Звіт з лабораторної роботи №1

З курсу «Комп'ютерні системи»

Студентки 3 курсу КІ-СА

Уваренко Олени

1) Оберіть 3 текстових файла різного тематичного та лінгвістичного спрямування (наприклад, вірш Тараса Шевченка "Мені тринадцятий минало", "Казка про рєпку" Леся Подерв'янського та специфікацію інерфейсу PCI)

   Мої файли:
   i)   Стаття «Події поза планами кандидатів у президенти: Зеленський, Порошенко»
   ii) «Убога мова в нас» А.Фета
   iii) Класифікація штучних клапанів для серця

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

```
File to analize: fet.txt

Total summ of sumbols in file: 355

Relative frequency of appearing letter "а" in the text = 0.0563380281690141 ; Concrete letter present in the text: 20 times.
Relative frequency of appearing letter "б" in the text = 0.0112676056338028 ; Concrete letter present in the text: 4 times.
Relative frequency of appearing letter "в" in the text = 0.0535211267605634 ; Concrete letter present in the text: 19 times.
Relative frequency of appearing letter "г" in the text = 0.0169014084507042 ; Concrete letter present in the text: 6 times.
Relative frequency of appearing letter "ґ" in the text = 0 ; Concrete letter present in the text: 0 times.
Relative frequency of appearing letter "д" in the text = 0.0225352112676056 ; Concrete letter present in the text: 8 times.
Relative frequency of appearing letter "е" in the text = 0.0647887323943662 ; Concrete letter present in the text: 23 times.
Relative frequency of appearing letter "є" in the text = 0.00563380281690141 ; Concrete letter present in the text: 2 times.
Relative frequency of appearing letter "ж" in the text = 0.0140845070422535 ; Concrete letter present in the text: 5 times.
Relative frequency of appearing letter "з" in the text = 0.0225352112676056 ; Concrete letter present in the text: 8 times.
Relative frequency of appearing letter "и" in the text = 0.0929577464788732 ; Concrete letter present in the text: 33 times.
Relative frequency of appearing letter "і" in the text = 0.0507042253521127 ; Concrete letter present in the text: 18 times.
Relative frequency of appearing letter "ї" in the text = 0 ; Concrete letter present in the text: 0 times.
Relative frequency of appearing letter "й" in the text = 0.0112676056338028 ; Concrete letter present in the text: 4 times.
Relative frequency of appearing letter "к" in the text = 0.0197183098591549 ; Concrete letter present in the text: 7 times.
Relative frequency of appearing letter "л" in the text = 0.0394366197183099 ; Concrete letter present in the text: 14 times.
Relative frequency of appearing letter "м" in the text = 0.0450704225352113 ; Concrete letter present in the text: 16 times.
Relative frequency of appearing letter "н" in the text = 0.0619718309859155 ; Concrete letter present in the text: 22 times.
Relative frequency of appearing letter "о" in the text = 0.0704225352112676 ; Concrete letter present in the text: 25 times.
Relative frequency of appearing letter "п" in the text = 0.0366197183098592 ; Concrete letter present in the text: 13 times.
Relative frequency of appearing letter "р" in the text = 0.047887323943662 ; Concrete letter present in the text: 17 times.
Relative frequency of appearing letter "с" in the text = 0.0422535211267606 ; Concrete letter present in the text: 15 times.
Relative frequency of appearing letter "т" in the text = 0.0507042253521127 ; Concrete letter present in the text: 18 times.
Relative frequency of appearing letter "у" in the text = 0.028169014084507 ; Concrete letter present in the text: 10 times.
Relative frequency of appearing letter "ф" in the text = 0 ; Concrete letter present in the text: 0 times.
Relative frequency of appearing letter "х" in the text = 0.0225352112676056 ; Concrete letter present in the text: 8 times.
Relative frequency of appearing letter "ц" in the text = 0.00845070422535211 ; Concrete letter present in the text: 3 times.
Relative frequency of appearing letter "ч" in the text = 0.00845070422535211 ; Concrete letter present in the text: 3 times.
Relative frequency of appearing letter "ш" in the text = 0.0140845070422535 ; Concrete letter present in the text: 5 times.
Relative frequency of appearing letter "щ" in the text = 0.0028169014084507 ; Concrete letter present in the text: 1 times.
Relative frequency of appearing letter "ь" in the text = 0.0197183098591549 ; Concrete letter present in the text: 7 times.
