

**МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»**

**(РУТ (МИИТ)**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Институт транспортной техники и систем управления**

**Кафедра «Управление и защита информации»**

Отчет по лабораторной работе №1

**«Сравнение энтропии распакованных и упакованных файлов»**

по дисциплине

**«Технология реверс-инжиниринга»**

**Выполнили:** студенты ТКИ-341

Козлов А. Д. и Дьячков Д.О.

**Проверили:** Профессор “УиЗИ”Сидоренко В.Г.

Профессор “УиЗИ” Сафронов А. И.

Москва 2025 г.

ОГЛАВЛЕНИЕ

[**Цель работы** 3](#_Toc191324519)

[**Поставленные задачи** 3](#_Toc191324520)

[**Представления различных файлов и программ в hex-редакторе** 4](#_Toc191324521)

[**Сравнение энтропии распакованных и упакованных исполняемых файлов программ** 7](#_Toc191324522)

[**Гистограммы и энтропия С++** 7](#_Toc191324523)

[**Гистограммы и энтропия С#** 10](#_Toc191324524)

[**Гистограммы и энтропия Python** 12](#_Toc191324525)

[**Сравнение энтропии аудиофайлов разных форматов** 15](#_Toc191324526)

[**Сравнение энтропии файлов изображений разных форматов** 18](#_Toc191324527)

[**Сравнение энтропии текстовых файлов разных форматов** 21](#_Toc191324528)

[**Сравнение энтропии файлов таблиц разных форматов** 24](#_Toc191324529)

[**Таблицы сравнений энтропии** 28](#_Toc191324530)

[**Вывод** 30](#_Toc191324531)

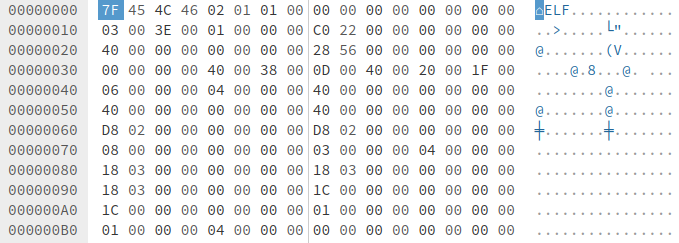
# **Цель работы**

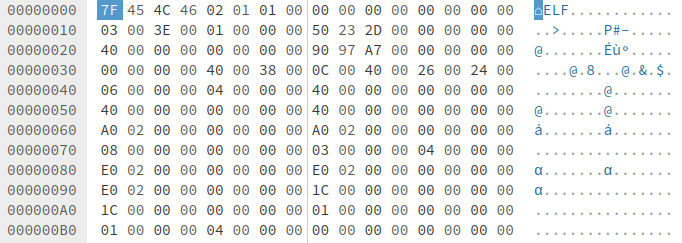
Исследование файлов разных форматов и разной степенью сжатия с целью изучения их структуры и определения влияния сжатия на энтропию этих файлов.

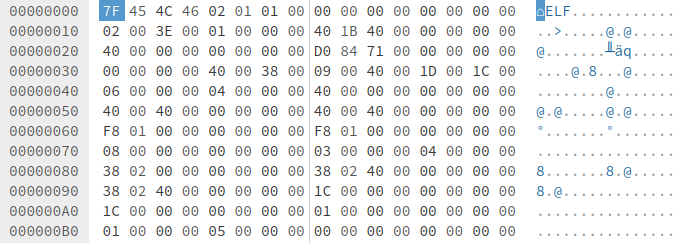
# **Поставленные задачи**

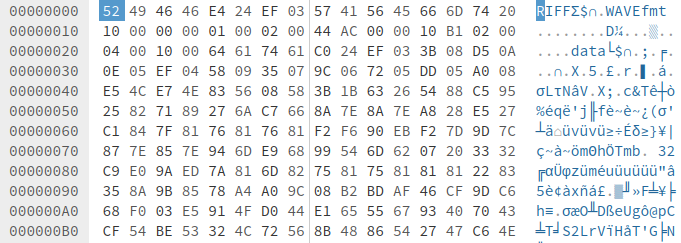
1. Анализ файлов в hex-редакторе. Выбрать несколько файлов разных типов близкого размера (изображение, музыка, таблица, текстовый файл, тексты программ, реализующих одни и те же действия, написанные на разных языках программирования, и их исполняемые файлы), просмотреть их в hex-редакторе.
2. Сравнение энтропии распакованных и упакованных исполняемых файлов программ, а также сравнение файлов одного типа в разных форматах данных. Упаковать выбранный файл с использованием UPX, просмотреть его в hex-редакторе, сравнить результаты, полученные в п.п.1 и 2. При помощи программы radare2, NIST, или другого ПО рассчитать энтропию исходного и упакованного файлов, провести анализ полученных результатов.

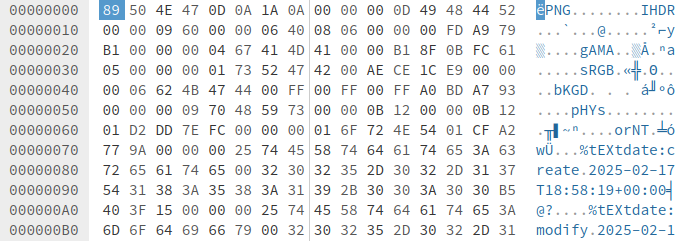
# **Представления различных файлов и программ в hex-редакторе**

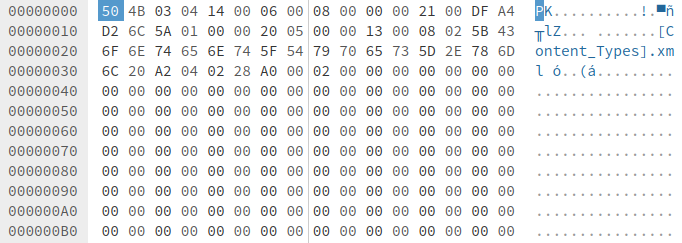
  
Рис 1. Представление исполняемого файла с++ в hex-редакторе

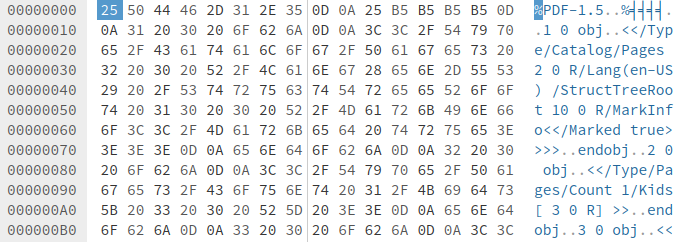
  
Рис 2. Представление исполняемого файла с# в hex-редакторе

  
Рис 3. Представление исполняемого файла python в hex-редакторе

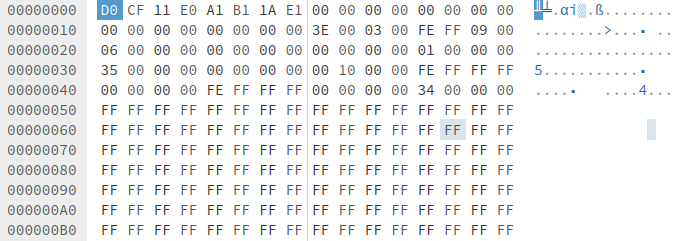
  
Рис 4. Представление аудиофайла в hex-редакторе

  
Рис 5. Представление файла изображения в hex-редакторе

  
Рис 6. Представление текстового файла формата .docx в hex-редакторе

  
Рис 7. Представление текстового файла формата .pdf в hex-редакторе

D:\Miit-Labs\TRE-Labs\Lab1\HexEditor\hex-text-txt.png  
Рис 8. Представление текстового файла формата .txt в hex-редакторе

