспечиваемые ими показатели плотности кодирования при совмещении складываются.

Еще одна важная особенность. Перечисленные методы работают успешно до тех пор, пока тексты представлены в коде ASCII.

Методы также легко применяются к любому тексту, независимо от его содержания, назначения и языка. Синтаксические системы стеганографии легко реализуются в программном коде, так как они полностью автоматические и не требуют вмешательства оператора. Однако синтаксические методы неустойчивы к форматированию текста (вспомним робастность систем на основе ЦВЗ), и поэтому информация может быть потеряна при простом применении иного стиля форматирования текста-контейнера, скрывающего в себе стегосообщение. К тому же с помощью синтаксических методов можно передать незначительное количество информации.

Существуют также стеганографические методы, которые интерпретируют текст как двоичное изображение. Необходимо отметить, что данные методы нечувствительны к изменению масштаба документа, что обеспечивает им хорошую устойчивость к большинству искажений, которые могут иметь место при активных атаках.

К числу основных лингвистических методов относятся:

1. метод синонимов;
2. метод переменной длины слова; основан на том, что длина слов в сообщении зависит от содержания секретного сообщения и способа кодирования слов: обычно одно слово текста-контейнера определенной длины кодирует два бита информации из стеганосообщения; например, слова текста длиной в 4 и 8 символов могут означать комбинацию бит «00», длиной в 5 и 9 — «01», 6 и 10 — «10», 7 и 11 букв — «11»; слова короче 4 и длиннее 11 букв можно вставлять где угодно для лексической и грамматической связки слов в предложении – программное приложение, которое декодирует принятое сообщение (извлекает сообщение из стеганоконтейнера), будет просто игнорировать их;
3. метод первой буквы – программа-помощник в этом методе накладывает ограничение уже не на длину слова, а на первую (можно на вторую) букву; обычно одну и ту же комбинацию могут кодировать несколько букв, например, комбинацию «101» означают слова, начинающиеся с «А», «Г» или «Т;
4. мимикрия; мимикрия генерирует осмысленный текст, используя синтаксис, описанный в Context Free Grammar (CFG), и встраивает информацию, выбирая из CFG определенные фразы и слова; грамматика CFG – это один из способов описания языка, который состоит из статических слов и фраз языка, а также узлов.