Relative frequency of appearing letter "ю" in the text = 0.00563380281690141 ; Concrete letter present in the text: 2 times.
Relative frequency of appearing letter "я" in the text = 0.0169014084507042 ; Concrete letter present in the text: 6 times.

Average entropy of the unequipotent alphabet in the text: 4.43257982826731

Amount of information in the text: 196.695729879362
```

```
File to analize: heart.txt

Total summ of sumbols in file: 1929

Relative frequency of appearing letter "а" in the text = 0.0762052877138414 ; Concrete letter present in the text: 147 times.
Relative frequency of appearing letter "б" in the text = 0.00933125972006221 ; Concrete letter present in the text: 18 times.
Relative frequency of appearing letter "в" in the text = 0.046656298600311 ; Concrete letter present in the text: 90 times.
Relative frequency of appearing letter "г" in the text = 0.0145152928978745 ; Concrete letter present in the text: 28 times.
Relative frequency of appearing letter "ґ" in the text = 0 ; Concrete letter present in the text: 0 times.
Relative frequency of appearing letter "д" in the text = 0.0207361327112494 ; Concrete letter present in the text: 40 times.
Relative frequency of appearing letter "е" in the text = 0.0580611715914982 ; Concrete letter present in the text: 112 times.
Relative frequency of appearing letter "є" in the text = 0.00881285640228098 ; Concrete letter present in the text: 17 times.
Relative frequency of appearing letter "ж" in the text = 0.0046656298600311 ; Concrete letter present in the text: 9 times.
Relative frequency of appearing letter "з" in the text = 0.0217729393468118 ; Concrete letter present in the text: 42 times.
Relative frequency of appearing letter "и" in the text = 0.0585795749092794 ; Concrete letter present in the text: 113 times.
Relative frequency of appearing letter "і" in the text = 0.0570243649559357 ; Concrete letter present in the text: 110 times.
Relative frequency of appearing letter "ї" in the text = 0.0067392431313115604 ; Concrete letter present in the text: 13 times.
Relative frequency of appearing letter "й" in the text = 0.00829445308449974 ; Concrete letter present in the text: 16 times.
Relative frequency of appearing letter "к" in the text = 0.0451010886469673 ; Concrete letter present in the text: 87 times.
Relative frequency of appearing letter "л" in the text = 0.0368066355624676 ; Concrete letter present in the text: 71 times.
Relative frequency of appearing letter "м" in the text = 0.0285121824779679 ; Concrete letter present in the text: 55 times.
Relative frequency of appearing letter "н" in the text = 0.0787973043027475 ; Concrete letter present in the text: 152 times.
Relative frequency of appearing letter "о" in the text = 0.0855365474339036 ; Concrete letter present in the text: 165 times.
Relative frequency of appearing letter "п" in the text = 0.0393986521513738 ; Concrete letter present in the text: 76 times.
Relative frequency of appearing letter "р" in the text = 0.055469155002592 ; Concrete letter present in the text: 107 times.
Relative frequency of appearing letter "с" in the text = 0.0378434421980301 ; Concrete letter present in the text: 73 times.
Relative frequency of appearing letter "т" in the text = 0.0502851218247797 ; Concrete letter present in the text: 97 times.
Relative frequency of appearing letter "у" in the text = 0.0305857957490928 ; Concrete letter present in the text: 59 times.
Relative frequency of appearing letter "ф" in the text = 0.0031104199066874 ; Concrete letter present in the text: 6 times.
Relative frequency of appearing letter "х" in the text = 0.0139968895800933 ; Concrete letter present in the text: 27 times.
Relative frequency of appearing letter "ц" in the text = 0.0202177293934681 ; Concrete letter present in the text: 39 times.
Relative frequency of appearing letter "ч" in the text = 0.0171073094867807 ; Concrete letter present in the text: 33 times.
Relative frequency of appearing letter "ш" in the text = 0.0062208398133748 ; Concrete letter present in the text: 12 times.
Relative frequency of appearing letter "щ" in the text = 0.00259201658890617 ; Concrete letter present in the text: 5 times.
Relative frequency of appearing letter "ь" in the text = 0.0134784862623121 ; Concrete letter present in the text: 26 times.
Relative frequency of appearing letter "ю" in the text = 0.0134784862623121 ; Concrete letter present in the text: 26 times.
Relative frequency of appearing letter "я" in the text = 0.0212545360290306 ; Concrete letter present in the text: 41 times.