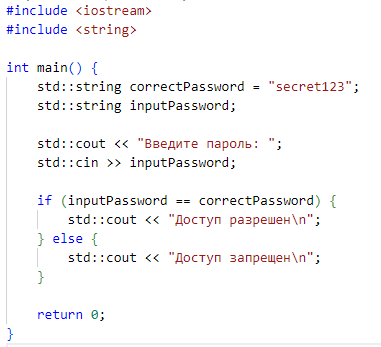
  
Рис 9. Представление файла таблицы формата xls в hex-редакторе

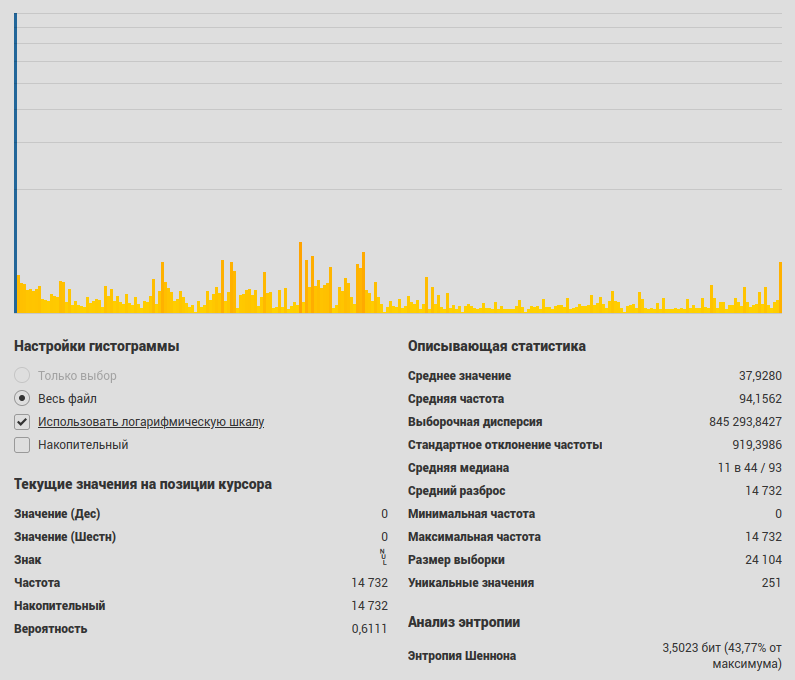
# **Сравнение энтропии распакованных и упакованных исполняемых файлов программ**

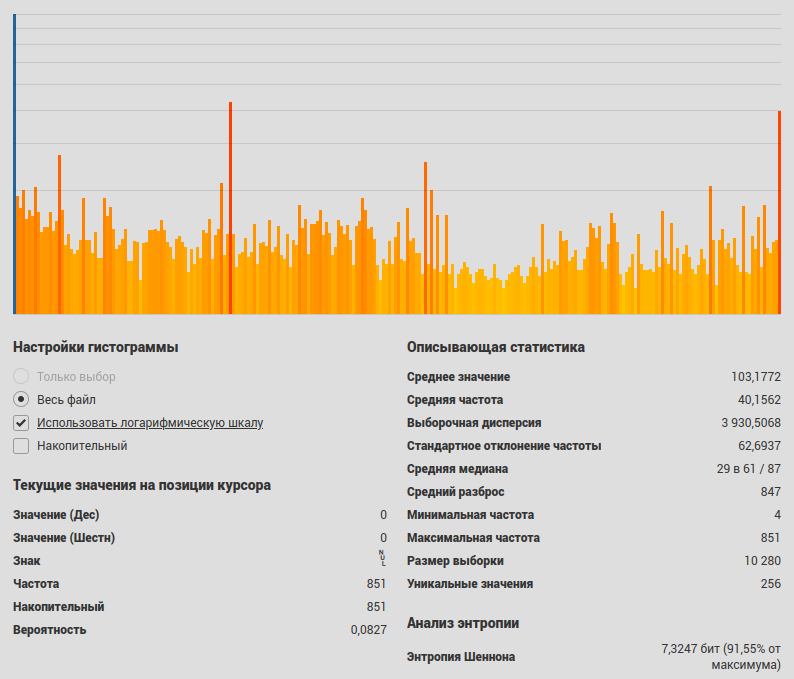
Для анализа исполняемых файлов были подготовлены и скомпилированы программы, написанные на языках c++, c# и python, а также были подготовлены упакованные версии этих программ с использованием упаковщика UPX и GZEXE. С использованием hex-редактора и утилиты ENT мы получили сравнили энтропию файлов программ и их запакованных версий. Данное сравнение помогает нам представить степень сжатия данных, а также насколько снижена предсказуемость данных.

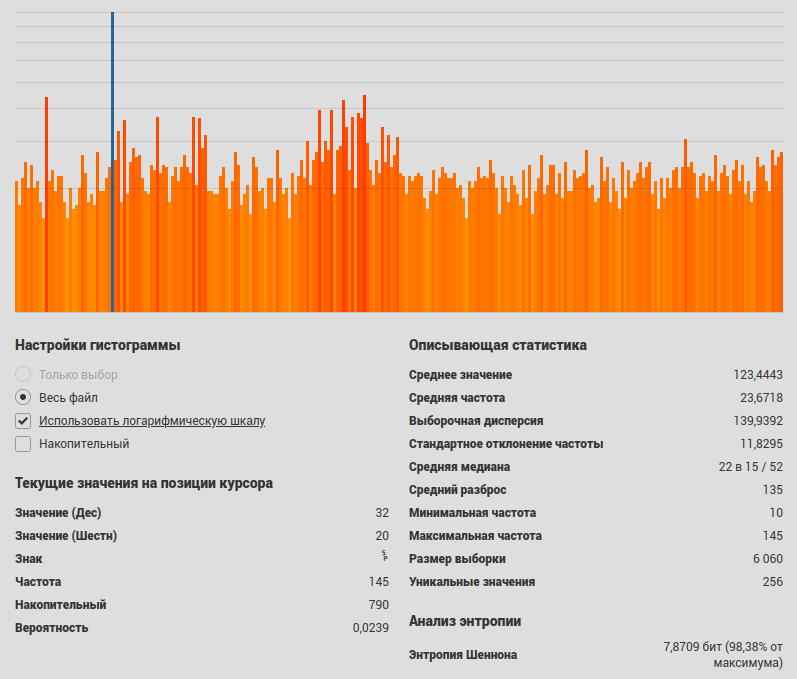
Построенные ниже диаграммы позволяют наблюдать за изменением распределения частот значений байтов в файле.

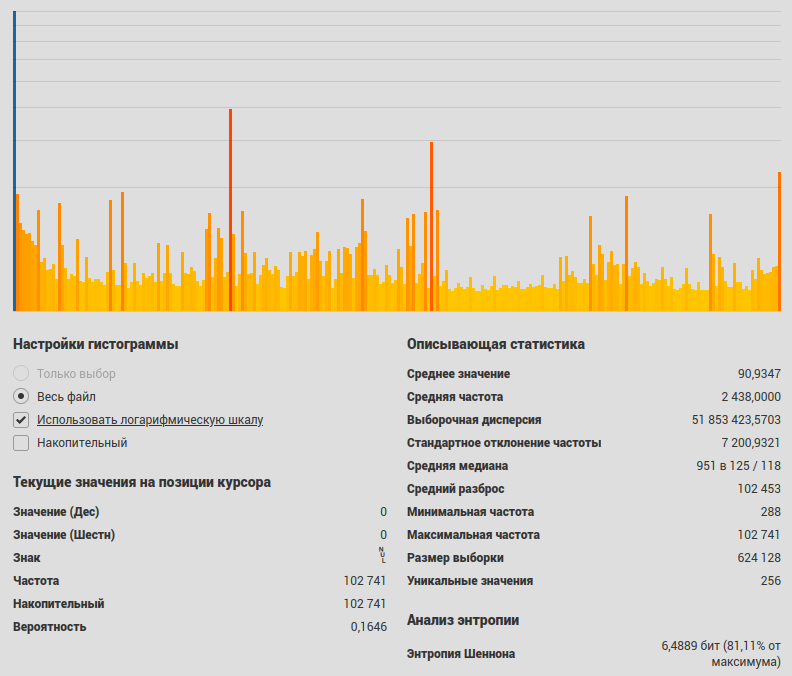
## **Гистограммы и энтропия С++**

  
Рис 10. Листинг программы C++

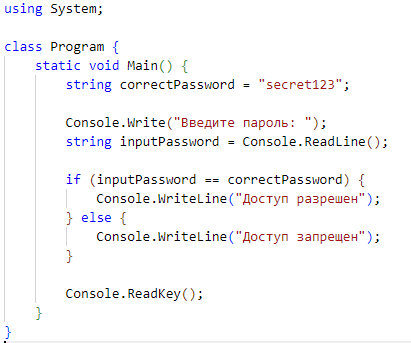
  
Рис 11. Гистограмма и энтропия исполняемого файла C++ (без упаковки)