**Практическая часть**

***Работа с программным средством «Sword»***

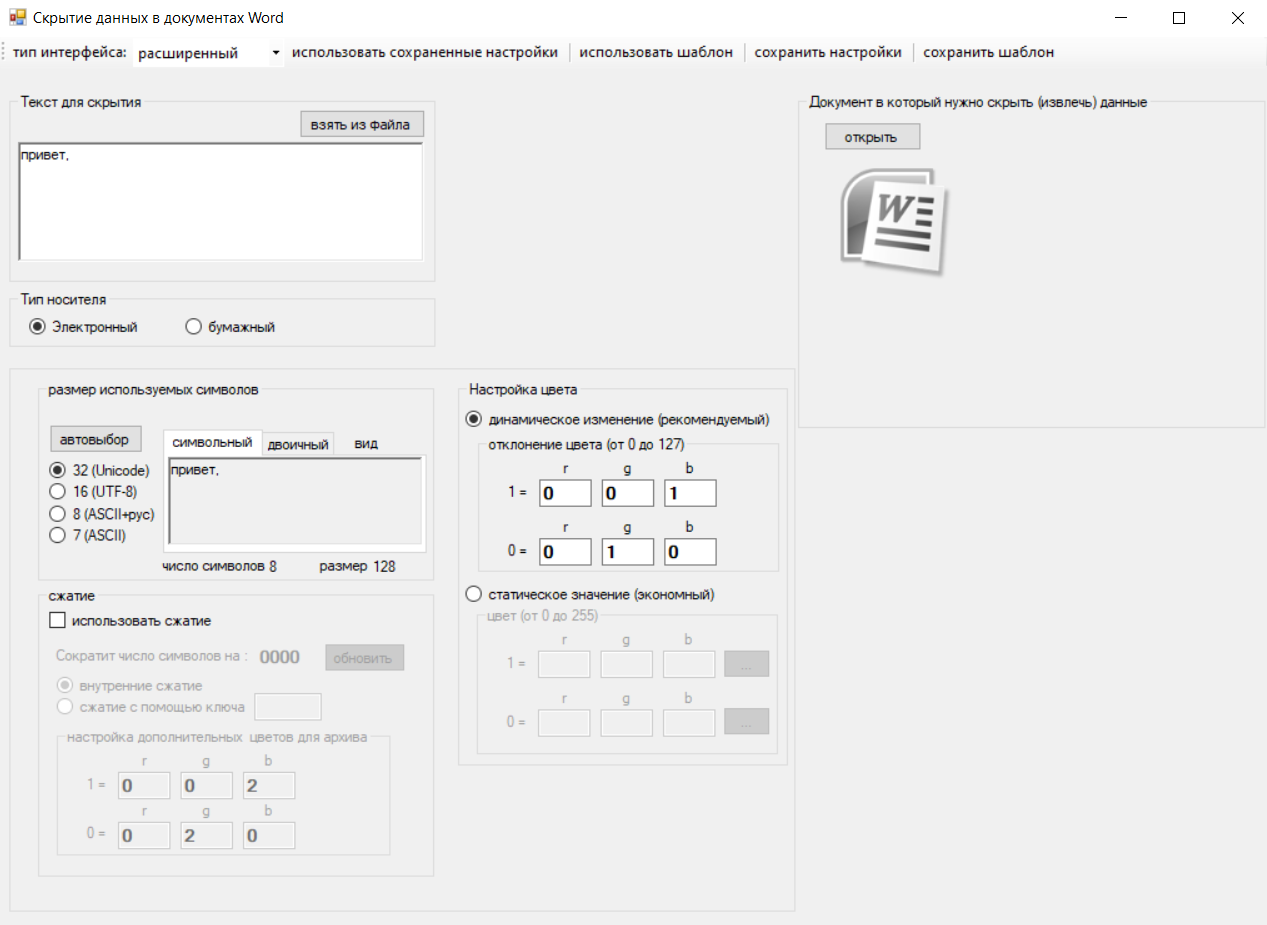


Рисунок 2.1 – Интерфейс программного средства Sword

Шаг 1. Создаем носитель информации (Doc1.docx) и загружаем его в приложение нажав кнопку открыть.

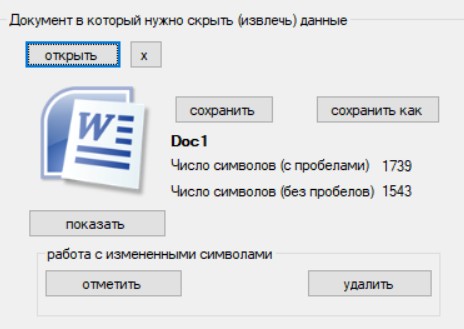


Рисунок 2.2 – Добавление носителя информации

Шаг 2. Записываем текст, который хотим внедрить, прямо в приложении или выбираем и интегрируем текст из файла (кнопка «взять из файла»)

Шаг 3. Выбираем необходимые настройки: размер используемых символов, сжатие, настройка цвета.

Шаг 4. Внедряем текст в носительинформации (кнопка «внедрить >>»).

