

**Київський національний університет імені Тараса Шевченка  
факультет радіофізики, електроніки та комп'ютерних систем**

**Лабораторна робота №1**

**Тема: «Дослідження кількості інформації при різних варіантах  
кодування»**

Роботу виконав  
студент 3 курсу  
КІ-СА  
Федоренко Олег

Київ 2020

**Мета:** Дослідити імовірнісні параметри української мови для оцінки кількості інформації текстів. Дослідити вплив різних методів кодування інформації на її кількість.

Репозиторій на GitHub:

### Хід роботи

#### **Дослідження кількості інформації в тексті**

1. Оберіть 3 текстових файла різного тематичного та лінгвістичного спрямування.
2. Переконайтесь, що тексти, які ви використовуєте є унікальними і не повторюються у ваших колег! Використовуйте наявні електронні засоби зв'язку та документообігу, щоб уникнути дублювання! Вдруге аналіз того самого тексту не зараховується!

Файли:

1. Створіть програму (будь-якою зручною для вас мовою), яка в якості вхідних даних приймає текстовий файл, та аналізуючи його вміст:
  - a. обраховує частоти (імовірності) появи символів в тексті
  - b. обраховує середню ентропію алфавіту для даного тексту
  - c. виходячи з ентропії визначає кількість інформації та порівнює її з розмірами файлів
  - d. виводить на екран значення частот, ентропії та кількості інформації

Вивід роботи програми:

```

D:\Универ\3 курс\Комп Сиситемы\1 лаба\Олег>python part1.py text1.txt
{' ', 'м', '2', '\n', 'л', 'к', 'п', 'с', 'ш', 'г', 'и', '.', 'к', 'ч', 'м', 'е', 'н', 'ё', 'ы', 'ж', 'б', 'я', 'щ', 'ю',
, 'в', 'с', 'ь', '(', 'б', 'е', '-', 'т', 'о', 'б', 'х', '1', '"', '[', 'а', 'о', '-', 'а', 'ь', 'ф', ' ', 'э', ' ', 'я',
, '4', 'ч', 'г', 'р', 'з', 'ш', 'н', 'л', ' )', '!', 'д', 'з', 'у', '-', 'т', '?', 'й', 'э', 'д', 'и', 'н'}
Total number of symbols in text: 3029
Probability for ]: 0.0036315615714757345
Probability for м: 0.02079894354572466
Probability for 2: 0.0003301419610432486
Probability for \n: 0.030042918454935622
Probability for л: 0.0006602839220864972
Probability for к: 0.034334763948497854
Probability for п: 0.0019808517662594917
Probability for с: 0.03367448002641136
Probability for ш: 0.0036315615714757345
Probability for г: 0.0006602839220864972
Probability for и: 0.002641135688345989
Probability for .: 0.008913832948167713
Probability for к: 0.0049521294156487285
Probability for ч: 0.007923407065037967
Probability for м: 0.0006602839220864972
Probability for е: 0.0009904258831297458
Probability for н: 0.001650709805216243
Probability for ё: 0.005282271376691978
Probability for ы: 0.012545394519643446
Probability for ж: 0.004291845493562232
Probability for б: 0.0003301419610432486
Probability for я: 0.003301419610432486
Probability for щ: 0.0006602839220864972
Probability for ю: 0.006602839220864972
Probability for в: 0.034664905909541105
Probability for с: 0.005282271376691978
Probability for ь: 0.012215252558600198
Probability for (: 0.00957411687025421
Probability for б: 0.0009904258831297458
Probability for е: 0.04027731924727633
Probability for -: 0.006932981181908221
Probability for т: 0.052162429844833276

```

```

Probability for т: 0.052162429844833276
Probability for о: 0.09706173654671509
Probability for б: 0.018157807857378673
Probability for х: 0.006602839220864972
Probability for 1: 0.0003301419610432486
Probability for ": 0.0006602839220864972
Probability for [: 0.0036315615714757345
Probability for а: 0.09409045889732585
Probability for О: 0.0003301419610432486
Probability for -: 0.0013205678441729944
Probability for А: 0.0009904258831297458
Probability for ь: 0.007263123142951469
Probability for ф: 0.0003301419610432486
Probability for ,: 0.013205678441729944
Probability for э: 0.001650709805216243
Probability for : 0.14691317266424564
Probability for я: 0.014196104324859689
Probability for 4: 0.0003301419610432486
Probability for Ч: 0.0003301419610432486
Probability for г: 0.00924397490921096
Probability for р: 0.04522944866292506
Probability for 3: 0.0003301419610432486
Probability for Ш: 0.0003301419610432486
Probability for п: 0.023109937273027403
Probability for л: 0.022449653350940905
Probability for ): 0.00957411687025421
Probability for !: 0.004291845493562232
Probability for д: 0.008583690987124463
Probability for 3: 0.0003301419610432486
Probability for у: 0.012215252558600198
Probability for -: 0.0019808517662594917
Probability for Т: 0.008583690987124463
Probability for ?: 0.00231099372730274
Probability for й: 0.00957411687025421
Probability for э: 0.00924397490921096
Probability for Д: 0.0009904258831297458
Probability for и: 0.027731924727632883

```

```

Probability for и: 0.027731924727632883
Probability for н: 0.043908880818752064
Entropy: 4.815399723383712
Size of file: 3120 bytes
Quantity of information: 1823.230720266158 bytes