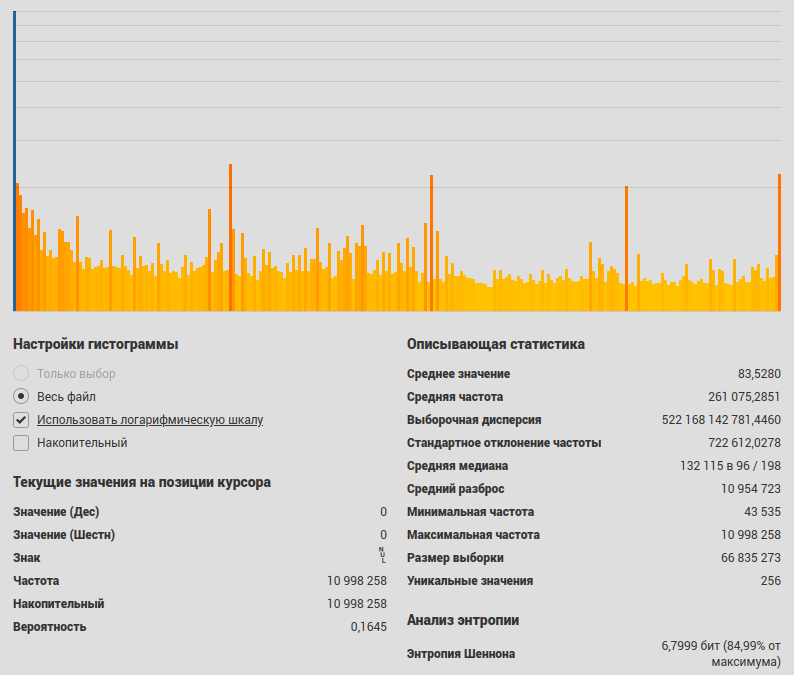
  
Рис 12. Гистограмма и энтропия исполняемого файла C++ (UPX)

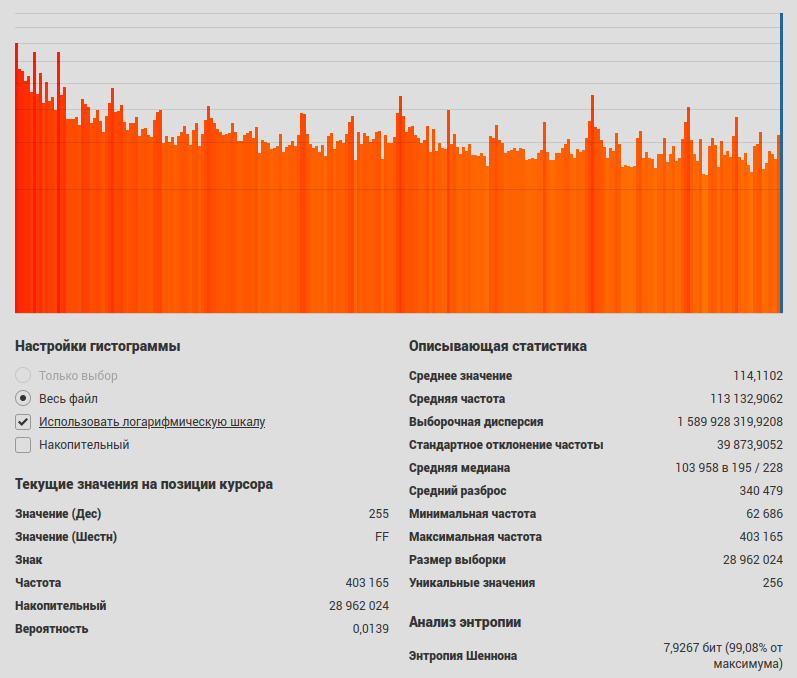
  
Рис 13. Гистограмма и энтропия исполняемого файла C++ (GZEXE)

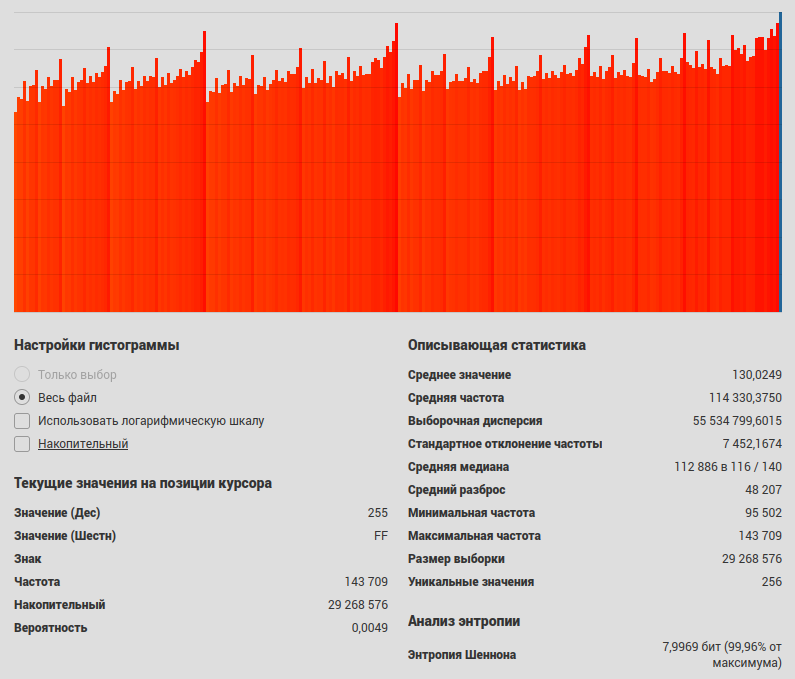
  
Рис 14. Гистограмма и энтропия исполняемого файла C++ (упаковщик Перепелкина Н.)

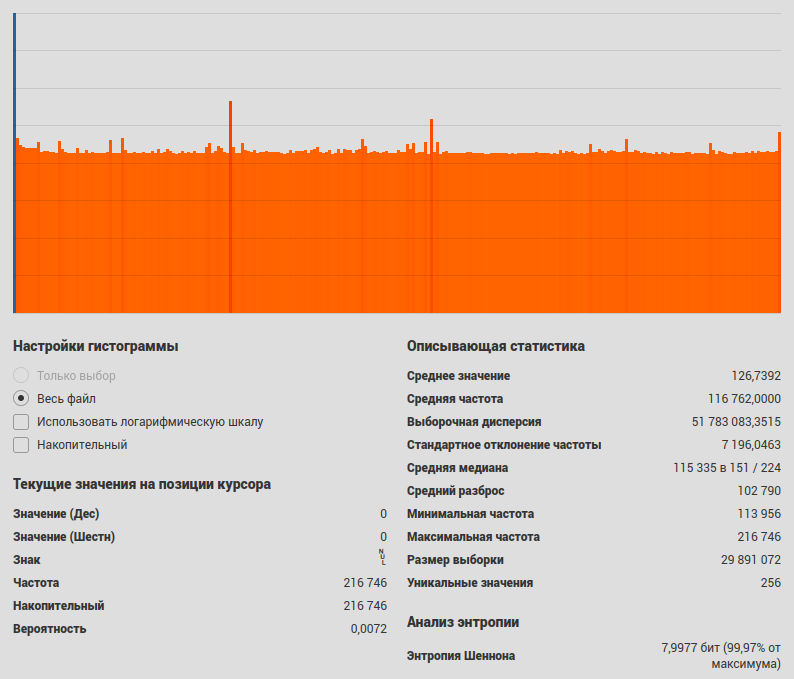
## **Гистограммы и энтропия С#**

  
Рис 15. Листинг программы С#

  
Рис 16. Гистограмма и энтропия исполняемого файла C# (без упаковки)

  
Рис 17. Гистограмма и энтропия исполняемого файла C# (UPX)

  
Рис 18. Гистограмма и энтропия исполняемого файла C# (GZEXE)

  
Рис 19. Гистограмма и энтропия исполняемого файла C# (упаковщик Перепелкина Н.)