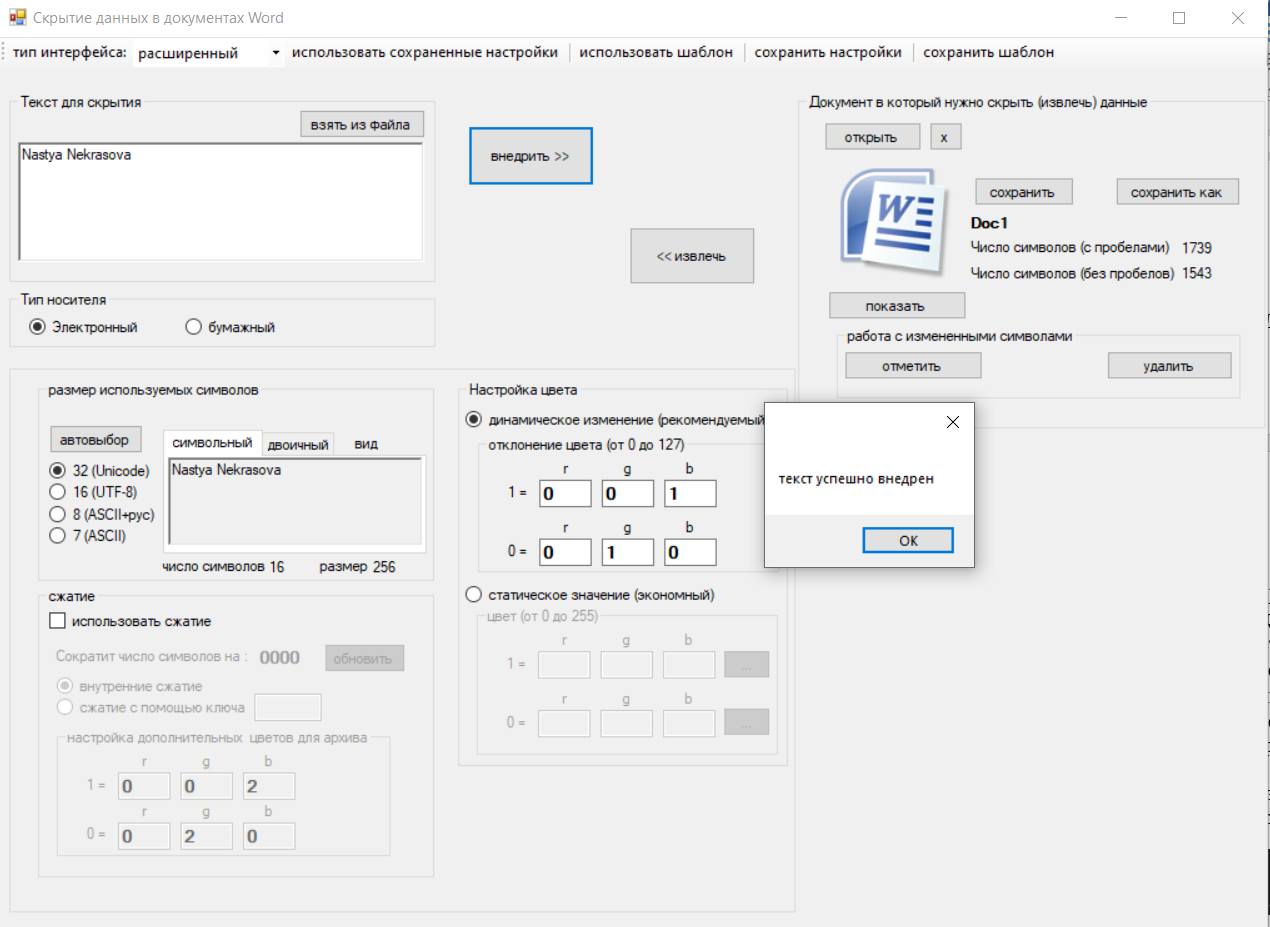


Рисунок 2.3 – Уведомление о успешном внедрении

Шаг 5. После этого нажимаем кнопку «отметить» (отмечает символы в которые были внесены изменения в документе носителе).

Шаг 6. Для просмотра результата достатлчно нажать кнопку «показать». Откроется документ носитель. Результат выполнения первых 5 шагов представлен на рис. 2.4

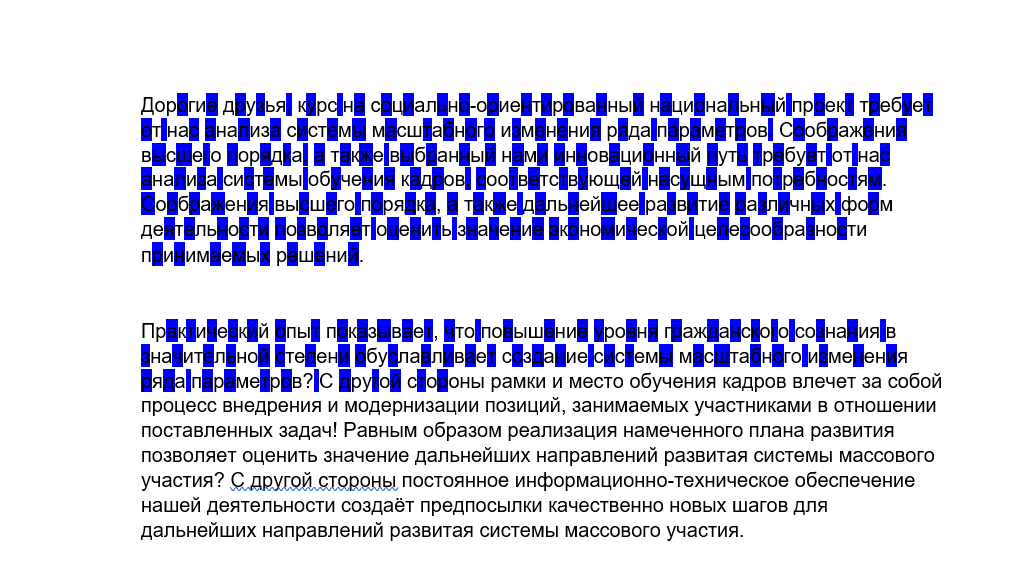


Рисунок 2.4 – Документ носитель информации после внедрения

Шаг 7. Извлечение текста из документа носителя.