```

```
D:\Универ\3 курс\Комп Сиситемы\1 лаба\Олег>python part1.py text2.txt
{'q', 'I', 'x', 'i', 'H', '!', 'L', '"', 'b', 'J', 'd', 'p', 'M', 'c', 'O', '-', '\n', 'S', 'B', 'E', 'u', 'm', 'l', 'a',
, 'k', 'C', 'N', '.', 'e', ';', 'w', 'R', 't', 'g', 'f', 'U', 'D', '?', 's', 'n', 'z', 'o', 'v', 'y', ':', ' ', ' ', ' ', 'n'
, 'G', 'T', 'W', 'A', 'h', 'F'}
Total number of symbols in text: 2702
Probability for q: 0.0014803849000740192
Probability for I: 0.0051813471502590676
Probability for x: 0.0007401924500370096
Probability for i: 0.04922279792746114
Probability for H: 0.003700962250185048
Probability for !: 0.0029607698001480384
Probability for L: 0.0014803849000740192
Probability for ': 0.005921539600296077
Probability for b: 0.00999259807549963
Probability for J: 0.0003700962250185048
Probability for d: 0.028867505551443375
Probability for p: 0.013693560325684678
Probability for M: 0.0011102886750555144
Probability for c: 0.015544041450777202
Probability for O: 0.0055514433752775726
Probability for -: 0.0044411547002220575
Probability for \n: 0.02849740932642487
Probability for S: 0.003700962250185048
Probability for B: 0.0014803849000740192
Probability for E: 0.0022205773501110288
Probability for u: 0.020355292376017766
Probability for m: 0.018134715025906734
Probability for l: 0.02553663952627683
Probability for a: 0.056254626202812734
Probability for k: 0.008512213175425611
Probability for C: 0.0014803849000740192
Probability for N: 0.0022205773501110288
Probability for .: 0.0033308660251665434
Probability for e: 0.08290155440414508
Probability for ;: 0.0029607698001480384
Probability for w: 0.01221317542561066
Probability for R: 0.0048112509252405625
```

```
Probability for t: 0.076980014803849
Probability for g: 0.01147298297557365
Probability for f: 0.017024426350851222
Probability for U: 0.0014803849000740192
Probability for D: 0.0029607698001480384
Probability for ?: 0.0003700962250185048
Probability for s: 0.054404145077720206
Probability for n: 0.04145077720207254
Probability for z: 0.0003700962250185048
Probability for o: 0.06254626202812731
Probability for v: 0.0055514433752775726
Probability for y: 0.013323464100666173
Probability for |: 0.0029607698001480384
Probability for ,: 0.15581051073279054
Probability for ,: 0.019615099925980754
Probability for r: 0.04330125832716506
Probability for G: 0.0007401924500370096
Probability for T: 0.0055514433752775726
Probability for W: 0.0029607698001480384
Probability for A: 0.007031828275351592
Probability for h: 0.046632124352331605
Probability for F: 0.0025906735751295338
Entropy: 4.628005929989433
Size of file: 2779 bytes
Quantity of information: 1563.109002853931 bytes
```

```

D:\Универ\3 курс\Комп Системы\1 лаба\Олег>python part1.py text3.txt
{'i', 'g', '"', 'F', '"', 'X', 'k', 'c', 'a', 'n', 'h', '-', 'S', '2', '.', 'G', 'f', 'l', 'r', 'v', 'B', 'N', 'y', 't',
'Q', 'p', '(', 'e', 'R', 'j', 'D', '4', 'C', 'q', 'x', 'W', ')', '0', '/', 'o', 'T', 'E', '=', 'b', 'I', 'A', 'O', 'U',
'm', ' ', 'c', '"', 'H', 'd', 'u', ' ', ' ', 'w', '\n', 'Y', 'l', 's', 'L'}
Total number of symbols in text: 5300
Probability for i: 0.06
Probability for g: 0.016226415094339624
Probability for ": 0.009433962264150943
Probability for F: 0.0030188679245283017
Probability for ": 0.009433962264150943
Probability for X: 0.0003773584905660377
Probability for k: 0.0033962264150943396
Probability for c: 0.0003773584905660377
Probability for a: 0.0530188679245283
Probability for n: 0.04226415094339623
Probability for h: 0.030754716981132076
Probability for -: 0.005094339622641509
Probability for S: 0.007735849056603773
Probability for 2: 0.00018867924528301886
Probability for .: 0.008490566037735849
Probability for G: 0.0011320754716981133
Probability for f: 0.017924528301886792
Probability for l: 0.04188679245283019
Probability for r: 0.026415094339622643
Probability for v: 0.00830188679245283
Probability for B: 0.0007547169811320754
Probability for N: 0.0047169811320754715
Probability for y: 0.005660377358490566
Probability for t: 0.0809433962264151
Probability for Q: 0.004528301886792453
Probability for p: 0.026415094339622643
Probability for (: 0.0026415094339622643
Probability for e: 0.08452830188679246
Probability for R: 0.006603773584905661
Probability for j: 0.00018867924528301886
Probability for D: 0.0022641509433962265