## **Гистограммы и энтропия Python**

  
Рис 20. Листинг программы Python

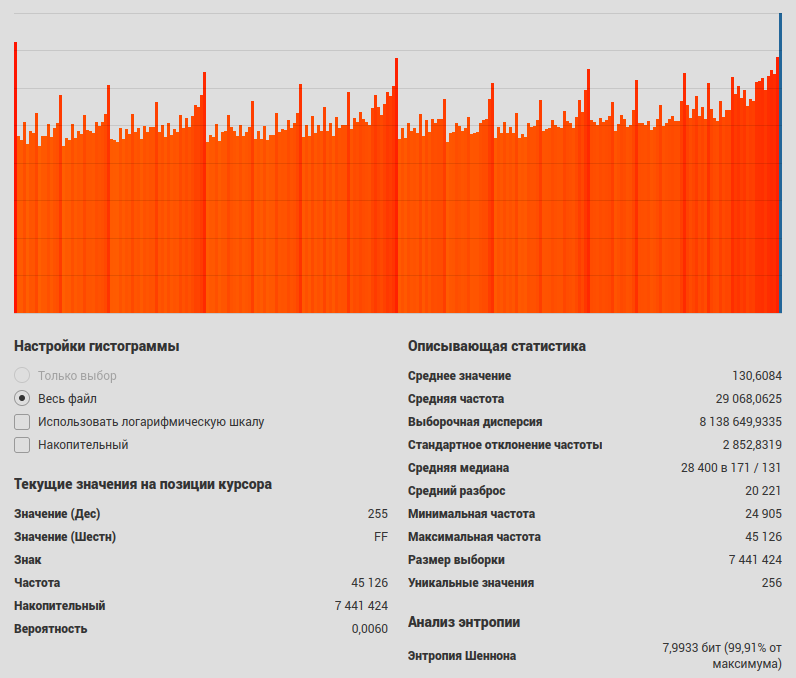
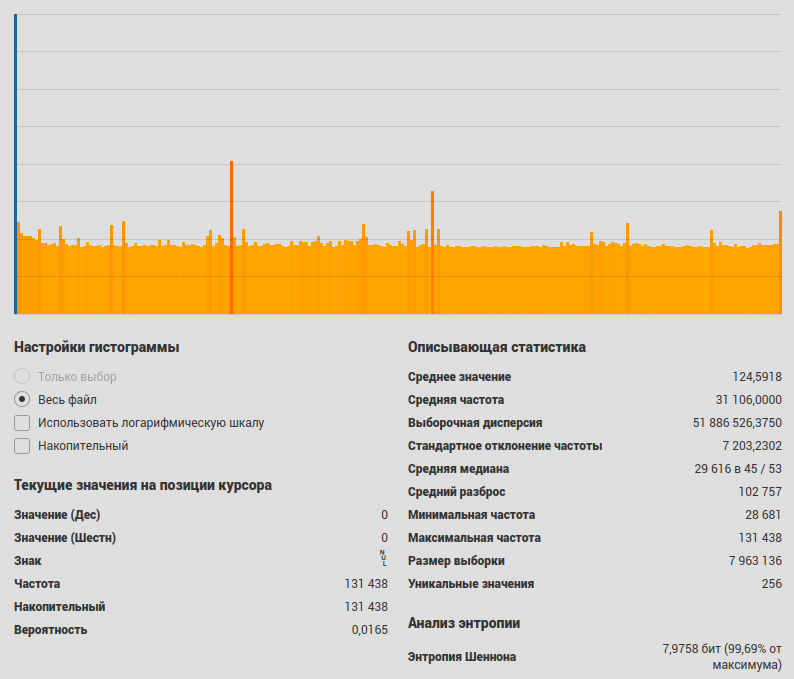
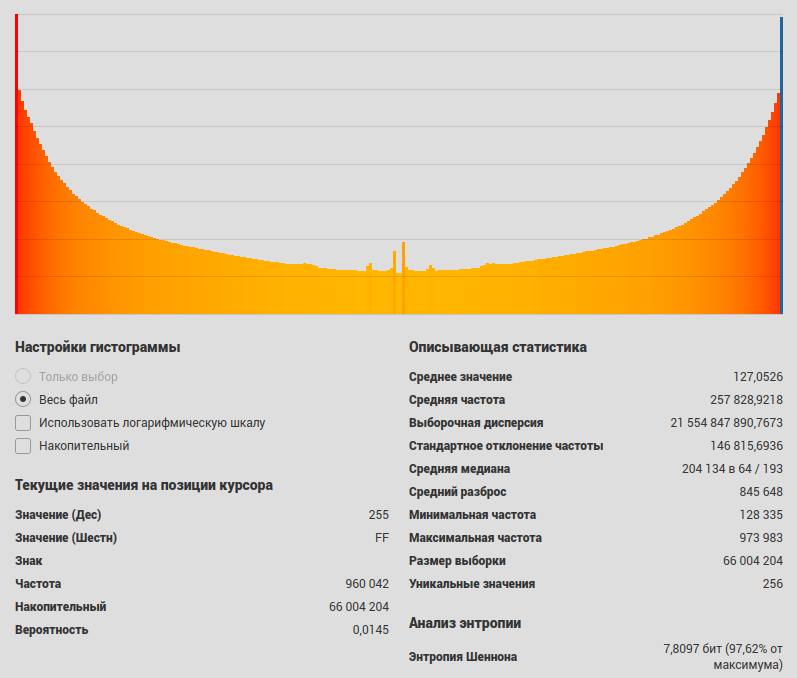
  
Рис 21. Гистограмма и энтропия исполняемого файла Python (без упаковки)

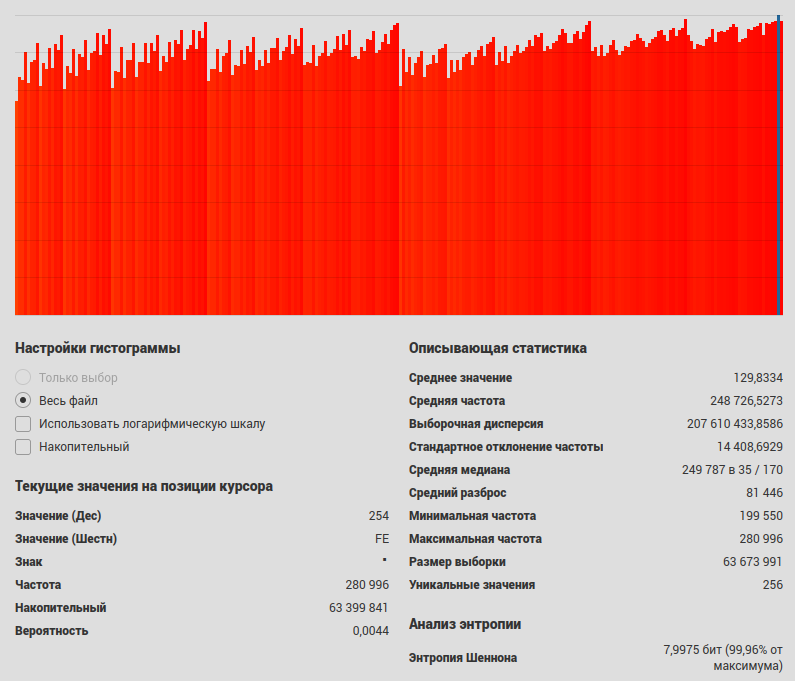
  
Рис 22. Гистограмма и энтропия исполняемого файла Python (UPX)

  
Рис 23. Гистограмма и энтропия исполняемого файла Python (GZEXE)

  
Рис 24. Гистограмма и энтропия исполняемого файла Python (упаковщик Перепелкина Н.)

# **Сравнение энтропии аудиофайлов разных форматов**

  
Рис 25. Гистограмма и энтропия аудиофайла формата .wav

  
Рис 26. Гистограмма и энтропия аудиофайла формата .wav (zip)

