

DOI: 10.18372/2310-5461.42.13801

УДК 621.327:681.5 (045)

І. М. Тупиця

Харківський національний університет Повітряних Сил імені Івана Кожедуба

orcid.org/0000-0001-6806-4914

e-mail: ivan20081982@gmail.com

МЕТОДОЛОГІЯ РЕСТРУКТУРИЗАЦІЇ ДАНИХ ІНФОРМАЦІЙНОГО РЕСУРСУ ДЛЯ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ СТАТИСТИЧНОГО КОДУВАННЯ

Вступ

У сучасних алгоритмах кодування для більш вигідного представлення кодованих даних активно використовується «реструктуризація» даних. Під цим поняттям розуміється трансформація вихідних даних в більш зручну форму з метою підвищення ефективності представлення кодованих даних [1–9]. Пропонується проаналізувати підходи до реструктуризації даних, які використовуються у сучасних алгоритмах стиснення інформації, виявити їх переваги та недоліки.

Аналіз останніх досліджень і публікацій

Проаналізувавши останні наукові публікації можна зробити висновок, що на ефективність статистичного кодування даних інформаційного ресурсу впливають такі чинники: характер закону розподілу ймовірностей появи елементів у повідомленні (рівномірний, нерівномірний); стратегія позиціонування кодових конструкцій окремих елементів у загальній кодовій послідовності [10–13].

Так характер закону розподілу ймовірностей появи елементів у повідомленні має значущий вплив на захист даних від несанкціонованого доступу, а саме:

– при рівномірному законі розподілу ймовірностей у разі несанкціонованого доступу кодову послідовність можна розбити на окремі кодові конструкції однакової довжини таким чином, що вони будуть збігатися з реальними кодовими конструкціями. Це дозволить визначити позиції кодових конструкцій у кодовій послідовності і при доступі до таблиці частот забезпечити несанкціоноване декодування елементів кодової послідовності.

– при нерівномірному законі розподілу ймовірностей існує можливість підбору зловмисником з безлічі таблиць частот (законів розподілу ймовірностей) таких варіантів законів розподілу ймовірностей, які присвоюють елементам повідомлення в процесі статистичного кодування однакові коди.

У свою чергу, стратегія позиціонування кодових конструкцій окремих елементів у кодовій послідовності впливає на ефективність статистичного кодування з позиції скорочення довжини для представлення інформації. Так використання при статистичному кодуванні маркерних роздільників та властивості префіксності для визначення позицій кодових конструкцій окремих елементів у загальній кодовій послідовності, призводить до зростання довжини для представлення інформації [14].

Для вирішення вищевказаних проблем необхідно забезпечити виконання таких умов:

1) забезпечити нерівномірний розподіл ймовірностей появи елементів повідомлення таким чином, щоб виключити можливість безпомилкового підбору зловмисником закону розподілу ймовірностей;

2) виключити вплив низькочастотних складових (елементів) повідомлення на формування кодограми з метою скорочення як кодових конструкцій, що присвоюються окремим елементам повідомлення, так і кодовій послідовності в цілому.

З метою вирішення вищезазначених проблем актуальним стає питання дослідження методів реструктуризації даних інформаційного ресурсу для підвищення ефективності статистичного кодування з позиції підвищення захисту та скорочення довжини на подання інформації.

Мета статті (постановка завдання)
Дослідження існуючих та пошук нових підходів до реструктуризації даних інформаційного ресурсу для підвищення ефективності статистичного кодування.

Виклад основного матеріалу

На даний час існують два підходи до проведення реструктуризації даних інформаційного ресурсу:

– перший підхід — зовнішня реструктуризація даних. Зовнішня реструктуризація — перетворення вихідних даних (їх трансформація)

для підвищення ефективності представлення кодованих даних;

– другий підхід — внутрішня реструктуризація даних — виявлення закономірностей у внутрішній структурі даних інформаційного ресурсу.

Пропонується проаналізувати методи зовнішньої реструктуризації, в залежності від етапів, на яких відбувається трансформація даних:

1) на етапі підготовки вихідних даних до кодування;

2) всі елементи кодованих даних піддаються трансформації. Одним з прикладів даного варіанту зовнішньої реструктуризації є дискретно-косинусне перетворення;

3) спільне використання декількох варіантів реструктуризації;

4) на етапі кодування:

– кодування окремих елементів за спільною ознакою;

– кодування окремих елементів за динамічними діапазонами;

– трансформація (перетворення) кодованих даних;

– спільне використання декількох варіантів.

Пропонується проаналізувати застосування методів зовнішньої реструктуризації даних на прикладі формату JPEG, так як в ньому присутні всі вищевказані методи зовнішньої реструктуризації даних.

