1. 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).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 7z | bz2 | gz | rar | zip | Обч. інфо  в байтах | Реальний об’єм |
| Text-1 | 1290 | 1047 | 1208 | 1305 | 1335 | 1112 | 3056 |
| Text-2 | 1509 | 1098 | 1376 | 1478 | 1508 | 1252 | 3583 |
| Text-3 | 1157 | 890 | 1043 | 1135 | 1175 | 1314 | 3815 |

Для кращого порівняння цих величин введемо поняття **коефіцієнта стиснення**

Text-1: вірт. коефіціент стиснення = реальний об’єм / обчислена інформація = 2.75

Text-2: вірт. коефіціент стиснення = реальний об’єм / обчислена інформація = 2.86

Text-3: вірт. коефіціент стиснення = реальний об’єм / обчислена інформація = 2.9

**Висновок**: так як в Text-1 наявний текст про Base64 специфікацію, яка має як латинські так і кириличні літери в словах, із цього слідує що середня ентропія буде вищою, а отже і віртуальний коефіцієнт стиснення менший(2.75). На противагу цьому, в Text-2 та Text-3 використовується лише кириличний алфавіт, а отже і віртуальний коефіцієнт стиснення більший.(2.86 i 2.9 відповідно). Якщо обрахувати фактичний середній коефіцієнт стиснення за 5-тьма архіваторами для кожного тексту:

Text-1: сер. коефіцієнт стиснення = сума коефіцієнтів / к-сть архіваторів

= (2.37 + 2.92 + 2.53 + 2.34 + 2.29) / 5 = 12.45 / 5 = 2.49

Text-2: сер. коефіцієнт стиснення = сума коефіцієнтів / к-сть архіваторів

= (2.37 + 3.26 + 2.6 + 2.42 + 2.38) / 5 = 13.03 / 5 = 2.606

Text-3: сер. коефіцієнт стиснення = сума коефіцієнтів / к-сть архіваторів

= (3.3 + 4.29 + 3.66 + 3.36 + 3.25) / 5 = 17.86 / 5 = 3.572

Таким чином середні коефіцієнти стиснення для Text-1 та Text-2 менші за віртуальні значення(різниця в 0.26), але для Text-3 значно більший сер. коефіцієнт за віртуальний (різниця в 0.67)

1. Закодуйте в Base64 обрані вами текстові файли. Обрахуйте кількість інформації в base64-закодованому варіанті файлу.Порівняйте отримане значення з кількістю інформації вихідного файлу.Зробіть висновки з отриманого результату.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | К-сть Інформації вихідного ф-ла | Об’єм вихідного ф-ла | К-сть Інформації Base64-закодованого | Об’єм Base64-закодованого |
| Text-1 | 1112 | 3056 | 2514 | 4076 |
| Text-2 | 1252 | 3583 | 2968 | 4780 |
| Text-3 | 1314 | 3815 | 3161 | 5088 |

Text-1: Base64 вірт. коефіціент стиснення=реальний об’єм Base64 / обчислена інформація Base64 = 1.6

Text-2: Base64 вірт. коефіціент стиснення=реальний об’єм Base64 / обчислена інформація Base64 = 1.61

Text-3: Base64 вірт. коефіціент стиснення=реальний об’єм Base64 / обчислена інформація Base64 = 1.609

По-перше, можна помітити що Base64 віртуальний коефіцієнт стиснення для різних файлів є майже однаковим. Це так тому, що середня ентропія в них майже однакова(бо однакова к-сть символів алфавіту) , а к-сть символів відрізняється.

По-друге, стиснення Вихідного файла краще, за стиснення Base64-закодованого файла. Це так, бо незважаючи на меншу ентропію(4.9 проти 5.1), Base64-закодований файл має в рази більшу к-сть символів(4076 проти 1735). З цього слідує, що якщо здійснити кодування Base64 таким чином, що кожним 3-ьом символам відповідає 4-и символа патерну, то завдяки прямій пропорційності кількість інформації для Base64 збільшиться в дану к-сть разів.

1. Закодуйте в Base64 стиснені кращим з алгоритмів текстові файли. Обрахуйте кількість інформації в base64-закодованому варіанті стисненого файлу. Порівняйте отримане значення з кількістю інформації вихідного файлу та base64-закодованого файлу. Зробіть висновки з отриманого результату.

Беручи до уваги таблицю 1, найкращий алгоритм стиснення – **bz2.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | К-сть Інформації Base-64 стисненого ф-ла | Об’єм стисненого  файлу | К-сть інфо Вихідного файла | К-сть Інформації  Base-64 ф-ла | Об’єм Base64-закодованого |
| Text-1 | 1038 | 1047 | 1112 | 2514 | 4076 |
| Text-2 | 1089 | 1098 | 1252 | 2968 | 4780 |
| Text-3 | 885 | 890 | 1314 | 3161 | 5088 |

Кількість інформації в Base64-закодованому варіанті стисненого файлу майже не відрізняється від об’єму стисненого файла, тому більшого стиснення не відбудеться, так як можлива втрата даних. А так як об’єм стисненого файла менша за к-сть інформації вихідного файла - це означає, що алгоритм стиснення bz2 гарно стискає дані.