Average entropy of the unequipotent alphabet in the text: 4.55328529664514

Amount of information in the text: 1097.91091715356
```

```
File to analize: zelen.txt

Total summ of sumbols in file: 2221

Relative frequency of appearing letter "а" in the text = 0.0805943268797839 ; Concrete letter present in the text: 179 times.
Relative frequency of appearing letter "б" in the text = 0.0175596578117965 ; Concrete letter present in the text: 39 times.
Relative frequency of appearing letter "в" in the text = 0.0495272399819901 ; Concrete letter present in the text: 110 times.
Relative frequency of appearing letter "г" in the text = 0.00945520036019811 ; Concrete letter present in the text: 21 times.
Relative frequency of appearing letter "ґ" in the text = 0 ; Concrete letter present in the text: 0 times.
Relative frequency of appearing letter "д" in the text = 0.0396217919855921 ; Concrete letter present in the text: 88 times.
Relative frequency of appearing letter "е" in the text = 0.0463755065285907 ; Concrete letter present in the text: 103 times.
Relative frequency of appearing letter "є" in the text = 0.0031517345339937 ; Concrete letter present in the text: 7 times.
Relative frequency of appearing letter "ж" in the text = 0.00945520036019811 ; Concrete letter present in the text: 21 times.
Relative frequency of appearing letter "з" in the text = 0.0162089149031968 ; Concrete letter present in the text: 36 times.
Relative frequency of appearing letter "и" in the text = 0.0675371454299865 ; Concrete letter present in the text: 150 times.
Relative frequency of appearing letter "і" in the text = 0.0693381359747861 ; Concrete letter present in the text: 154 times.
Relative frequency of appearing letter "ї" in the text = 0.00495272399819901 ; Concrete letter present in the text: 11 times.
Relative frequency of appearing letter "й" in the text = 0.0135074290859973 ; Concrete letter present in the text: 30 times.
Relative frequency of appearing letter "к" in the text = 0.0409725348941918 ; Concrete letter present in the text: 91 times.
Relative frequency of appearing letter "л" in the text = 0.040072039621792 ; Concrete letter present in the text: 89 times.
Relative frequency of appearing letter "м" in the text = 0.0292660963529941 ; Concrete letter present in the text: 65 times.
Relative frequency of appearing letter "н" in the text = 0.0630346690679874 ; Concrete letter present in the text: 140 times.
Relative frequency of appearing letter "о" in the text = 0.0855470508779829 ; Concrete letter present in the text: 190 times.
Relative frequency of appearing letter "п" in the text = 0.0238631247185952 ; Concrete letter present in the text: 53 times.
Relative frequency of appearing letter "р" in the text = 0.045024763619991 ; Concrete letter present in the text: 100 times.
Relative frequency of appearing letter "с" in the text = 0.0328680774425934 ; Concrete letter present in the text: 73 times.
Relative frequency of appearing letter "т" in the text = 0.0540297163439892 ; Concrete letter present in the text: 120 times.
Relative frequency of appearing letter "у" in the text = 0.0315173345339937 ; Concrete letter present in the text: 70 times.
Relative frequency of appearing letter "ф" in the text = 0.00720396217919856 ; Concrete letter present in the text: 16 times.
Relative frequency of appearing letter "х" in the text = 0.00810445745159838 ; Concrete letter present in the text: 18 times.
Relative frequency of appearing letter "ц" in the text = 0.0090049527239982 ; Concrete letter present in the text: 20 times.
Relative frequency of appearing letter "ч" in the text = 0.00810445745159838 ; Concrete letter present in the text: 18 times.
Relative frequency of appearing letter "ш" in the text = 0.0103556956325979 ; Concrete letter present in the text: 23 times.
Relative frequency of appearing letter "щ" in the text = 0.00720396217919856 ; Concrete letter present in the text: 16 times.
Relative frequency of appearing letter "ь" in the text = 0.0193606483565961 ; Concrete letter present in the text: 43 times.
Relative frequency of appearing letter "ю" in the text = 0.00720396217919856 ; Concrete letter present in the text: 16 times.
Relative frequency of appearing letter "я" in the text = 0.0202611436289959 ; Concrete letter present in the text: 45 times.

Average entropy of the unequipotent alphabet in the text: 4.47647389634111

Amount of information in the text: 1242.7810654717
```