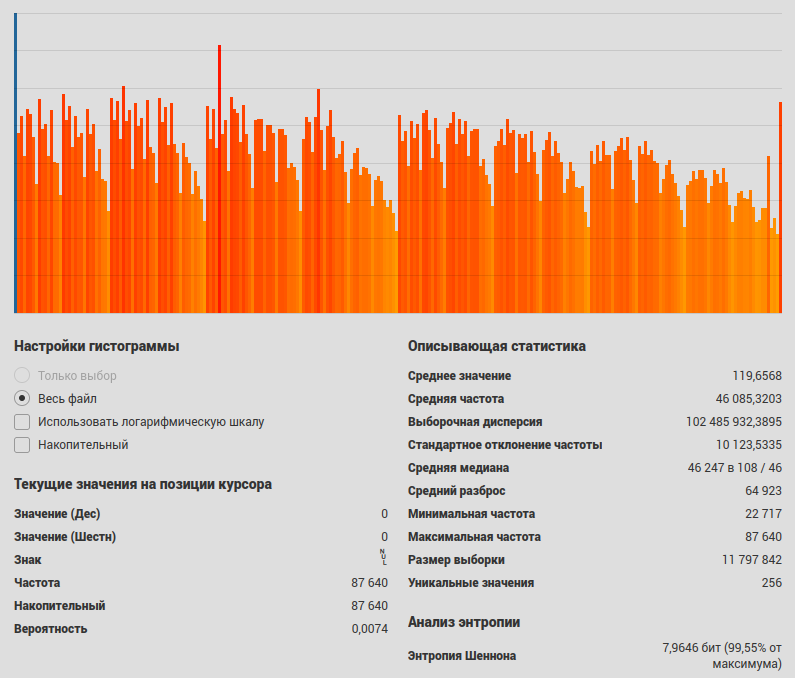
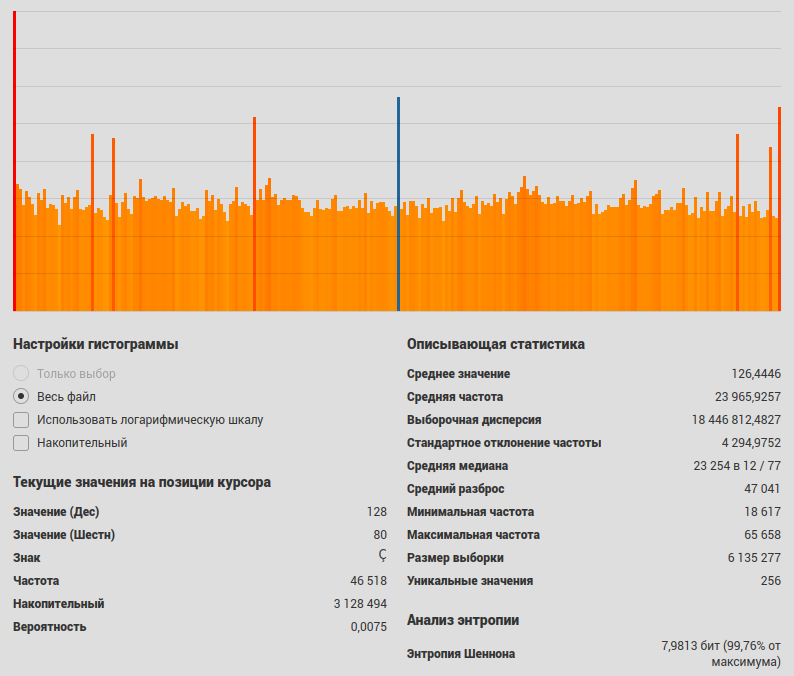
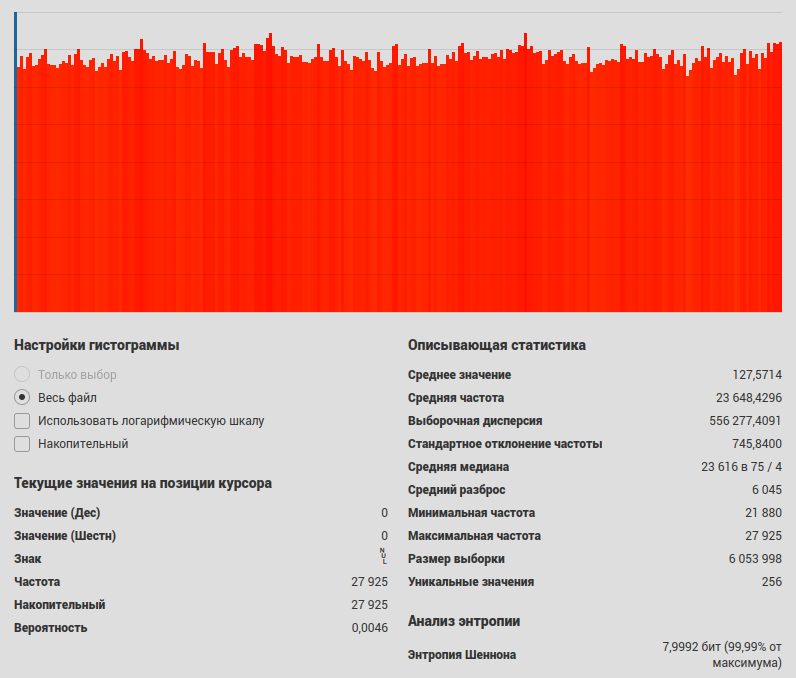
  
Рис 27. Гистограмма и энтропия аудиофайла формата .mp3

  
Рис 28. Гистограмма и энтропия аудиофайла формата .mp3 (zip)

  
Рис 29. Гистограмма и энтропия аудиофайла формата .aac

  
Рис 30. Гистограмма и энтропия аудиофайла формата .aac (zip)

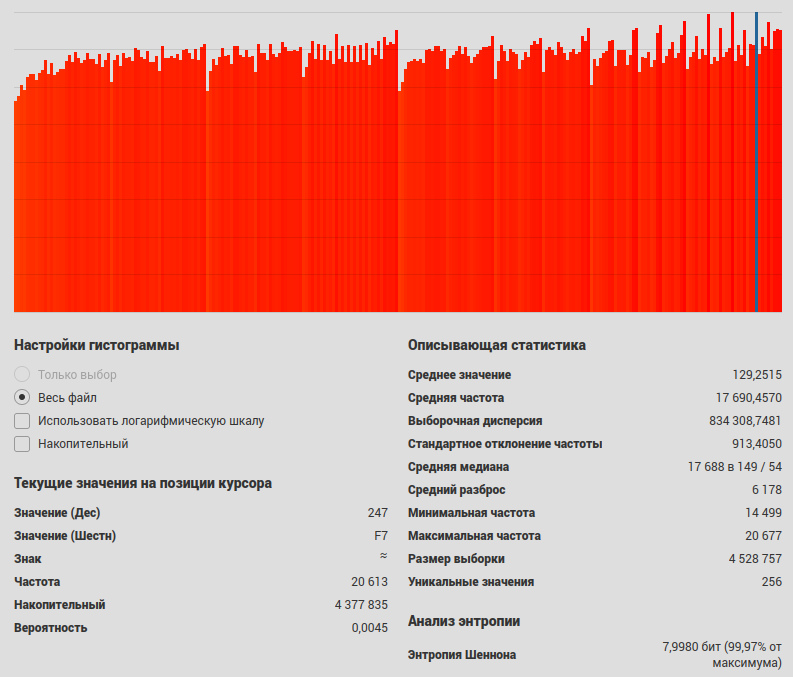
# **Сравнение энтропии файлов изображений разных форматов**

  
Рис 31. Используемое для анализа энтропии изображение

  
Рис 32. Гистограмма и энтропия файла изображения формата .bmp

  
Рис 33. Гистограмма и энтропия файла изображения формата .bmp (zip)

  
Рис 34. Гистограмма и энтропия файла изображения формата .png

  
Рис 35. Гистограмма и энтропия файла изображения формата .png (zip)