```

```

Probability for 4: 0.00018867924528301886
Probability for C: 0.00018867924528301886
Probability for q: 0.0007547169811320754
Probability for x: 0.0003773584905660377
Probability for W: 0.0003773584905660377
Probability for ): 0.0026415094339622643
Probability for 0: 0.0030188679245283017
Probability for /: 0.0007547169811320754
Probability for o: 0.0509433962264151
Probability for T: 0.004339622641509434
Probability for E: 0.0005660377358490566
Probability for =: 0.0024528301886792454
Probability for b: 0.012075471698113207
Probability for I: 0.0020754716981132076
Probability for A: 0.0033962264150943396
Probability for O: 0.0013207547169811322
Probability for U: 0.00018867924528301886
Probability for m: 0.01
Probability for : 0.1830188679245283
Probability for c: 0.030943396226415093
Probability for ': 0.0007547169811320754
Probability for H: 0.0013207547169811322
Probability for d: 0.017735849056603775
Probability for u: 0.02320754716981132
Probability for ,: 0.008867924528301888
Probability for w: 0.010566037735849057
Probability for \n: 0.007924528301886792
Probability for Y: 0.0005660377358490566
Probability for l: 0.0035849056603773585
Probability for s: 0.04981132075471698
Probability for L: 0.0013207547169811322
Entropy: 4.5684756191714895
Size of file: 5342 bytes
Quantity of information: 3026.6150977011116 bytes

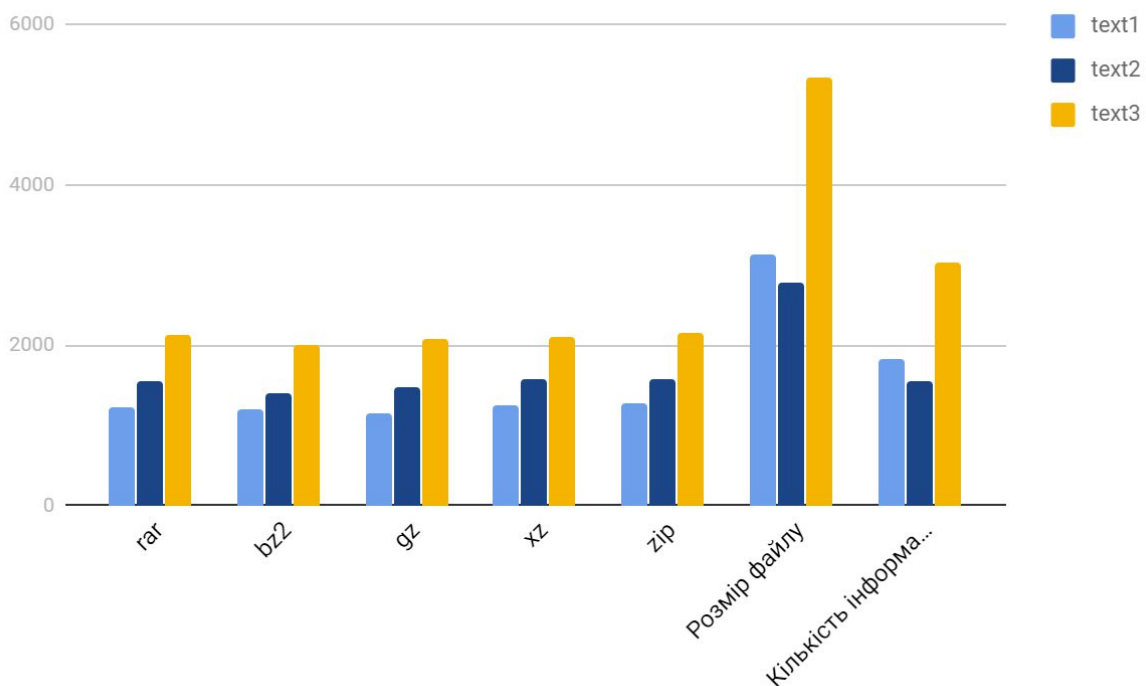
```

4. Проведіть стиснення кожного вхідного файлу за допомогою 5 різних алгоритмів стиснення (zip, rar, gzip, bzip2, xz, або будь-які інші на ваш вибір, можна використовувати готові програмні засоби для стиснення).



5. Порівняйте результуючі обсяги архівів з обчисленою кількістю інформації та **наведіть у звіті висновки** щодо кореляції цих величин для обраних вами файлів (яка відмінність, що вийшло більше і чому)

Файл	text1.txt, bytes	text2.txt, bytes	text3.txt, bytes
rar	1233	1561	2136
bz2	1208	1399	2012
gz	1155	1470	2068
xz	1248	1572	2108
zip	1267	1567	2157
Розмір без стиснення	3120	2779	5342
Кількість інформації	1823	1563	3026



За умови ідеального стиснення розмір файлу повинен бути рівним кількості інформації. Але у реальності розміри архівованих файлів у більшості випадків відрізняються від кількості інформації: для великих текстів кількість інформації більша, ніж розмір стисненого файлу, проте для невеликого тексту ситуація протилежна.

Це можна пояснити тим, що алгоритми для стиснення використовують повторювані частини тексту, а у великому тексті таких частин повинно бути більше.

### **Дослідження способів кодування інформації на прикладі Base64**

1. Ознайомтесь зі стандартом [RFC4648](#)
2. Для практичного засвоєння методу кодування, створіть програму, що кодує довільний файл в Base64 (шляхом реалізації алгоритму вручну, а не виклику бібліотечної функції)
3. Закодуйте в Base64 обрані вами текстові файли
  - a. Обрахуйте кількість інформації в base64-закодованому варіанті файлу
  - b. Порівняйте отримане значення з кількістю інформації вихідного файлу
  - c. Зробіть висновки з отриманого результату
4. Закодуйте в Base64 стиснені кращим з алгоритмів текстові файли
  - a. Обрахуйте кількість інформації в base64-закодованому варіанті стисненого файлу
  - b. Порівняйте отримане значення з кількістю інформації вихідного файлу та base64-закодованого файлу.
  - c. Зробіть висновки з отриманого результату

## Text1:

```
D:\Универ\3 курс\Комп Сиситемы\1 лаба\Олег>python part2.py text1.txt
File Name: text1.txt
```

Source text:

[Текст песни "Собака съела товар"]

[Интро]

- Братья и сёстры, вы готовы услышать историю про собаку?!