3) 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).

4) 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)

Кількість інформації - мінімальна кількість біт, необхідна для кодування всіх можливих значень повідомлення. Вона буде меншою за розмір наший файлів.
З розповсюджених визначень, можна навести таке:
Кількість інформації можна розгядати як міру зменшення невизначенності знання
При отриманні інформаційних поідомень.

Файл	fet.txt, bytes	heart.txt, bytes	zelen.txt, bytes
rar	524	1,527	1,994
tar	2,560	6,144	6,656

bz2	383	1,201	1,601
zip	577	1,579	2,051
xz	508	1,452	1,880
ORIGINAL	817	4,222	4,982
Кількість інформації	196	1097	1242
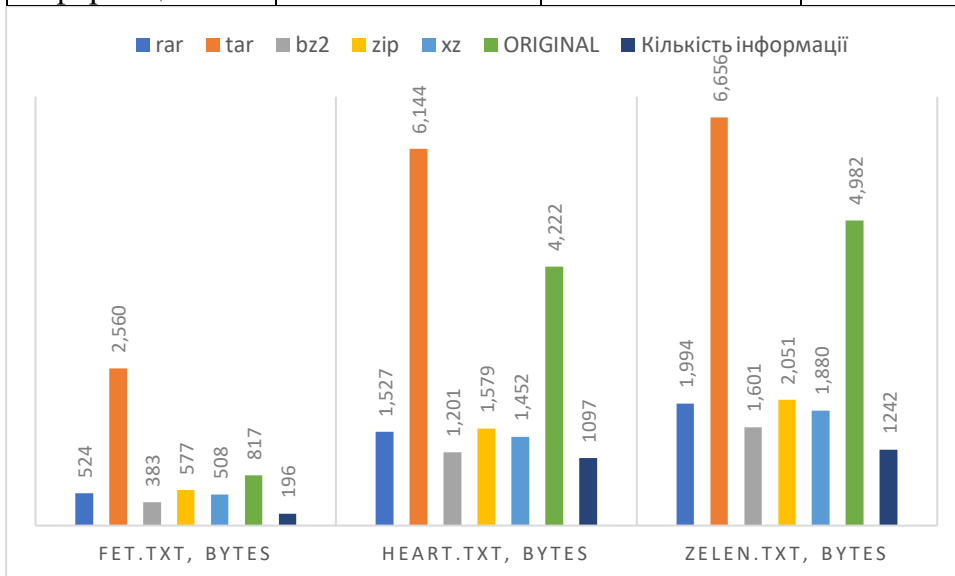
Розміри архівованих файлів у трохи більші за кількість інформації. Однак, коли ми прицюємо з великим текстом, помітне сильніше стиснення.