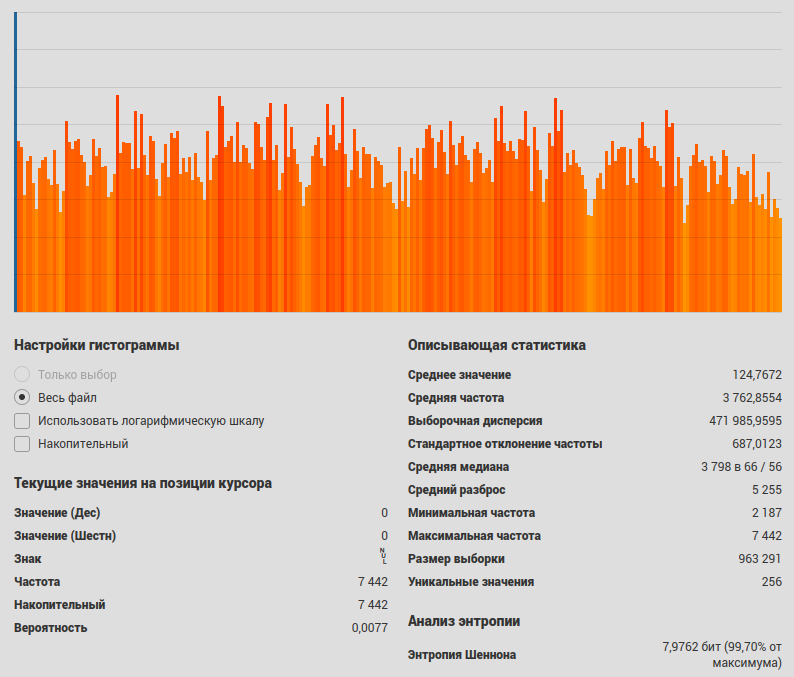
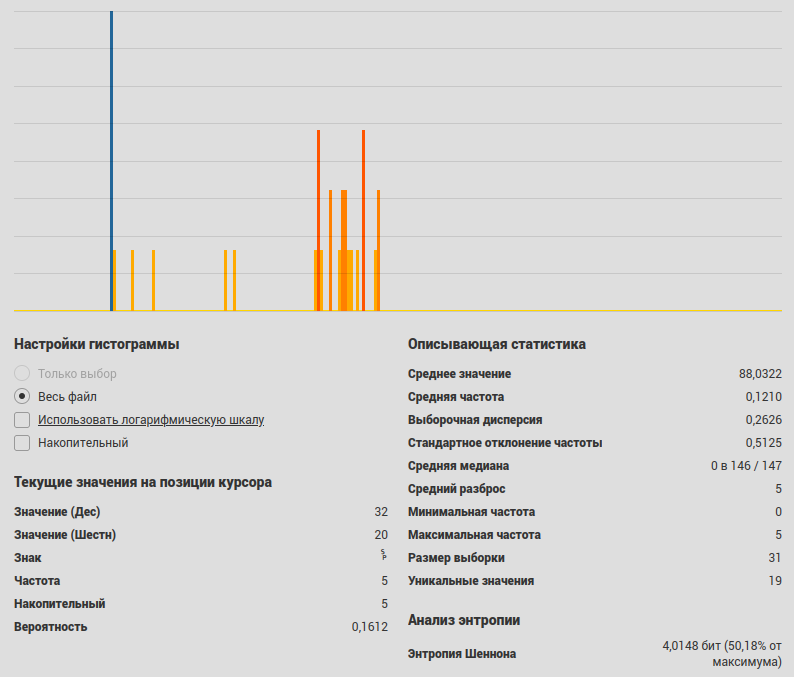
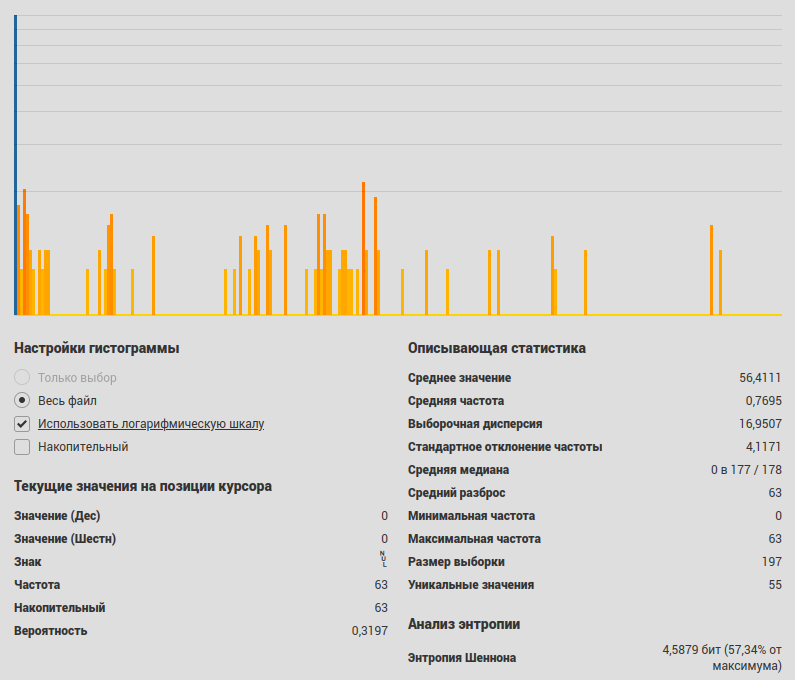
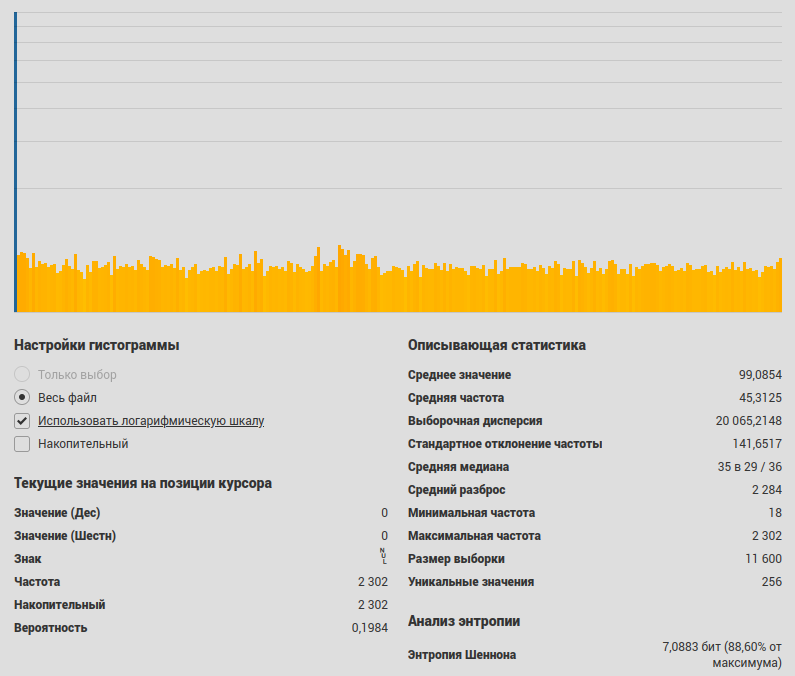
  
Рис 36. Гистограмма и энтропия файла изображения формата .jpeg

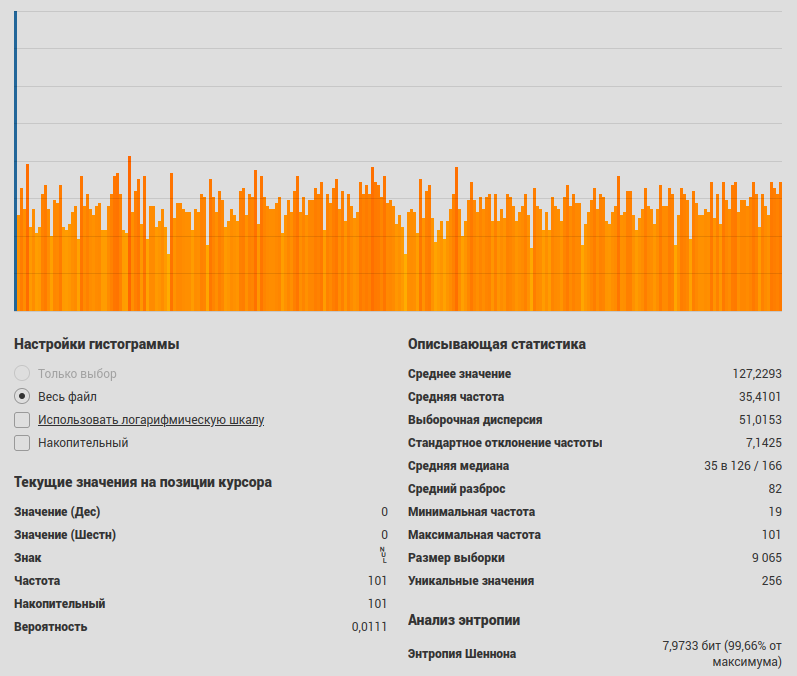
  
Рис 37. Гистограмма и энтропия файла изображения формата .jpeg (zip)

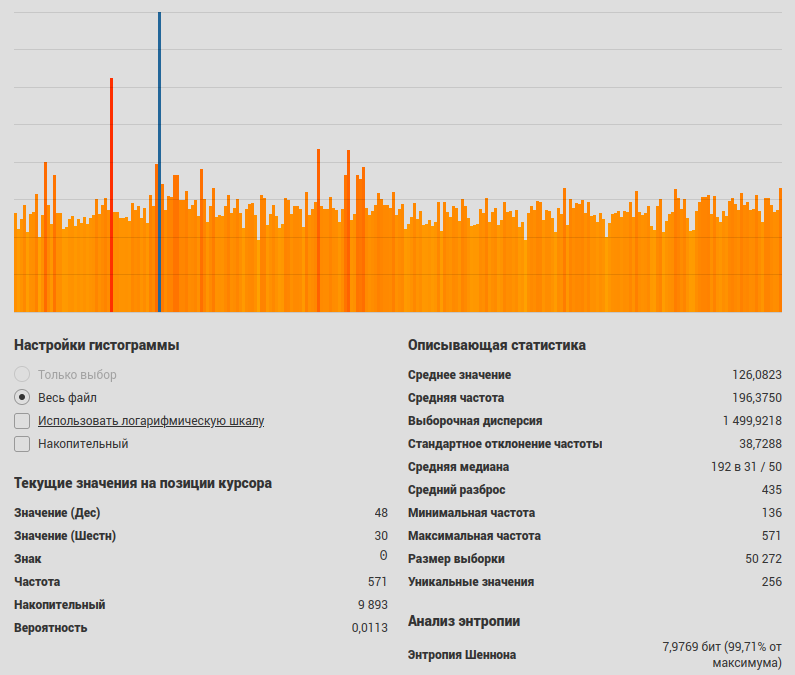
# **Сравнение энтропии текстовых файлов разных форматов**

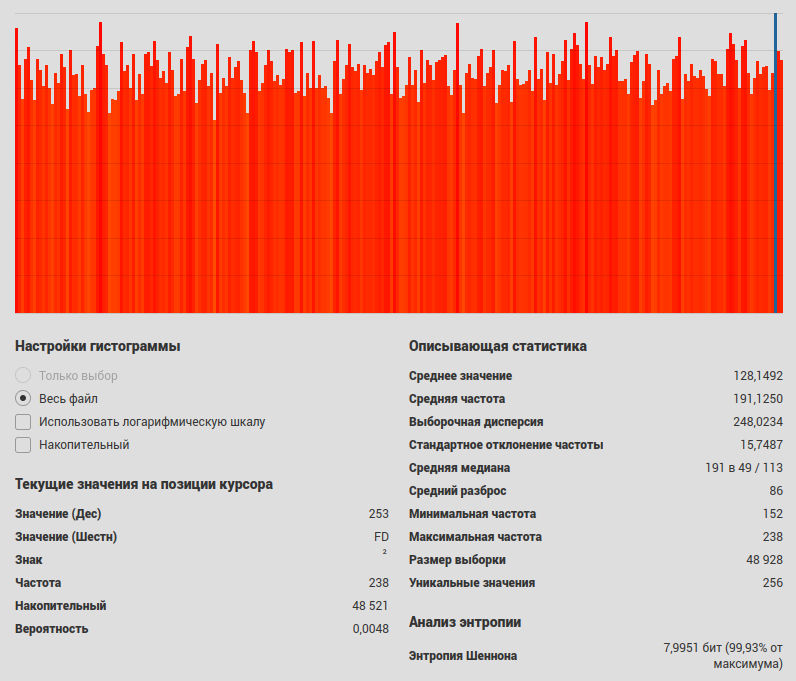
  
Рис 38. Гистограмма и энтропия текстового файла формата .txt

  
Рис 39. Гистограмма и энтропия текстового файла формата .txt (zip)