Так у форматі JPEG залежно від етапу на якому дані піддаються трансформації, використовуються наступні методи зовнішньої реструктуризації даних:

1) на етапі підготовки вихідних даних до кодування:

– дискретно-косинусне перетворення;

– квантування;

– впорядкована лінеаризація.

2) на етапі кодування:

– групове кодування;

– кодування по динамічним діапазонам.

Структурна схема застосування методів зовнішньої реструктуризації даних в форматі JPEG представлена на рис. 1.

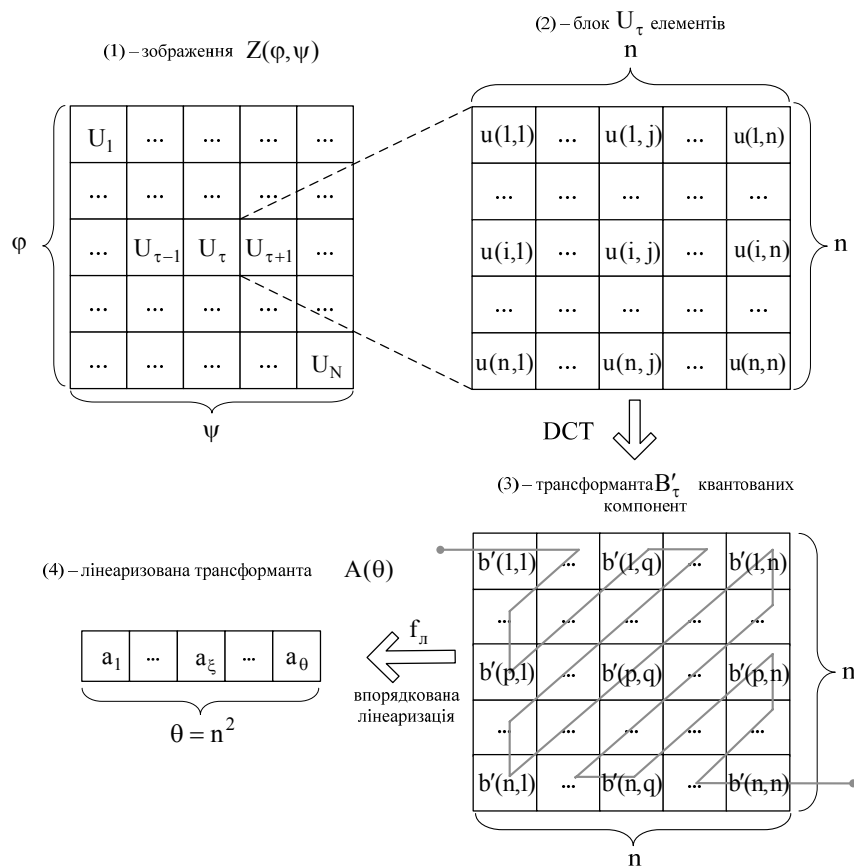


Рис. 1. Структурна схема процесу використання методів зовнішньої реструктуризації даних у форматі JPEG

Аналіз методів зовнішньої реструктуризації даних, які використовуються на етапі підготовки інформаційного ресурсу до кодування

На етапі підготовки вихідних даних до кодування застосовується такий варіант зовнішньої реструктуризації даних — коли всі елементи кодованих даних піддаються трансформації.

Як було зазначено вище, у форматі JPEG даний варіант зовнішньої реструктуризації даних реалізується з використанням наступних методів: дискретно-косинусне перетворення, квантування, впорядкована лінеаризація.

Так одним з методів зовнішньої реструктуризації даних, який застосовується в форматі JPEG на етапі підготовки даних до кодування, є **дискретно-косинусне перетворення** (ДКП або Discrete Cosine Transform). Дискретно-косинусне перетворення являє собою різновид перетворення Фур'є. Дискретне косинусне перетворення дозволяє переходити від просторового представлення зображення до його спектрального подання і назад. Впливаючи на спектральне подання зображення, що складається з «гармонік», тобто, відкидаючи найменш значущі з них, можна балансувати між якістю відновлення і ступенем стиснення [15].

У форматі JPEG пряме ДКП застосовується не до всього зображення $Z(\varphi, \psi)$, а до блоків U_τ елементів $u(i, j)$ розміром $n \times n$ (де $n = 8$) з декількох причин [16]:

1) оскільки зі збільшенням розміру оброблюваного блоку U_τ зображення $Z(\varphi, \psi)$ зростає складність обчислень;

2) використання при ДКП блоків U_τ елементів $u(i, j)$ розміром 8×8 не призводить до істотних втрат якості відновленого зображення.

