Міністерство освіти і науки України

Черкаський державний технологічний університет

Кафедра програмного забезпечення автоматизованих систем

**ЗВІТ**

з лабораторної роботи №1

з предмету «Основи інтеграції інформаційних потоків»

|  |  |
| --- | --- |
| Перевірив:  Д. т. н., проф., зав. каф. ПЗАС  Первунінський С.М.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  «\_\_\_\_» \_\_\_\_\_\_\_\_\_\_ 2020 р. | Виконав:  студент 1-го курсу  групи МПЗ-1904  Гаврилюк В. Є. |

Черкаси 2020

**Лабораторна робота №1**

**Тема роботи:** Дослідження ентропії джерел повідомлень.

**Мета роботи:** Закріпити теоретичні знання й придбати навички по дослідженню ентропії джерел повідомлень дискретного й безперервного типів.

**Завдання:**

* вивчити методи розрахунків ентропії джерел інформації;
* ознайомитися з описом лабораторної роботи;
* підготувати бланк звіту;
* підготувати відповіді на контрольні питання.

**Теоретичні відомості**

Розвиток теорії інформації як науки став можливим завдяки введенню поняття кількості інформації*,* щовизначає математичні властивості переданих і прийнятих повідомлень.

Поняття про інформацію припускає абстрагування від багатьох властивостей реальних подій і об'єктів. Границі абстрагування досить умовні й визначаються специфікою розв'язуваних завдань.

Матеріальна форма втілення інформаціїназивається *повідомленням.* Повідомлення можуть бути представлені у вигляді показань приладів, станів фізичних елементів, друкованого тексту і т.д. Повідомлення, безпосередньо призначене для передачі інформації називається сигналом. Широке застосування одержали сигнали електричні, електромагнітні, звукові й ін.

Повідомлення й сигнали можуть бути безперервними й дискретними як по станах так і в часі. Великий клас дискретних повідомлень, які передаються у системах зв'язку, можна представити в загальному виді, у вигляді *n* елементів, кожний з яких може перебувати в одному з *m* різних фіксованих станів. Такі повідомлення будемо називати дискретними по станам елементів.

Графічно дискретне повідомлення може бути презентовано у вигляді діаграми, зображеної на рис. 1, де по горизонтальній осі розміщені елементи повідомлення а по вертикальній - стан цих елементів.



*Рис. 1. Діаграма станів елементів повідомлення*

Помітимо, що іноді *повідомлення* називають словом, елементи повідомлення - символами, стан елементів - буквами, а всю сукупність можливих станів -алфавітом. Тому розглянуте вище повідомлення, представлене в загальному виді, можна трактувати як слово, що полягає з *n* символів, заданих алфавітом з *m* букв.

Якщо повідомлення складені з *n* елементів, то кількість можливих різних повідомлень, що відрізняються станами елементів, дорівнює .

У різних додатках теорії інформації зручно оперувати як із загальною кількістю інформації, рівною , так і з кількістю інформації, що доводяться на один елемент повідомлень.

Кількість інформації, що доводиться на один елемент, називається питомою інформативністю або ентропією повідомлень і позначається буквою *H*:

.

Нехай у повідомленні, представленому на рис. 1, кожний з елементів є незалежним і може мати кожне з *m* станів  з імовірностями  відповідно. Якщо для деякого поодинокого повідомлення число елементів, що прийняли стан, дорівнює , число елементів, що прийняли стан, дорівнює  й т.д, то таке повідомлення може бути презентовано наступною схемою:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Стан елементів |  |  |  | ... |  | ... |  |
| Імовірності станів |  |  |  | ... |  | ... |  |
| Число елементів |  |  |  | ... |  | ... |  |

Якщо число елементів утворюючих повідомлення досить велике, то за ймовірність прийняти  елементам стан  можна взяти частость, рівну відношенню числа елементів, що прийняли цей стан, до загального числа елементів у повідомленні: .

Нескладно підрахувати й середню ймовірність передачі одного повідомлення з параметрами, зазначеними в таблиці,

.

По середній імовірності  передачі одного повідомлення можна обчислити середнє число всіх можливих повідомлень:

.

Знаючи середнє число  всіх можливих повідомлень, неважко визначити й середня кількість інформації , що міститься в одному повідомленні:

. (1)

Розділивши середню кількість інформації, що міститься в одному повідомленні, на число його елементів, одержимо середню ентропію  повідомлень:

. (2)

Надалі для стислості замість середня ентропія й середня кількість інформації будемо казати просто ентропія й кількість інформації, маючи наувазі, що мова йде про середні величини.