- Конечно!

- Готовы?!

- Да!

- Историю про собаку, которая съела товар!

[Припев]

Собака съела товар. Теперь, она — наркоман

Товар ей понравился точно. Я знаю много таких собак (да) Собака съела товар (это бесподобно). Теперь, она — наркома (гав!)

Товар ей понравился точно. Я знаю много таких собак (ещё бы!)

Собака съела товар. Теперь, она — наркоман (ну)

...

Encoded text:

W4RQ1h1EGIRBD+GsQYew4ICKEMPoYwwh1DAgiDEqGsO4YEEQofQyhHEAiXQoKW4M  
Q9iFECHy6FQCZBBGIEhRMieQQ04IIgxRiDEKIESywh1EsghnD6IUPoZRLIIhxB  
h3EuJEMiHrMIiCRBiFD6IEOIInEEP4gQ+IIgw+hjDCHUQz8hCoBMgg1D6HsNYjw9h  
8QhUAmQQTh9EKH0Mo1n5CFQCZBBSGBCFQCZBBiIMQofRAHxE4gh/ECHxBEGH0MY  
Q6iGWEE0oFRCh9ECGETyCIMSow7hgQRCh9DKGEQCEKCluD8QIcQ/hrDJdCoQw+h  
jDCHUMCCIMSow7hgQRCh9DKGEQC4ghFDwH8NYgRMLCCH0PYEEAoQQ9hhECHUPo  
eQwh6FQih9DKGEQCCGsOSCH8PoexAhhDKHEO4gxPIIhQ+iPD2HxcQQvIIbw9hhE4  
gh5D2H0M4FEEQoYQ6hxEUgiDD6GMMIdEBRDSGBSQQhh9DGGE0oYEEQY1Q1h3DAgi  
FD6GUMIGEBRE2IUPiCGMNYgw/h9DSH0MYew+KS4ghFDwH8NYgRMLCCH0PYEEAoQ  
Q9hhECHUPoeQwICiGcMIZEJSFQih9DKGEQCCGsOSCH8PoexAhhDKHEO4gxPIIhQ+  
iPD2HxcQQvIIbw9hhE4gh5D2H0M4FEEQoYQ6hxEUgiDD6GMMIdEBRDWJMUSCGMSy  
EpCoQw+hjDCHUMCCIMSow7hgQRCh9DKGEQC4ghFDwH8NYgRMLCCH0PYEEAoQQ9  
hhECHUPoeQwh6QFEPYh1IVCKH0MoYRAIIaw5IIIfw+h7ECGEMocQ7iDE8giFD6I8P  
YFFxBc8ghvD2GETiCHkPYfQzh8QRChDqHERSCIMPoYwwh0QFENIYFIVCGH0MYQ  
6hgQRBiVDWHcMCCIUPoZQwiAXEEIoaw/hrECJhYQQ+h7DaggFCCHsMIgQ6h9DyGE  
PQqEUPoZQwiAQQ1hyQQ/h9D2IEMIZQ4h3EGJ5BEKH0R4ew+LiCF5BDeHsMIInEPI  
ew+hnD4giFDCHUOIikEQYfQxhhDogKIhQ+iPD2HxSFBS3BqIcP4dw1iEQGK6FQhh  
9DGGE0oYEEQY1Q1h3DAgiFD6GUMIGEBRDCGcMCwghgQRHiiQRFh9ECH0SiCQ5IIh  
Q+h1DCIB+UhuFIYFhBE2IUPiCGMS4dkERYfRAH9EiHEOSCIUPoZQwiAQFEMYfQ7i  
ZEiHEOSCIUPoZQwiAU1hBDKGsO4cQ6h9DuGsP4exLhyQRCh9DKGEQCAohjD6HcTI  
kQ4hyQRCh9DKGEQCKg9EeGsPYmEEPYhw2h7EuHJBEKH0MoYRAICiHkNIYFJYQQ6  
h9EKH0QI1w5IIexDhtD2HxBD+IEPoaQwiFEwKgwQQ6iHD+HEQomEEQYawxhqQRHi  
FD4tiFD4giFDCHUPoalhbEeIUPiCHkPobQ9h8QQ/iBD6GkMIhRMCoNQ+hnDSGBBD  
+IEPoaQwh6QRCh9DKGEQAqEMPoYwwh1DAgiDEqGsO4YEEQofQyhHEAggFCCGkPog