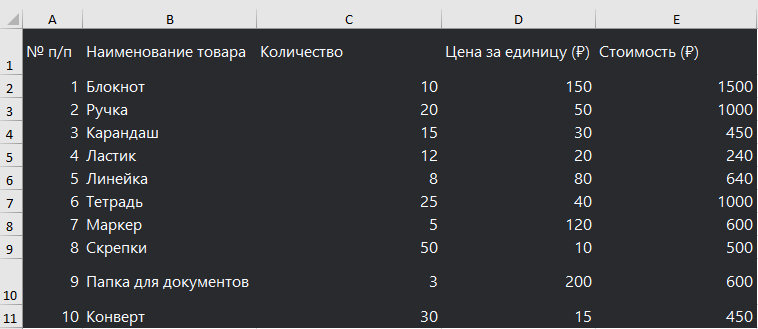
  
Рис 40. Гистограмма и энтропия текстового файла формата .docx

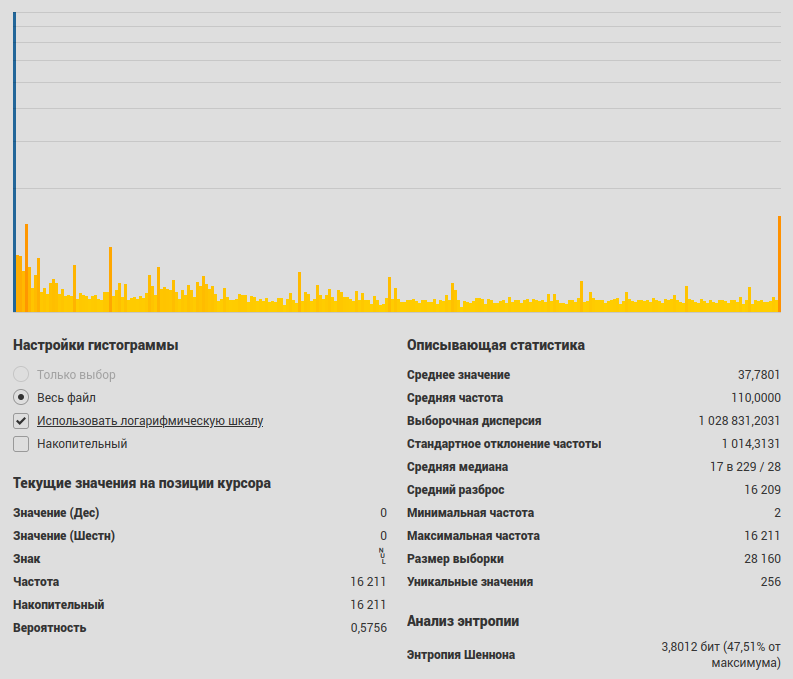
  
Рис 41. Гистограмма и энтропия текстового файла формата .docx (zip)

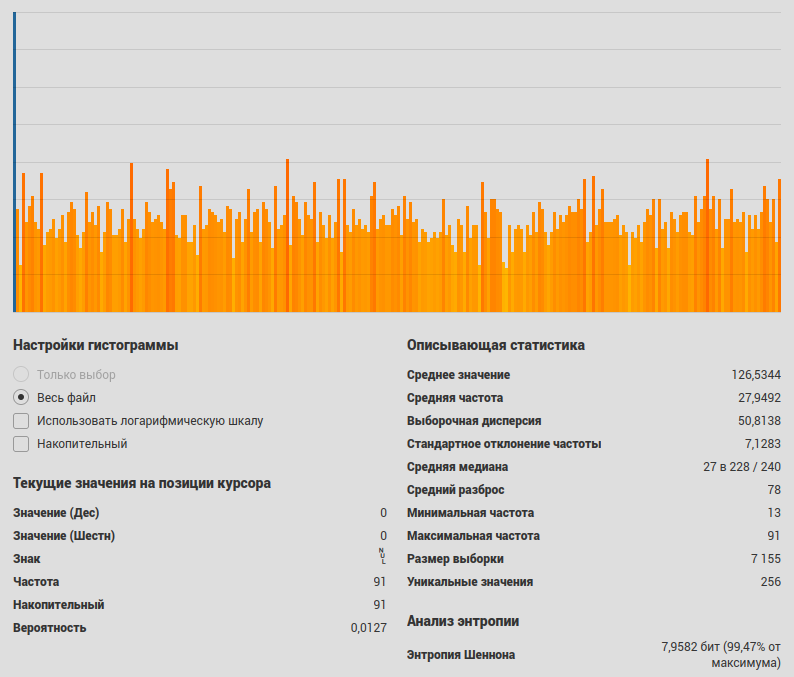
  
Рис 42. Гистограмма и энтропия текстового файла формата .pdf

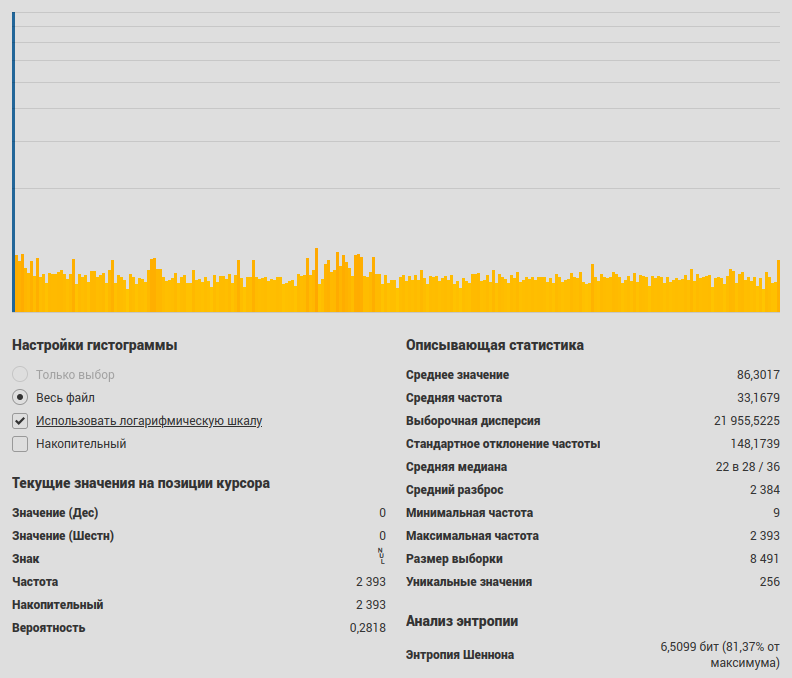
  
Рис 43. Гистограмма и энтропия текстового файла формата .pdf (zip)

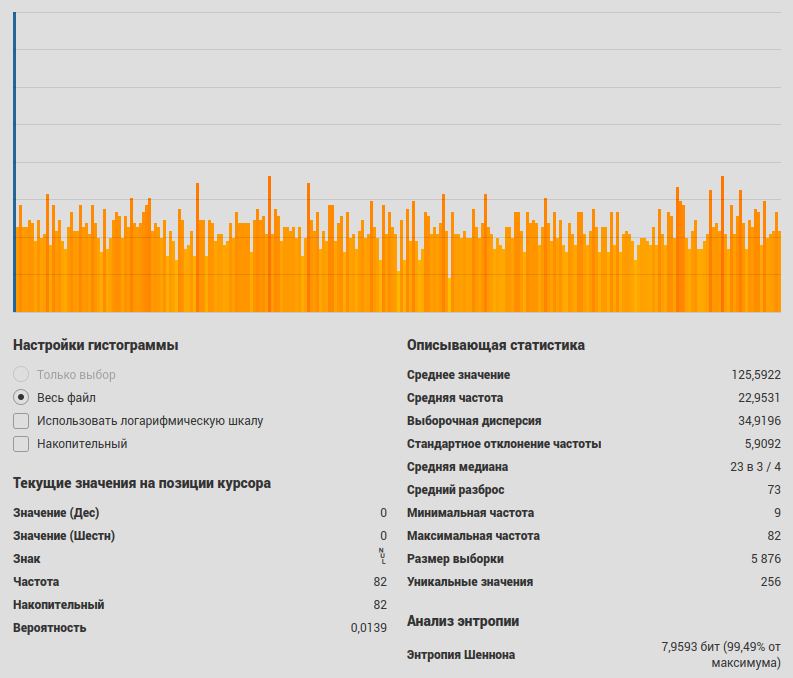
# **Сравнение энтропии файлов таблиц разных форматов**

