

УДК 004.627

Грицай Р. – ст. гр. СНМ-51

Тернопільський національний технічний університет імені Івана Пулюя

МЕТОДИ СТИСНЕННЯ ІНФОРМАЦІЇ

Науковий керівник: к.т.н., доцент Козак Р.О.

Hryzaj R.

Ternopil Ivan Pul'uj National Technical University

METHODS COMPRESSION OF INFORMATION

Supervisor: Kozak R.

Ключові слова: стиснення, метод, дані

Keywords: compression, method, data

Методи стиснення даних можна розділити на два типи:

1. Методи без спотворення (loseless) - методи стиснення (інша назва: методи стиснення без втрат) гарантують, що декодовані дані будуть в точності збігатися з вихідними;

2. Методи з втратами (lossy) - методи стиснення (інша назва: методи стиснення з втратами) можуть спотворювати вихідні дані, наприклад за рахунок видалення несуттєвої частини даних, після чого повне відновлення неможливе.

В основу методів стиснення даних без втрат інформації покладено усунення надмірності подання інформації. Економне кодування досягається за рахунок подання малоімовірних подій більш довгими словами, ніж подій з високою ймовірністю настання. Якщо ймовірність настання події дорівнює P , то, відповідно до теореми Шеннона про кодування джерела інформації, таку подію найкраще кодувати словом завдовжки $-\log_2 P$ біт. Методи стиснення даних спираються на цей факт.

В результаті процесу економного кодування одиниці вихідних даних (символу, слова, рядку, числа і т.п.) ставиться у відповідність так зване кодове слово. Кодове слово складається з послідовності цифр, як правило двійкових. Сукупність усіх кодових слів утворює код. Якщо довжини всіх кодових слів однакові, то використовуваний код має фіксовану (постійну) довжину, в іншому випадку - змінну. Якщо вихідні дані можуть бути однозначно відновлені по масиву відповідних кодових слів, то кодування не призводить до втрат інформації.

Ефективність стиснення як характеристика скорочення розміру подання інформації щодо початкового визначається ступенем стиснення. Ступінь стиснення приймається рівною відношенню об'єму вихідних даних до об'єму відповідних їм стиснутих даних і вимірюється в разях.

Всі методи стиснення прийнято розділяти на два класи: методи статистичного кодування і методи словникового стиснення. В схемах стиснення часто використовуються допоміжні перетворення, що забезпечують або сприяють виконанню етапу економного кодування.

В результаті проведених досліджень найбільш поширених алгоритмів стиснення даних можна зробити висновок, що найбільш ефективним методом стиснення даних є метод PPM (Prediction by Partial Matching).