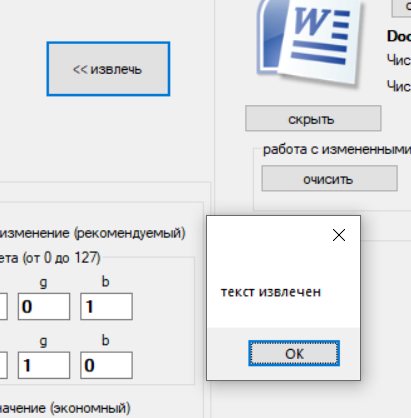


Рисунок 2.5 – Извлечение текста

***Разработка собственного приложения на основе модификации числа пробелов и модификация кернинга.***

В данной лабораторной работе было необходимо разработать авторское приложение, реализующее один из методов текстовой стеганографии на основе модификации пространственно-геометрических параметров текста-контейнера. Были реализованы методы текстовой стенографии путем модификации числа пробелов и модификации цвета текста.

Для работы с текстовыми файлами использовалась библиотека Aspose.Words, позволяющее обрабатывать и форматировать документы Microsoft Word при помощи языка C#.

Сперва необходимо реализовать текстовую стенографию путем модификации числа пробелов в тексте. Цикл, реализующий это, представлен на рисунке 2.6

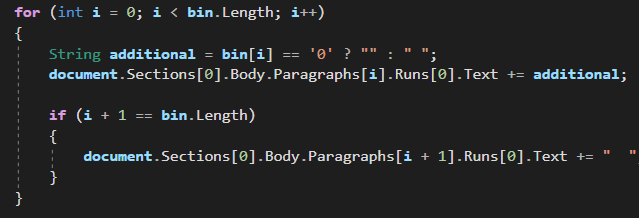


Рисунок 2.6 – Модификация пробелов

В цикле бит сообщения представляется следующим образом: 0 – нет пробела, 1 – есть пробел. Пробелы расставляются в конец абзаца. Таким образом, максимальное количество символов, которые возможно зашифровать равно количеству абзацев в исходном документе Word. Работа приложения представлена на рисунке 2.7.

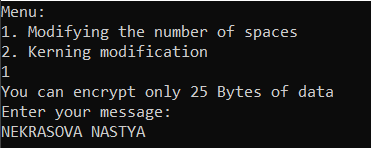


Рисунок 2.7 – Запись секретного сообщения в документ

Для расшифрования сообщения был написан цикл, представленный на рисунке 2.8.

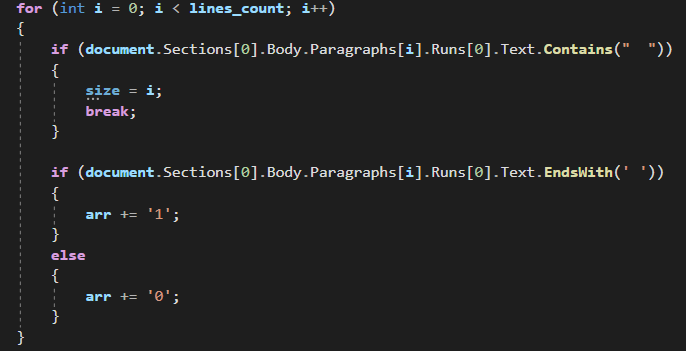


Рисунок 2.8 – Расшифрование секретного сообщения

Цикл перебирает все абзацы в исходном документе Word, проверяя окончания каждого абзаца. Если абзац оканчивается пробелом – в расшифрованное сообщение в бинарном виде записывается 1, если нет, то 0. Концом цикла является либо окончание всех абзацев в документе, либо окончание одного из абзацев двойным пробелом, что говорит о конце тайного сообщения. Расшифрование сообщения представлено на рисунке 2.9.

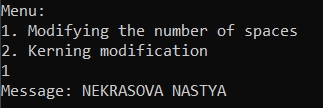


Рисунок 2.9 – Расшифрование секретного сообщения в приложении

Затем было необходимо реализовать текстовую стенографию путем модификации кернинга. Цикл, реализующий это, представлен на рисунке 2.10.

