**Київський національний університет імені Тараса Шевченка**

**радіофізичний факультет**

лабораторна робота № 1

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

Роботу виконав

студент 3 курсу

Комп’ютерна інженерія

Онищук Родіон Сергійович

Київ 2018

***Дослідження кількості інформації в тексті***

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Мені тринадцятий минало | | Казка про рєпку, або Хулі не ясно | | Специфікація інтерфейсу PCI | |
| Символ | Частота | Символ | Частота | Символ | Частота |
| м | 0,0239149689991143 | д | 0,0285925726587729 | p | 0,00930343511450382 |
| е | 0,0496014171833481 | i | 6,72766415500538E-05 | c | 0,012881679389313 |
| н | 0,0566873339238264 | я | 0,0176937567276642 | i | 0,00954198473282443 |
| і | 0,0389725420726306 | « » | 0,147672228202368 | « » | 0,145992366412214 |
| « » | 0,131089459698849 | п | 0,0189047362755651 | 2 | 0,00739503816793893 |
| т | 0,029229406554473 | е | 0,0265069967707212 | . | 0,0109732824427481 |
| р | 0,0230292294065545 | р | 0,0342438105489774 | 0 | 0,00333969465648855 |
| и | 0,0380868024800709 | ш | 0,00814047362755651 | - | 0,00763358778625954 |
| а | 0,0637732506643047 | а | 0,0806646932185145 | п | 0,0197996183206107 |
| д | 0,0203720106288751 | . | 0,0224031216361679 | е | 0,0238549618320611 |
| ц | 0,0053144375553587 | \n | 0,0102260495156082 | р | 0,039837786259542 |
| я | 0,0265721877767936 | х | 0,00773681377825619 | ш | 0,00596374045801527 |
| й | 0,0070859167404783 | т | 0,0359930032292788 | а | 0,0589217557251908 |
| л | 0,0513728963684677 | с | 0,0395586652314316 | в | 0,0333969465648855 |
| о | 0,0903454384410983 | в | 0,0445371367061356 | с | 0,0357824427480916 |
| . | 0,0318866253321523 | и | 0,0573196986006459 | і | 0,0448473282442748 |
| \n | 0,0442869796279894 | о | 0,0771663078579117 | я | 0,0133587786259542 |
| п | 0,0212577502214349 | н | 0,0475645855758881 | б | 0,0119274809160305 |
| с | 0,0327723649247121 | ч | 0,016280947255113 | з | 0,0176526717557252 |
| г | 0,0230292294065545 | - | 0,0026237890204521 | г | 0,00715648854961832 |
| з | 0,0115146147032772 | з | 0,0184337997847147 | т | 0,0462786259541985 |
| ч | 0,0186005314437555 | б | 0,0144644779332616 | н | 0,0546278625954198 |
| к | 0,016829052258636 | ц | 0,00464208826695371 | д | 0,0193225190839695 |
| , | 0,0265721877767936 | м | 0,0253632938643703 | у | 0,0207538167938931 |
| б | 0,0194862710363153 | і | 0,0346474703982777 | , | 0,0112118320610687 |
| у | 0,0203720106288751 | , | 0,0215285252960172 | щ | 0,00262404580152672 |
| ? | 0,0008857395925597 | к | 0,0332346609257266 | ж | 0,00524809160305344 |
| ю | 0,0106288751107174 | у | 0,0247578040904198 | л | 0,0188454198473282 |
| в | 0,0265721877767936 | є | 0,00867868675995694 | и | 0,0384064885496183 |
| ж | 0,0053144375553587 | л | 0,0267088266953714 | к | 0,0236164122137405 |
| ' | 0,0008857395925597 | ь | 0,0146663078579117 | ь | 0,0124045801526718 |
| ь | 0,0106288751107174 | г | 0,008745963401507 | ю | 0,00834923664122137 |
| є | 0,0035429583702391 | ж | 0,00706404736275565 | 5 | 0,00286259541984733 |
| ! | 0,0079716563330380 | ю | 0,00645855758880517 | v | 0,00500954198473282 |
| : | 0,0008857395925597 | щ | 0,00195102260495156 | 1 | 0,00500954198473282 |
| - | 0,0017714791851195 | ї | 0,00356566200215285 | 3 | 0,0147900763358779 |
| ї | 0,0026572187776793 | й | 0,0078713670613563 | м | 0,0233778625954198 |
| х | 0,0026572187776793 | ф | 0,00282561894510226 | / | 0,00310114503816794 |
| щ | 0,0008857395925597 | ( | 0,00154736275565124 | ; | 0,00429389312977099 |
| ш | 0,0008857395925597 | ) | 0,00154736275565124 | \n | 0,0045324427480916 |
| [ | 0,0008857395925597 | ! | 0,00275834230355221 | - | 0,0066793893129771 |
| ] | 0,0008857395925597 | ' | 0,000740043057050592 | ч | 0,00596374045801527 |
|  | | ? | 0,00289289558665231 | ї | 0,00357824427480916 |
| " | 0,000269106566200215 | х | 0,00811068702290076 |
| : | 0,000336383207750269 | b | 0,000715648854961832 |
| 5 | 0,000134553283100108 | u | 0,000954198473282443 |
| 2 | 0,000134553283100108 | s | 0,000954198473282443 |
| 0 | 0,000134553283100108 | m | 0,00143129770992366 |
|  | | a | 0,00381679389312977 |
| t | 0,00143129770992366 |
| e | 0,00381679389312977 |
| r | 0,00357824427480916 |
| ( | 0,00238549618320611 |
| й | 0,00644083969465649 |
| ) | 0,00238549618320611 |
| ц | 0,00787213740458015 |
| 6 | 0,0104961832061069 |
| ' | 0,000954198473282443 |
| є | 0,00596374045801527 |
| 4 | 0,00477099236641221 |
| ф | 0,00310114503816794 |
| x | 0,000238549618320611 |
| : | 0,000715648854961832 |
| n | 0,00238549618320611 |
| o | 0,00190839694656489 |
| g | 0,000477099236641221 |
| d | 0,00166984732824427 |
| h | 0,000238549618320611 |
| k | 0,000477099236641221 |
| 9 | 0,000477099236641221 |
| ‘ | 0,000238549618320611 |
| l | 0,000477099236641221 |
| z | 0,000477099236641221 |
| 8 | 0,000238549618320611 |
| Ентропія – 4 біт | | Ентропія – 4 біт | | Ентропія – 5 біт | |
| Кількість інф. – 658 байт | | Кількість інф. – 8710 байт | | Кількість інф. – 2689 байт | |