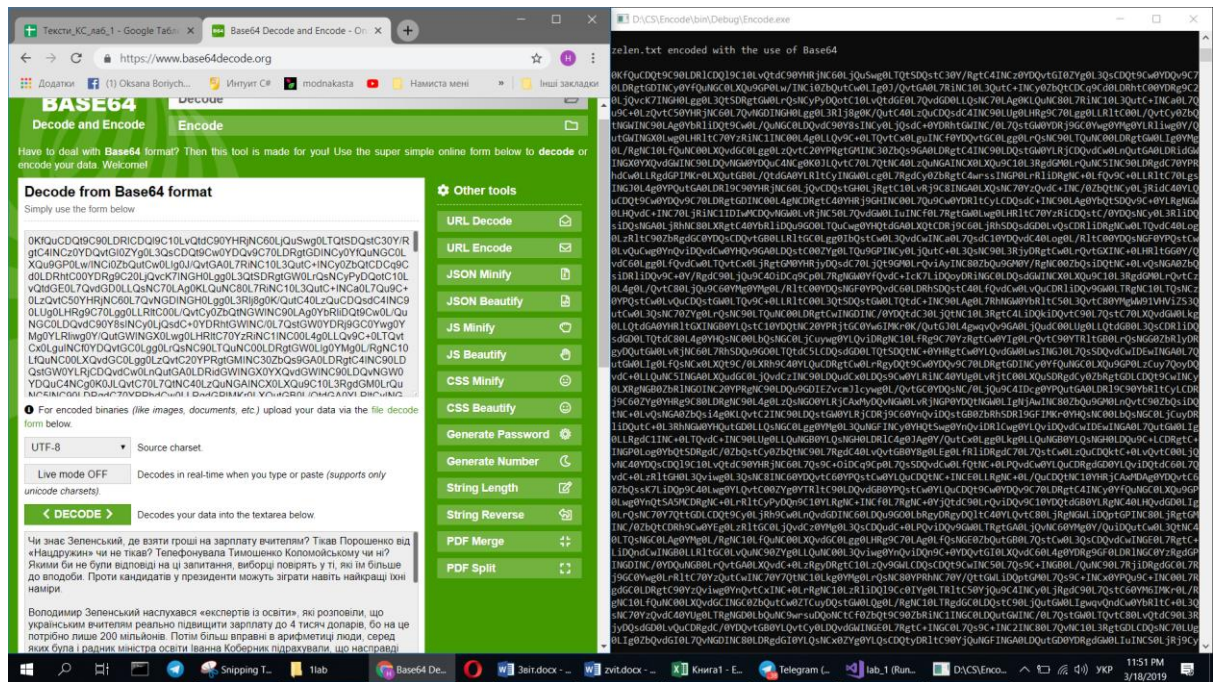2. Дослідження способів кодування інформації на прикладі Base64

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)
перевірте коректність роботи програми, порівнявши результат з існуючими програмними засобами (наприклад, openssl enc -base64)

1) Закодуйте в Base64 обрані вами текстові файли

2) Правильність кодування перевірено онлайн. Тексти співпадають.
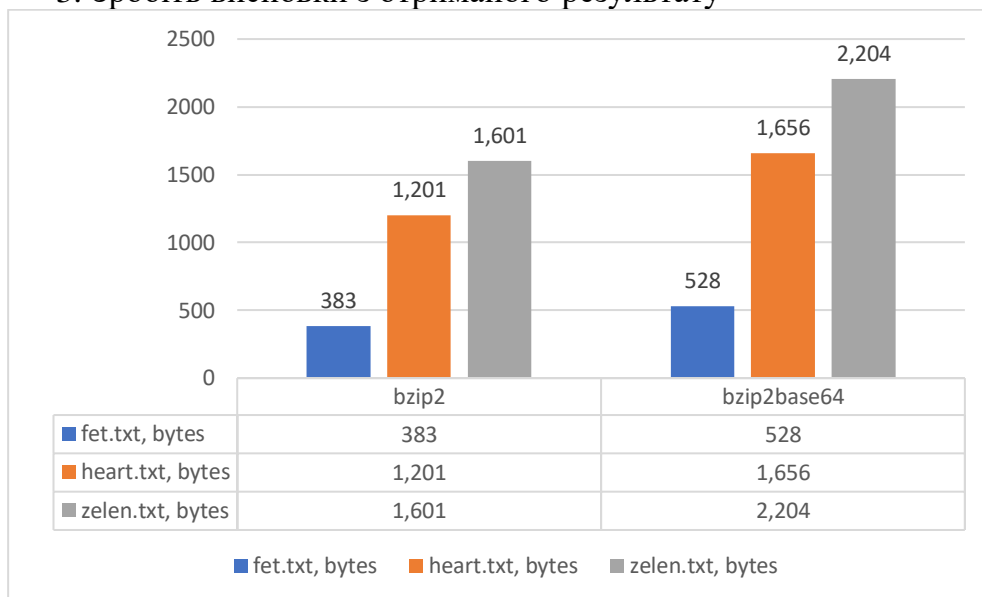
   i)   Обрахуйте кількість інформації в base64-закодованому варіанті файлу