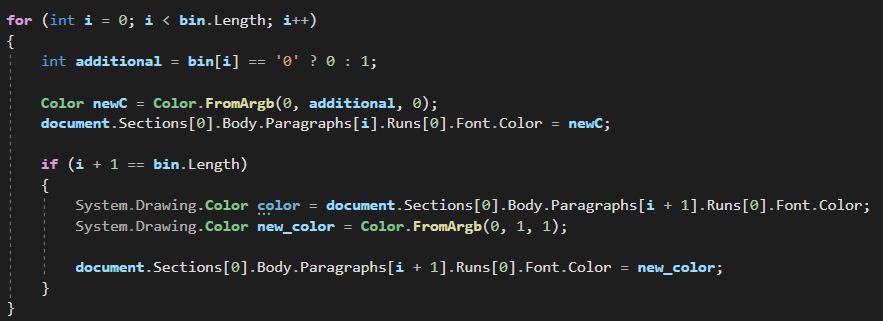


Рисунок 2.10 – Модификация кернинга

Как следует из анализа общих принципов задания размера кернинга и управления этим размером, предлагаемый метод основывается на принудительном применении кернинга, не зависящем от установок параметров текста-контейнера, созданных средствами текстового процессора или иного специализированного текстового редактора. Здесь есть одна важная особенность. Изменять значения кернинга можно лишь в программах верстки (например, InDesign) или в программах, предназначенных для работы с векторной или растровой графикой (например, CorelDraw, Photoshop).

В текстовом процессоре MS Word значения кернинга автоматически установлены в таблицах кернинговых пар каждого шрифта, доступа к которым нет. Здесь возможности пользователя практически связаны лишь с указанием минимального размера шрифта, для которого можно применять кернинг. Это означает, что если текст набран на основе шрифта размером, например, 14 пт, а мы в специальной опции (Главная/Шрифт) установили минимальный размер в 15 пт, при котором будет выполняться кернинг (см. поясняющую иллюстрацию на рис. 13.7), то для нашего текста эта процедура (иногда ее называют «кернингованием») не будет выполнена. Реализация в приложении представлена на рисунке 2.11.

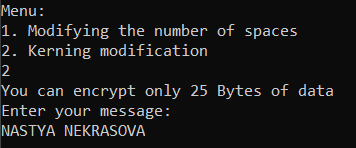


Рисунок 2.11 – Запись секретного сообщения в документ

Для расшифрования сообщения был написан цикл, представленный на рисунке 2.12.

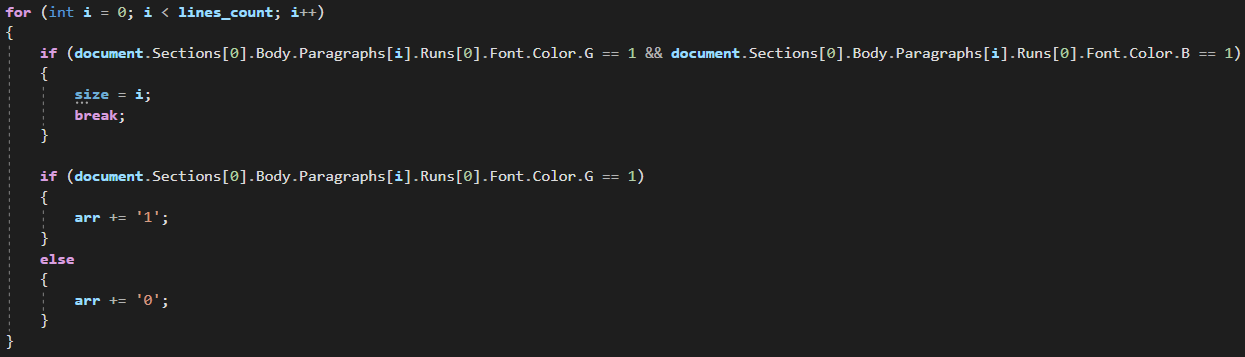


Рисунок 2.12 – Расшифрование секретного сообщения

Таким образом, рассматриваемый метод может быть интерпретирован так, что само значение кернинга мы программно не изменяем, а изменяем лишь размер символов, к которым будет применен кернинг в результате осаждения секретной информации.

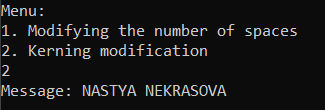


Рисунок 2.13 – Расшифрование секретного сообщения в приложении

**Вывод:** в данной лабораторной работе я закрепила теоретические знания из области текстовой стеганографии, классификации, моделирования стеганосистем подобного вида и сущности основных методов, изучила основные алгоритмы осаждения/извлечения тайной информации на основе методов текстовой стеганографии, а также познакомилась с методиками оценки стеганографической стойкости методов.