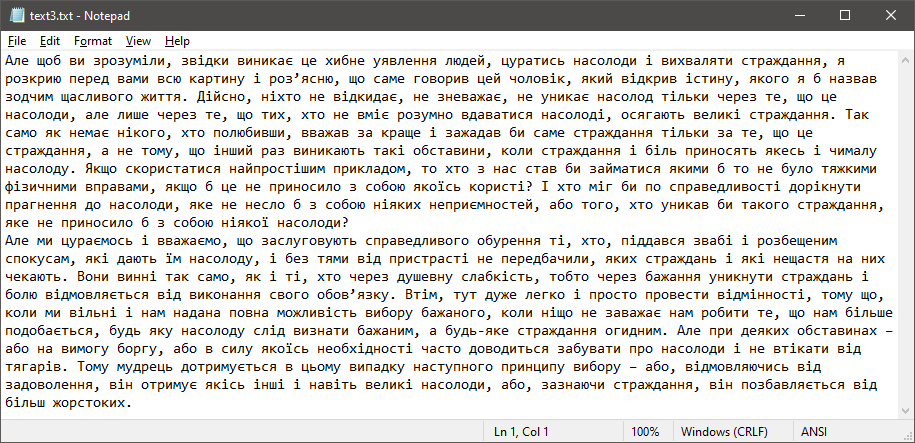
Вираз для обчислення ентропії джерела повідомлень із безперервним розподілом станів елементів визначається по формулі

, (3)

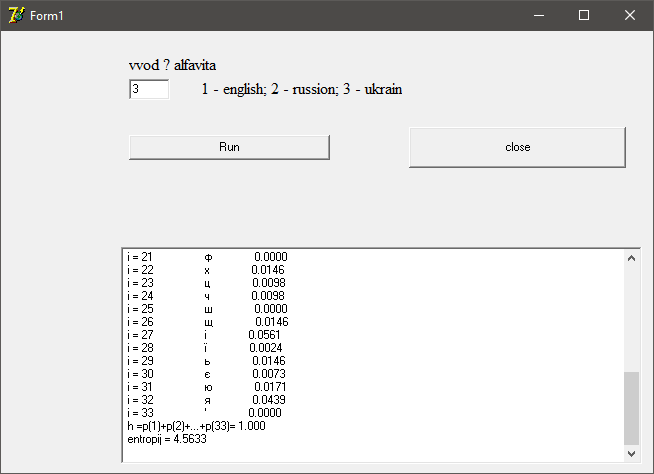
де  - наведена ентропія джерела повідомлень;  - щільність розподілу станів;  - інтервал квантування станів джерела повідомлень.

**Хід роботи**

1. Визначити ентропію повідомлень, представлених машинописним текстом на український російський й англійських мовах. Дослідити залежність величини ентропії повідомлень від кількості символів у них. Результати досліджень представити графічними залежностями.
   1. Обробка тексту на українській мові.



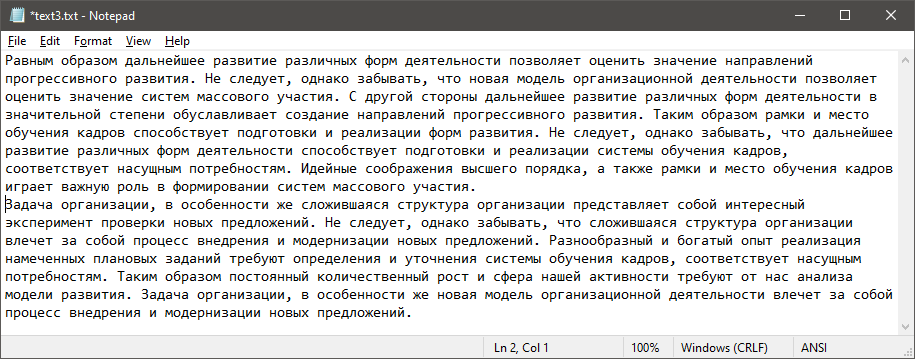
*Рис. 1. Додавання тексту українською мовою у файл text3.txt*



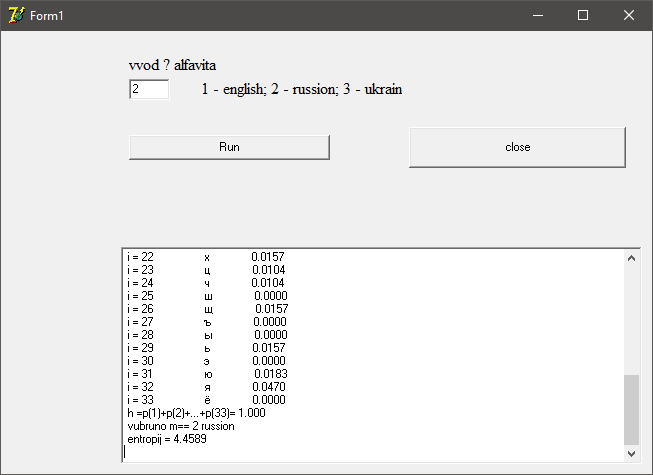
*Рис. 2. Результат аналізу тексту на українській мові*

Результат:

* кількість символів: 410;
* ентропія: 4.5633.
  1. Обробка тексту на російській мові.