   ii) Порівняйте отримане значення з кількістю інформації вихідного файлу

   iii) Зробіть висновки з отриманого результату

	Original	Base64
fet.txt, bytes	817	1,092
heart.txt, bytes	4,222	5,632
zelen.txt, bytes	4,982	6,650
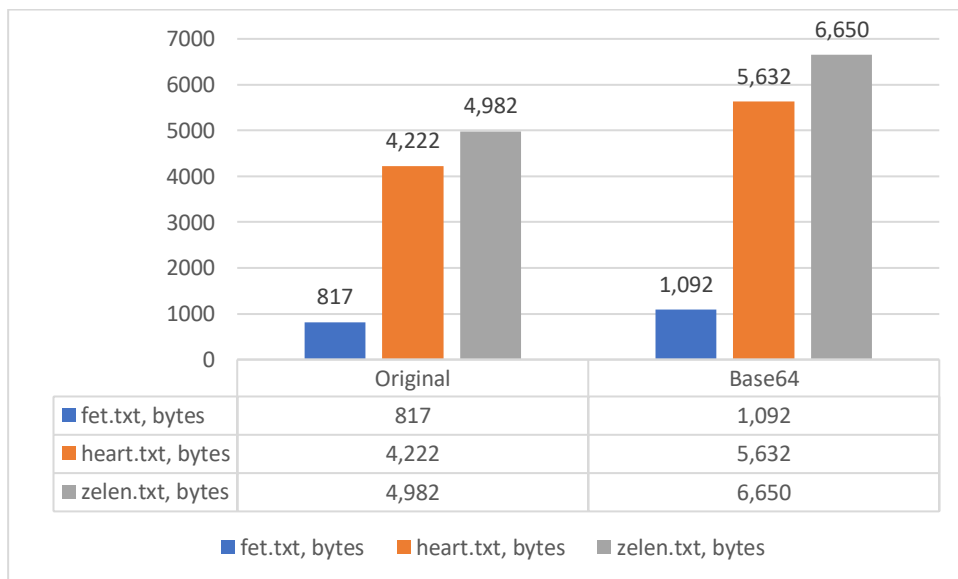
Бачимо, що кількість інформації у закодованих файлах зросла приблизно на третину (особливості алгоритму base64).

1. Закодуйте в Base64 стиснені кращим з алгоритмів текстові файли

     1.Обрахуйте кількість інформації в base64-закодованому варіанті стисненого файлу

2. Порівняйте отримане значення з кількістю інформації вихідного файлу та base64-закодованого файлу
3. Зробіть висновки з отриманого результату

	bzip2	bzip2base64
■ fet.txt, bytes	383	528
■ heart.txt, bytes	1,201	1,656
■ zelen.txt, bytes	1,601	2,204

■ fet.txt, bytes   ■ heart.txt, bytes   ■ zelen.txt, bytes

Аналогічно до попереднього завдання, розмір файлів зріс на третину.

Посилання на код програм:

https://github.com/helenuvarenko/Computer-Systems/tree/master/1lab

Висновки до лабораторної роботи:

Під час виконання завдань лабораторної роботи №1 було здобуто навички із аналізування ефективності застосування різних типів архіваторів, застосовано знання для обрахунку ентропії та кодування за різними стандартами.