## Ентропія text1.txt.encoded:

```
D:\Універ\3 курс\Комп Сиситемы\1 лаба\Олер>python part1.py text1.txt.encoded
{'4', 'R', 'S', 'w', 'Z', 'U', 'q', 'k', '8', 'Y', 'F', 'M', '1', 'O', '3', 'm', 'h', 'I', 'e', 'j', '/', 'd', 'E', '9', 'L', '1', 'Q', '=', 'z', 'b', 'T', 's', 'J', '5', '7', 'G', 'D', '6', 'x', 'V', 'y', 'A', 'N', 'o', 'W', 't', 'p', '2', '+', 'H', '0', 'C', 'a', 'g', 'v', 'n', 'n', 'B', 'r', 'f', 'i', 'c', 'K', 'P', 'u', 'X'}
Total number of symbols in text: 5293
Probability for 4: 0.018515019837521254
Probability for R: 0.02210466654071415
Probability for S: 0.009446438692612885
Probability for w: 0.028528244851690913
Probability for Z: 0.006423578310976762
Probability for U: 0.017759304742112222
Probability for q: 0.0030228603816361234
Probability for k: 0.010202153788021915
Probability for 8: 0.009257509918760627
Probability for Y: 0.025883242017759305
Probability for F: 0.015492159455885132
Probability for M: 0.02380502550538447
Probability for 1: 0.006045720763272247
Probability for O: 0.013413942943510296
Probability for 3: 0.0030228603816361234
Probability for m: 0.0015114301908180617
Probability for h: 0.06461364065747213
Probability for I: 0.05705648970338183
Probability for e: 0.011335726431135461
Probability for j: 0.0035896467031928964
Probability for /: 0.004534290572454185
Probability for d: 0.005101076894010958
Probability for E: 0.06234649537124504
Probability for 9: 0.013791800491214812
Probability for L: 0.004534290572454185
Probability for l: 0.006990364632533535
```

```
Probability for N: 0.0049121481201587006
Probability for o: 0.028339316077838655
Probability for W: 0.005856791989419989
Probability for t: 0.0015114301908180617
Probability for p: 0.0017003589646703192
Probability for 2: 0.004534290572454185
Probability for +: 0.01076894010957869
Probability for H: 0.02796145853013414
Probability for 0: 0.008690723597203855
Probability for C: 0.045531834498394104
Probability for a: 0.006234649537124504
Probability for g: 0.032306820328736066
Probability for v: 0.001322501416965804
Probability for \n: 0.015303230682032874
Probability for n: 0.0041564330247496695
Probability for B: 0.023427167957679955
Probability for r: 0.0035896467031928964
Probability for f: 0.009824296240317401
Probability for i: 0.029850746268656716
Probability for c: 0.009824296240317401
Probability for K: 0.017003589646703194
Probability for P: 0.022671452862270923
Probability for u: 0.0041564330247496695
Probability for X: 0.001889287738522577
Entropy: 5.377462999548412
Size of file: 5374 bytes
Quantity of information: 3557.863957076218 bytes
```

## Text2:

```
D:\Универ\3 курс\Комп Системы\1 лаба\Олег>python part2.py text2.txt
File Name: text2.txt
```

Source text:

```

    HORATIO
That can I;
At least, the whisper goes so. Our last king,
Whose image even but now appear'd to us,
Was, as you know, by Fortinbras of Norway,
Thereto prick'd on by a most emulate pride,
Dared to the combat; in which our valiant Hamlet--
For so this side of our known world esteem'd him--
Did slay this Fortinbras; who by a seal'd compact,
Well ratified by law and heraldry,
Did forfeit, with his life, all those his lands
Which he stood seized of, to the conqueror:
Against the which, a moiety competent
Was gaged by our king; which had return'd
To the inheritance of Fortinbras,
Had he been vanquisher; as, by the same covenant,
And carriage of the article design'd,
```

Encoded text:

```
SE9SQVRJTwpUaGF0IGNhbiBJOwpBdCBsZWZdCwgdGh1IHdoaXNwZXIgz29lcyBz
by4gT3VyIGxhc3Qga2luZywkV2hvc2UgaW1hZ2UgZXZlbiBidXQgbm93IGFwcGVh
cidkIHRvIHVzLApXYXMsIGFzIHlvdSBrbm93LCBieSBGb3J0aW5icmFzIG9mIE5v
cndheSwKVGH1cmV0byBwcm1jaydkIG9uIGJ5IGEGbw9zdCB1bXVsYXRlIHByaWRl
LApEYXJlZCB0byB0aGUgY29tYmF0OyBpbjB3aGljaCBvdXIgdFsaWFudCBiYW1s
ZXQtLQpGb3Igc28gdGhpcyBzawRlIG9mIG91ciBrbm93biB3b3JsZCB1c3RlZW0n
ZCBoaw0tLQpEawQgc2xheSB0aGlzIEZvcnRpbmJyYXM7IHdobyBieSBhIHNIYWwn
ZCBjb21wYWN0LApXZlxsIHJhdGlmawVkIGJ5IGxhdyBhbmQgaGVyYwXkcncsCkRp
ZCBmb3JmZWl0LCB3aXRoIGhpcyBsaWZlLCBhbGwgdGhvc2UgaGlzIGxhbmRzClDo
aWNoIGhlIHNo29kIHNIaXplZCBvZiwdG8gdGh1IGNvbnF1ZXJvcjoKQWdhaw5z
dCB0aGUgd2hpy2gsIGEGbw9pZXR5IGNvbXBldGVudApXYXMGZ2FnZWQgYnkgb3Vy
IGtpbmc7IHdoaWNoIGhhZCBYXR1cm4nZApUbyB0aGUgaW5oZXJpdGFuY2Ugb2Yg
Rm9ydGluYnJhcycwKSGFkIGh1IGJlZW4gdmFucXVpc2hlcjsjYXMsIGJ5IHRoZSBz
YW1lIGNvdmVuYW50LApBbmQgY2FycmlhZ2Ugb2YgdGh1IGFydG1jbGUgZGVzaWdu
J2QsCkhpcyBmZWxsIHRvIEhhdGh1dC4gTm93LCBzaXIsIHlvdW5nIEZvcnRpbmJy
YXMsCk9mIHVuaW1wcm92ZWQgbW0dGx1IGhvdCBhbmQgZnVsbcwKSGF0aCBpbjB0
aGUgc2tpcnRzIG9mIE5vcndheSB0ZXJlIGFuZCB0aGVyZQpTaGFyaydkIHVwIGEG
bGlzdCBvZiBsYXdsZXNzIHJlc29sdXRlcycwKRm9yIGZvb2QgYW5kIGRpZXQsIHRv
IHNVbWUgZW50ZXJwcm1zZQpUaGF0IGhhdGggYSBzdG9tYwNoIGluJ3Q7IHdoaWNo
IGlzIG5vIG90aGVyLS0KQXMGaXQgZG90aCB3ZlxsIGFwcGVhciB1bnRvIG91ciBz
dGF0ZS0tckJ1dCB0byByZWVndmVyIG9mIHVzLCBieSBzdHJvbmcgaGFuZApBbmQg
dGVyYXMGY29tcHVsc2F0b3J5LCB0aG9zZSBmb3Jlc2FpZCBsYW5kcwpTbyBieSB0
aXMGZmF0aGVyIGxvc3Q6IGFuZCB0aGlzLCBjIHRha2UgaXQsCk1zIHRoZSBtYWlu
IG1vdG12ZSBvZiBvdXIgcHJlcGFyYXRpb25zLApUaGUgc291cmNlIG9mIHRoaXMG
b3VyIHdhdGNoIGFuZCB0aGUgY2hpbWYgaGVhZApPZiB0aGlzIHVvc3QtaGFzdGUg
YW5kIHJvbWFnZSBpbjB0aGUgbGFuZC4KQkVSTkFSRE8KSSB0aGluayBpdCBiZSBu
```