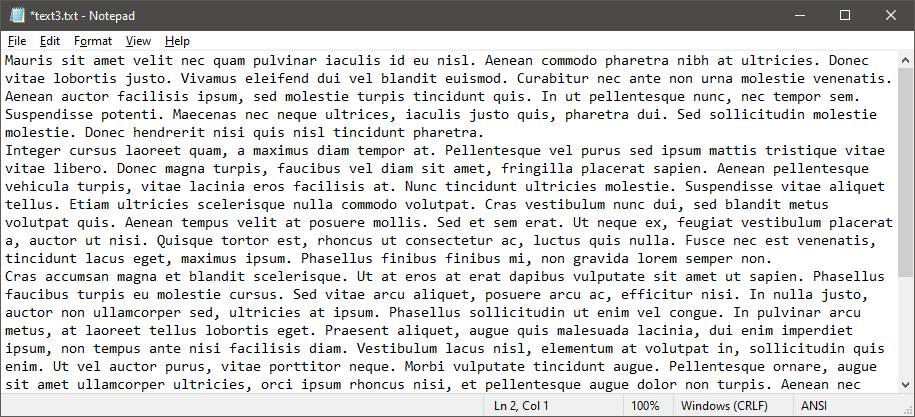
*Рис. 3. Додавання тексту російською мовою у файл text3.txt*



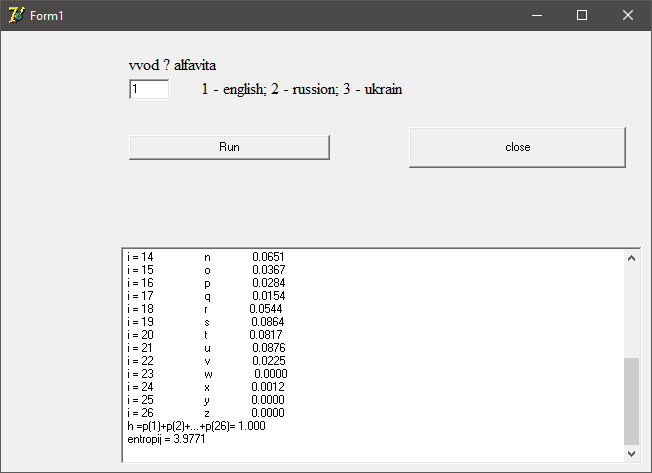
*Рис. 4. Результат аналізу тексту на російській мові*

Результат:

* кількість символів: 383;
* ентропія: 4.4589.
  1. Обробка тексту на англійській мові.



*Рис. 5. Додавання тексту англійською мовою у файл text3.txt*



*Рис. 6. Результат аналізу тексту на англійській мові*

Результат:

* кількість символів: 845;
* ентропія: 3.9771.
  1. Графічне відображення залежності величини ентропії повідомлень від кількості символів у них.

*Рис. 7. Залежність величини ентропії повідомлень від кількості символів у них*

1. Визначити ентропію джерела з безперервним розподілом станів елементів.

|  |  |  |
| --- | --- | --- |
| Кількість символів «N» | H очікуване | H отримане |
| 10 | 3,1751 | 3,1653 |
| 15 | 3,752 | 3,7535 |
| 20 | 4,159 | 4,1058 |
| 25 | 4,474 | 4,4033 |
| 30 | 4,73 | 4,6507 |
| 35 | 4,947 | 4,8447 |
| 40 | 5,135 | 5,0986 |
| 45 | 5,301 | 5,2165 |
| 50 | 5,449 | 5,3697 |
| 55 | 5,583 | 5,4921 |
| 60 | 5,705 | 5,6215 |

На основі отриманих даних було побудовано графік, який зображено на рисунку 8. Гістограма позначає значення отриманої ентропії у порівнянні із очікувано для кожного із значень змінної «N».

* синя лінія – значення очікуваної ентропії;
* оранжевий стовпець – значення отриманої ентропії для відповідного значення «N».

*Рис. 8. Графік залежності ентропії джерела з безперервним розподілом станів елементів*

**Висновок**: на даній лабораторній роботі я дослідив ентропію джерел повідомлень. Закріпив теоретичні знання й придбав навички по дослідженню ентропії джерел повідомлень дискретного й безперервного типів.