

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Сибирский государственный университет телекоммуникаций и
информатики»
(СибГУТИ)

Кафедра прикладной математики и кибернетики

Теория информации
Лабораторная работа №1

Выполнил: студент гр. ИП-711

Мартасов Илья Олегович

Проверила:

Доцент кафедры ПМиК

Мачикина Елена Павловна

Новосибирск 2021 г.

Содержание

| | |
|------------------------------------|---|
| Расчет теоретических значений..... | 3 |
| Результаты работы | 5 |
| Вывод..... | 6 |

Расчет теоретических значений

Алфавит источника: a, b, c, d

Файл F1:

Частоты отдельных символов:

$$P(a) = P(b) = P(c) = P(d) = 0.25$$

$$H(P(a), P(b), P(c), P(d)) = (-0.25 * \log[2, 0.25]) * 4 = 0.5 * 4 = \mathbf{2}$$

Частоты пар символов:

$$P(aa) = P(ab) = P(ac) = P(ad) = P(ba) = P(bb) = P(bc) = P(bd) = P(ca) = P(cb) = P(cc) = P(cd) = P(da) = P(db) = P(dc) = P(dd) = 0.0625$$

$$H(P(1), \dots, P(n)) = ((-0.0625 * \log[2, 0.0625]) * 16) / 2 = \mathbf{2}$$

Файл F2:

Частоты отдельных символов:

$$P(a) = 0.2$$

$$P(b) = 0.4$$

$$P(c) = 0.3$$

$$P(d) = 0.1$$

$$H(P(a), P(b), P(c), P(d)) = (-0.2 * \log[2, 0.2]) + (-0.4 * \log[2, 0.4]) + (-0.3 * \log[2, 0.3]) + (-0.1 * \log[2, 0.1]) = 0.4644 + 0.5288 + 0.5211 + 0.3322 = \mathbf{1.8465}$$

Частоты пар символов:

$$P(aa) = 0.04$$

$$P(bb) = 0.16$$

$$P(cc) = 0.09$$

$$P(dd) = 0.01$$

$$P(ab) = P(ba) = 0.08$$

$$P(ac) = P(ca) = 0.06$$

$$P(ad) = P(da) = 0.02$$

$$P(bc) = P(cb) = 0.12$$

$$P(bd) = P(db) = 0.04$$

$$P(cd) = P(dc) = 0.03$$

$$\begin{aligned} H(P(1), \dots, P(n)) = & ((-0.04 * \log[2, 0.04]) * 3 + (-0.16 * \log[2, 0.16]) + (-0.09 * \\ & \log[2, 0.09]) + (-0.01 * \log[2, 0.01]) + (-0.08 * \log[2, 0.08]) * 2) + (-0.06 * \\ & \log[2, 0.06]) * 2 + (-0.02 * \log[2, 0.02]) * 2 + (-0.12 * \log[2, 0.12]) * 2 + (-0.03 \\ & * \log[2, 0.03]) * 2) / 2 = (0.5574 + 0.4230 + 0.3127 + 0.0664 + 0.583 + 0.487 + \\ & 0.2258 + 0.7342 + 0.3036) / 2 = \mathbf{1.84655} \end{aligned}$$

Результаты работы

| | Оценка энтропии (частоты отдельных символов) | Теоретическое значение энтропии (отдельные символы) | Оценка энтропии (частоты пар символов) | Теоретическое значение энтропии (для пар символов) |
|----|--|---|--|---|
| F1 | 1.9998 | 2 | 1.9972 | 2 |
| F2 | 1.84326 | 1.8465 | 1.84315 | 1.84655 |

Вывод

Реальные значения достаточно близки к теоретическим. При этом энтропия пар символов не сильно отличается от энтропии, полученной при подсчете частот отдельных символов.

Также реальные значения энтропии для пар символов могут немного превышать теоретические ввиду того, что после генерации файла реальные частоты пар символов не всегда могут соответствовать теоретическим частотам.