Энтропія text2.txt.encoded:

```

D:\Универ\3 курс\Комп Системы\1 лаба\Олег>python part1.py text2.txt.encoded
{'W', 'G', 'd', '4', '8', 'Y', '5', 'F', 'R', 'x', 'Z', 'C', '3', 'L', 'b', 'g', 'P', 'u', 'U', 'M', '=', 'c', 'k', 'Q', 'h',
'K', '9', 'T', 'i', 'l', 'A', 'e', 'z', 'a', 'n', '0', 'B', 'o', 'y', 'O', '\n', 't', 'X', 'D', '1', 'm', 'E', 'S', 'I', 'r',
'7', 'p', 'j', 'w', 's', 'v', '6', '2', 'N', 'V', 'J', 'H'}
Total number of symbols in text: 3660
Probability for W: 0.027868852459016394
Probability for G: 0.051639344262295085
Probability for d: 0.03224043715846994
Probability for 4: 0.003278688524590164
Probability for 8: 0.002459016393442623
Probability for Y: 0.01994535519125683
Probability for 5: 0.010109289617486339
Probability for F: 0.016939890710382512
Probability for R: 0.020218579234972677
Probability for x: 0.006830601092896175
Probability for Z: 0.04316939890710383
Probability for C: 0.024863387978142075
Probability for 3: 0.015846994535519125
Probability for L: 0.007377049180327869
Probability for b: 0.03524590163934426
Probability for g: 0.03715846994535519
Probability for P: 0.001092896174863388
Probability for u: 0.01092896174863388
Probability for U: 0.015027322404371584
Probability for M: 0.006557377049180328
Probability for =: 0.000273224043715847
Probability for c: 0.027049180327868853

```

```

Probability for t: 0.006284153005464481
Probability for X: 0.025136612021857924
Probability for D: 0.000546448087431694
Probability for l: 0.009016393442622951
Probability for m: 0.019398907103825136
Probability for E: 0.006010928961748634
Probability for S: 0.016666666666666666
Probability for I: 0.040983606557377046
Probability for r: 0.001639344262295082
Probability for 7: 0.001366120218579235
Probability for p: 0.020218579234972677
Probability for j: 0.003825136612021858
Probability for w: 0.011475409836065573
Probability for s: 0.012841530054644808
Probability for v: 0.017759562841530054
Probability for 6: 0.000819672131147541
Probability for 2: 0.02185792349726776
Probability for N: 0.010109289617486339
Probability for V: 0.020765027322404372
Probability for J: 0.015300546448087432
Probability for H: 0.01721311475409836
Entropy: 5.53729781872946
Size of file: 3716 bytes
Quantity of information: 2533.313752068728 bytes

```

Text3:



```
D:\Универ\3 курс\Комп Сиситемы\1 лаба\Олег>python part2.py text3.txt
File Name: text3.txt
Source text:

    A flip flop is an electronic circuit with two stable states that can be used to store binary data.
Flip-flops(electronic)
The stored data can be changed by applying varying inputs.
Flip-flops and latches are fundamental building blocks of digital electronics systems used in computers, communications, and many other types of systems.
Flip-flops and latches are used as data storage elements.
It is the basic storage element in sequential logic. But first, let's clarify the difference between a latch and a flip-flop.

Flip flop v/s Latch
The basic difference between a latch and a flip-flop is a gating or clocking mechanism.

Read the full comparison of Flip Flop v/s latch here

For example, let us talk about SR latch and SR flip-flops.
In this circuit when you Set S as active the output Q would be high and Q' will be low.
```