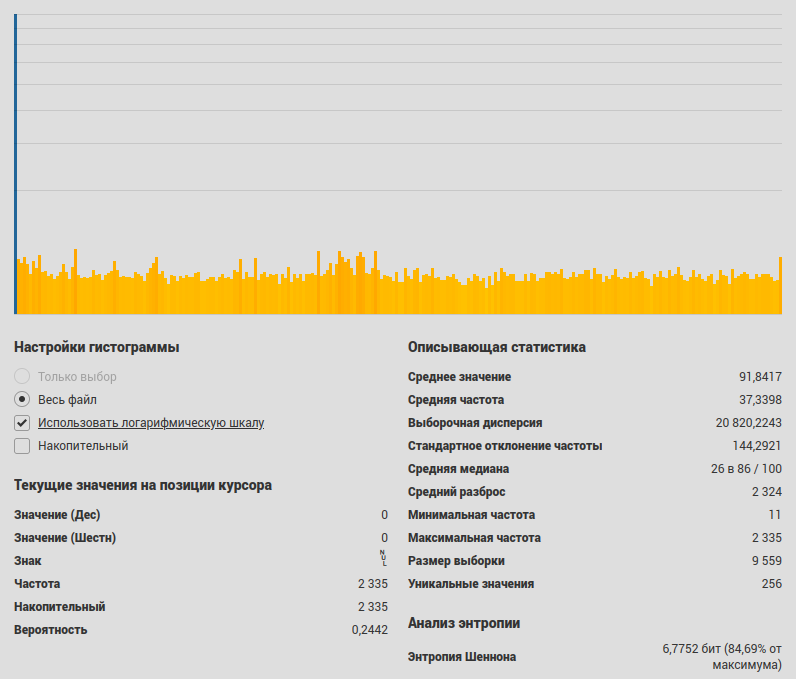
  
Рис 44. Представление используемой таблицы

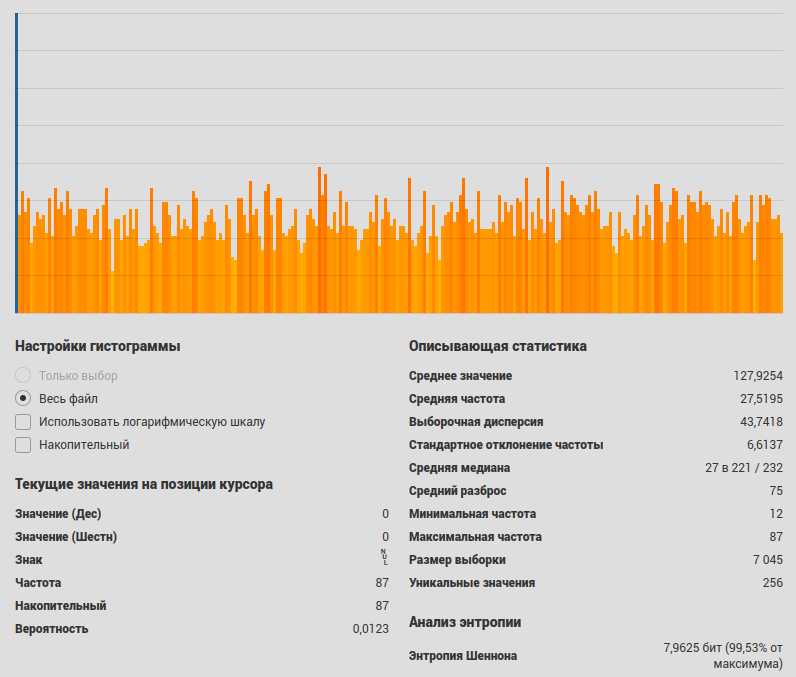
  
Рис 45. Гистограмма и энтропия текстового файла таблицы формата xls

  
Рис 46. Гистограмма и энтропия текстового файла таблицы формата xls (zip)

  
Рис 47. Гистограмма и энтропия текстового файла таблицы формата xlsb

  
Рис 48. Гистограмма и энтропия текстового файла таблицы формата xlsb (zip)

  
Рис 49. Гистограмма и энтропия текстового файла таблицы формата XLSX

  
Рис 50. Гистограмма и энтропия текстового файла таблицы формата XLSX (zip)

# **Таблицы сравнений энтропии**

|  |  |  |
| --- | --- | --- |
| Исполняемые файлы | | |
| Формат | Размер (кб) | Энтропия |
| С++ | 24 | 3.5023 |
| С++ (UPX) | 11 | 7.3247 |
| C++ (GZEXE) | 6 | 7.8709 |
| C++ (упаковщик Перепелкина Н.) | 610 | 6.4889 |
| C# | 65269 | 6.7999 |
| C# (UPX) | 28284 | 7.9267 |
| C# (GZEXE) | 28583 | 7.9969 |
| C# (упаковщик Перепелкина Н.) | 29191 | 7.9977 |
| Pyhon | 7268 | 7.9933 |
| Pyhon (UPX) | 7238 | 7.9938 |
| Pyhon (GZEXE) | 7172 | 7.9980 |
| Pyhon (упаковщик Перепелкина Н.) | 7777 | 7.9758 |

|  |  |  |
| --- | --- | --- |
| Аудиофайлы | | |
| Формат | Размер (мб) | Энтропия |
| wav | 64.458 | 7.8097 |
| wav (zip) | 62.182 | 7.9975 |
| mp3 | 11.522 | 7.9646 |
| mp3 (zip) | 11.485 | 7.9983 |
| aac | 5.992 | 7.9813 |
| aac (zip) | 5.913 | 7.9992 |

|  |  |  |
| --- | --- | --- |
| Файлы изображений | | |
| Формат | Размер (мб) | Энтропия |
| bmp | 11.254 | 7.6335 |
| bmp (zip) | 5.176 | 7.9966 |
| png | 4.427 | 7.9936 |
| png (zip) | 4.423 | 7.9980 |
| jpeg | 0.941 | 7.9762 |
| jpeg (zip) | 0.935 | 7.9923 |

|  |  |  |
| --- | --- | --- |
| Текстовые файлы | | |
| Формат | Размер (кб) | Энтропия |
| txt | 1 | 4.0148 |
| txt (zip) | 1 | 4.5879 |
| docx | 12 | 7.0883 |
| docx (zip) | 9 | 7.9733 |
| pdf | 50 | 7.9769 |
| pdf (zip) | 48 | 7.9951 |

|  |  |  |
| --- | --- | --- |
| Табличные файлы | | |
| Формат | Размер (кб) | Энтропия |
| xls | 28 | 3.8012 |
| xls (zip) | 7 | 7.9582 |
| xlsb | 9 | 6.5099 |
| xlsb (zip) | 6 | 7.9593 |
| XLSX | 10 | 6.7752 |
| XLSX (zip) | 7 | 7.9625 |

# **Вывод**

В ходе работы было исследовано изменение энтропии различных типов файлов, а также влияние упаковки исполняемых файлов на их значение энтропии. Проведённый анализ показал, что энтропия файла во многом зависит от его формата и структуры данных.

* Исполняемые файлы с++ и с# в исходном виде имели средний уровень энтропии, однако после упаковки их энтропия значительно увеличивалась, приближаясь к максимальному значению (около 8 бит на байт), что затрудняет анализ их содержимого. Программа на языке Python изначально имеет высокую энтропию, это означает, что при сборке программы в исполняемый файл уже использовался упаковщик и дальнейшая упаковка не приносит ощутимого прироста энтропии и уменьшения размера файла.
* Изображения, аудиофайлы и другие мультимедийные данные имели более высокую энтропию, особенно если использовали сжатые форматы (JPEG, MP3), что свидетельствует о минимизации избыточности данных. Упаковка же данных файлов ощутимых результатов не приносит.
* Текстовый формат txt продемонстрировал низкую энтропию из-за высокой повторяемости символов и предсказуемости структуры. Однако остальные выбранные форматы показали достаточно высокий уровень энтропии, поэтому можем предположить, что в этих форматах используется упаковка.

Таким образом, энтропийный анализ позволяет не только различать типы файлов, но и выявлять факт сжатия или шифрования данных. Эти знания могут быть полезны в области реверс-инжиниринга, информационной безопасности и анализа вредоносного ПО.