

МЕТОДИ УЩІЛЬНЕННЯ ДАНИХ, ЩО БАЗУЮТЬСЯ НА ОБЧИСЛЕННІ ВІДХИЛЕНЬ

Алексеева Т.М.

Науковий керівник - Каплун В.А.

В даній статті розглядається числова модель джерела інформації. Тобто дані, які підлягають ущільненню, незалежно від їх фактичного змісту, розглядаються як цілі числа. Пропонується класифікація методів ущільнення даних, основою яких є обчислення відхилень. Можна відійти від традиційного підходу: представляти дані для ущільнення не у вигляді потоку символів, а у вигляді додатніх чисел, тобто розбити весь потік вхідної інформації на послідовність 8-, 16-, 32- і т.д. розрядних чисел. У загальному випадку це будуть n -розрядні числа.

Всі способи ущільнення можна розділити на дві категорії: *ущільнення без втрат* (оборотне) і *ущільнення з втратами* (необоротне). Ущільнення без втрат завжди приводить до зниження об'єму вихідного потоку інформації без втрати інформаційної структури.

Можна використати значення сусідніх елементів вхідної послідовності, значення певним чином введених констант чи дискретні значення функцій апроксимації. Таким чином, усі методи ущільнення, що базуються на обчисленні відхилень, поділяються на такі, що враховують статистики, і такі, що їх не враховують.

Окрему групу методів, що враховують числові статистичні характеристики, утворюють методи, які базуються на розбитті значень елементів послідовності Q на *піддіапазони*. Як числа, відхилення від яких обчислюють і зберігають, можуть використовуватись *сусідні елементи* вхідної послідовності, які підлягають ущільненню. При застосуванні методів розбиття вхідного потоку даних в ущільненій послідовності доводиться зберігати інформацію або про належність до певного піддіапазону, або саму характеристику піддіапазону чи групи, або і те, і інше разом. Для обчислення методу відхилення від *середніх значень* у групах не потрібно обчислювати попередньо мінімальну і максимальну границі діапазону вхідних чисел, а взяти їх рівними 0 та $2^n - 1$ відповідно, можна не зберігати значення самих констант – їх значення можуть бути обчислені заздалегідь, і залежать вони від розрядності чисел, на які розбивається вхідний потік даних, і їх можна обчислити ще до здійснення ущільнення або до відновлення початкової інформації. Причому немає необхідності штучно вибирати розташування констант, можна вибирати їх так, щоб вони були рівномірно розкидані по діапазону введених чисел.