...

```
Encoded text:

QSBmbG1wIGZsb3AgaxMgYw4gZWx1Y3Ryb25pYyBjaXJjdWl0IHdpdGggdHdvIHNO
YWJsZSBzdGF0ZXMgdGhhdBjYw4gYmUgdXNlZCB0byBzdG9yZSBiaw5hcnkgZGF0
YS4gCkZsaXAtZmxvcHMozWx1Y3Ryb25pYyKvGh1IHNOB3JlZCBkYXRhIGNhbiBi
ZSBjaGFuZ2ZVIGJ5IGFwcGx5aW5nIHZhcmlpbmcgaW5wdXRzLiAKRmxpcC1mbG9w
cyBhbmQgbGF0Y2hlcYBhcmUgZnVuzGFtZW50YWwgYnVpbGRpbmcgYmxvY2tzIG9m
IGRpZ2l0YWwgZWx1Y3Ryb25pY3Mgc3lzdGVtcyB1c2VkIGluIGNvbXB1dGVycywg
Y29tbXVuaW50bG1vbnMsIGFuZCBtYW55IG90aGVyIHR5cGVzIG9mIHNSc3R1bXMu
IApGbG1wLWZsb3BzIGFuZCBsYXRjaGVzIGFyZSB1c2VkIGFzIGRhGEgc3RvcFn
ZSB1bGVtZW50cy4gCk10IGlzIHROZSBiYXNpYyBzdG9yYwdlIGVsZWl1bnQgalW4g
c2VxdWVudG1hbCBsb2dpYy4gQnV0IGZpcnN0LCBsZXSAZcyBjbGFyaWZ5IHRoZSB
kaWZmZXJlbnNlIGJldHd1ZW4gYSBsYXRjaCBhbmQgYSBmbG1wLWZsb3AuCgpGbG1
wIGZsb3Agdi9zIExhdGNoClRoZSBiYXNpYyBkaWZmZXJlbnNlIGJldHd1ZW4gYSB
sYXRjaCBhbmQgYSBmbG1wLWZsb3AgaxMgYSBnYXRpbmcgb3IyY2xvY2tpbmcbWV
jaGFuaXNlLgoKUmVhZCB0aGUgZnVsbcBjY21wYXJpc29uIG9mIEZsaXAgRmxvcCB
Zl3MgbGF0Y2ggaGVyZQoKRm9yIGV4YW1wbGUsIGx1dCB1cyB0YXNrIGF1b3V0IFN
SIGxhdGNoIGFuZCBTUibmbG1wLWZsb3BzLiAKSW4gdGhpcyBjaXJjdWl0IHdoZW4
geW91IFNldCBTIGFzIGFjdG12ZSB0aGUgb3V0cHV0IFEGd291bGQgYmUgaGlnaCB
hbmQgYUyBkgd2lscCBiZSBsb3cuIApUaGlzIGlzIGlycmVzcGVjdG12ZSBvZiBhbn
l0aGluZyB1bHNlLiAoVGhpcyBpcyBhbiBhY3RpdmtbG93IGNpcnN1aXQgc28gYW
N0aXZlIGh1cmUgbWVhbnMgbG93LCBldXQgZm9yIGFuIGFjdG12ZSB0aWdoIGNpcn
N1aXQgYWN0aXZlIHdvdkxkIG11YW4gaGlnaCkKSW4gdGhpcyBjaXJjdWl0IGRpYW
dyYW0sIHRoZSBvdXRwdXQgaXMgY2hhbmdlZCAoaS51LiB0aGUgc3RvcnVkaGRhdG
EgaXMgY2hhbmdlZCkgb25seSB3aGVuIHlvdSBnaXZlIGFuIGFjdG12ZSBjbG9jay
BzaWduYWwueIF90aGVyZ2l2ZSwgZXZlbiBpZiB0aGUgUyBvciB5IGlzIGFjdG12ZS
B0aGUgZGF0YSB3aWxsIG5vdCBjaGFuZ2UuIExldIBlzigxvb2sgYXQgdGh1IHR5c
GVzIG9mIGZsaXAtZmxvcHMgdG8gdW5kZXJzdGFuZCBiZXR0ZXIuCgpTUibG1wI
EZsb3AKVGHlcmUgYXJlIG1ham9ybHkgbCB0eXB1cyBvZiBmbG1wIGZsb3BzLCB3a
XR0IHROZSBtb3N0IGNvbW1vbiBvbmUgYmVpbmcgU1IgzMxpcCBmbG9wLiAKVGHpc
```