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)

|  |  |  |  |
| --- | --- | --- | --- |
| Формат | Мені тринадцятий минало | Казка про рєпку, або Хулі не ясно | Специфікація інтерфейсу PCI |
| .7z | 1 016 байт | 7 893 байт | 2 330 байт |
| .bz2 | 720 байт | 6 395 байт | 1 925 байт |
| .gz | 895 байт | 8 067 байт | 2 284 байт |
| .xz | 924 байт | 7 780 байт | 2 232 байт |
| .zip | 1 079 байт | 8 271 байт | 2 474 байт |

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

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

|  |  |  |  |
| --- | --- | --- | --- |
| Кодування .txt файлу | Мені тринадцятий минало | Казка про рєпку, або Хулі не ясно | Специфікація інтерфейсу PCI |
| Без | 658 байт | 8710 байт | 2689 байт |
| Base64 | 1668 байт | 21785 байт | 6034 байт |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Кодування | Мені тринадцятий минало | Казка про рєпку, або Хулі не ясно | Специфікація інтерфейсу PCI |
| Без (.txt) | 658 байт | 8710 байт | 2689 байт |
| Base64 (.txt) | 1668 байт | 21785 байт | 6034 байт |
| Base64 (.bz2) | 710 байт | 6383 байт | 1905 байт |

***Приклад роботи програми***