## Ентропія text2.txt.encoded:

```
D:\Универ\3 курс\Комп Сиситемы\1 лаба\Олег>python part1.py text3.txt.encoded
{'c', 'Y', 'p', 'a', '4', 'K', 'e', 'J', 'i', 'T', '\n', 'C', 'o', 'r', 'S', 'B', 'U', 'b', 'z', '=', 'H', 'k', 'm', 'l', 'L',
'D', 'X', '5', '3', 'x', 'w', 'R', 'u', 'E', 'n', 's', 'A', 'N', 'Q', 'W', 'j', 'F', '8', 'Z', 'G', '9', '1', 'V', 't', '2',
'h', 'I', 'g', 'O', 'M', '0', 'v', 'P', 'd', 'y'}
Total number of symbols in text: 7286
Probability for c: 0.028685149601976392
Probability for Y: 0.026489157287949493
Probability for p: 0.0168816909140818
Probability for a: 0.022646170738402416
Probability for 4: 0.004666483667307164
Probability for K: 0.0028822399121603073
Probability for e: 0.0008234971177600879
Probability for J: 0.008646719736480922
Probability for i: 0.010156464452374416
Probability for T: 0.0037057370299203953
Probability for \n: 0.015371946198188307
Probability for C: 0.021273675542135603
Probability for o: 0.007960472138347515
Probability for r: 0.0006862475981334065
Probability for S: 0.01894043370848202
Probability for B: 0.05160581937963217
Probability for U: 0.014960197639308262
Probability for b: 0.03692012077957727
Probability for z: 0.012901454844908043
Probability for =: 0.0001372495196266813
```

```

Probability for E: 0.00631347790282734
Probability for n: 0.008234971177600879
Probability for s: 0.012352456766401317
Probability for A: 0.00947021685424101
Probability for N: 0.012489706286028
Probability for Q: 0.009195717814987648
Probability for W: 0.024430414493549272
Probability for j: 0.00837222069722756
Probability for F: 0.0168816909140818
Probability for 8: 0.001784243755146857
Probability for Z: 0.03829261597584408
Probability for G: 0.05764479824320615
Probability for 9: 0.011803458687894593
Probability for 1: 0.01056821301125446
Probability for V: 0.019626681306615427
Probability for t: 0.0038429865495470767
Probability for 2: 0.020450178424375513
Probability for h: 0.0190776832281087
Probability for I: 0.05119407082075213
Probability for g: 0.0506450727422454
Probability for O: 0.0012352456766401318
Probability for M: 0.009744715893494373
Probability for 0: 0.022920669777655778
Probability for v: 0.012077957727147955
Probability for P: 0.0009607466373867692
Probability for d: 0.03842986549547077
Probability for y: 0.017705188031841887
Entropy: 5.461042483926823
Size of file: 7398 bytes
Quantity of information: 4973.644442236354 bytes

```

	Розмір файлу
text1.txt.encoded	5374
text2.txt.encoded	3716
text3.txt.encoded	7398

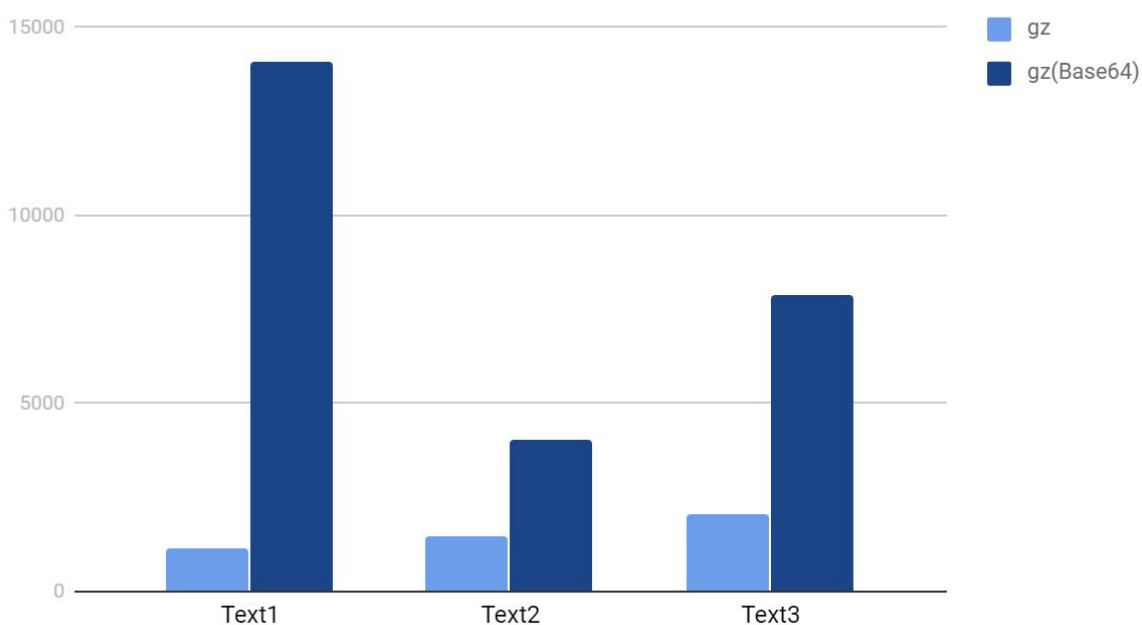
	Кількість інформації	Кількість інформації у закодованому файлі(Base64)
text1	1823	3557
text2	1563	2533
text3	3026	4973



У Base64 кількість інформації більша, оскільки задля реалізації алгоритма кодування 3 октети перетворюються у 4 сектети

	Розмір початкового файлу	Розмір файлу	Кількість інформації
text1.gz.arc.encoded	1155	1498	8355
text2.gz.arc.encoded	1470	4038	2719
text3.gz.arc.encoded	2068	7902	5377

Points scored



Таким чином, бачимо, що ситуація є аналогічною до попереднього пункту, проте розмір закодованого файлу значно більший, ніж розмір стиснутого.

**Висновок:** у ході виконання лабораторної роботи я ознайомився з поняттями ентропії та кількості інформації. Також практично побачив різницю між різними алгоритмами стиснення. Ознайомився з алгоритмом кодування Base64.