Двовимірне дискретно-косинусне перетворення, яке застосовується до кожного блоку U_τ елементів $u(i, j)$ задається такою формулою [16]:

$$b(p, q) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} u(i, j) c(p) c(q) \cos \left[\frac{(2i+1)p\pi}{2n} \right] \cos \left[\frac{(2j+1)q\pi}{2n} \right], \quad (1)$$

$$c(p) = \begin{cases} \sqrt{\frac{1}{n}} & p=0; \\ \sqrt{\frac{2}{n}} & p \neq 0. \end{cases}$$

де $b(p, q)$ — (p, q) -а компонента трансформанти B_τ ; $c(p)$ — коефіцієнт функції ДКП.

Результатом ДКП є трансформанти B_τ розміром $n \times n$, які мають такі властивості:

1. Розташовані в лівій верхній частині трансформанти B_τ компоненти $b(p, q)$ несуть інформацію про низькочастотні складові блоків U_τ зображення $Z(\varphi, \psi)$. Відповідно в правій нижній частині трансформанти B_τ розташовані компоненти $b(p, q)$ (АС-коефіцієнти), які несуть

інформацію про високочастотні складові блоків U_τ зображення $Z(\varphi, \psi)$.

2. Значення компонент $b(p, q)$ трансформанти B_τ мають тенденцію до зменшення при скануванні у зигзагнапрямку.

У форматі JPEG з метою зменшення діапазону (розрядності) отриманих після ДКП значень компонент $b(p, q)$ трансформанти B_τ застосовується ще один метод зовнішньої реструктуризації — **квантування**. Квантовані значення компонент $b(p, q)$ трансформанти B_τ (рис. 1, позначення (3)) отримуються шляхом ділення компонент $b(p, q)$ трансформанти B_τ на елементи $y(p, q)$ матриці $Y(n, n)$ коефіцієнтів квантування, що задається наступним виразом:

$$b'(p, q) = \left\lfloor \frac{b(p, q)}{y(p, q)} \right\rfloor. \quad (2)$$

Трансформанта B'_τ володіє такою властивістю, що чим більш віддалені від ДС-коефіцієнта (значення квантованої компоненти $b'(1,1)$ трансформанти B'_τ , розташованої в лівому верхньому кутку) АС — коефіцієнти (інші 63 квантовані компоненти $b'(p, q)$ трансформанти B'_τ), тим з більшою ймовірністю їх значення дорівнює нулю. Під віддаленістю слід розуміти суму індексів (p, q) компоненти $b'(p, q)$ трансформанти B'_τ : чим вона більше, тим далі розташована дана компонента $b'(p, q)$ від компоненти $b'(1,1)$. Слід зазначити, що на етапі квантування йде втрата інформації.

Аналіз застосування в форматі JPEG дискретно-косинусного перетворення свідчить про те, що на етапі підготовки даних інформаційного ресурсу до кодування використання одного конкретного методу зовнішньої реструктуризації даних недостатнє для більш вигідного подання кодованих даних. У даному випадку мова йде про спільне використання декількох методів зовнішньої реструктуризації — дискретно-косинусного перетворення та квантування.

Наступний метод зовнішньої реструктуризації даних, який застосовується у форматі JPEG — **впорядкована лінеаризація**. Даний метод зовнішньої реструктуризації даних у форматі JPEG використовується на заключному етапі підготовки даних до кодування.

У форматі JPEG більшість квантованих значень компонент $b'(p, q)$ трансформанти B'_τ при досить високому ступені стиснення зображення $Z(\varphi, \psi)$ виявляються нульовими.

Для того щоб згрупувати поруч якомога більше нульових компонент $b'(p, q)$ трансформанти B'_τ застосовується впорядкована лінеаризація. Впорядкована лінеаризація застосовується з метою забезпечення спрощення обробки даних та реалізується шляхом використання алгоритму зигзагсканування трансформанти B'_τ , який задається функцією лінеаризації f_n , а саме:

$$B'_\tau \xrightarrow{f_n} A(\theta), \quad (3)$$

де f_n — функція впорядкованої лінеаризації для алгоритму зигзагсканування; $A(\theta)$ — лінеаризована трансформанта B'_τ квантованих значень компонент $b'(p, q)$; θ — кількість компонент $b'(p, q)$ лінеаризованої трансформанти $A(\theta)$, $\theta = n^2$.

У результаті лінеаризації трансформанти B'_τ з використанням алгоритму зигзагсканування формується вектор $A(\theta)$ (рис. 1, позначення (4)), який задається таким виразом:

$$A(\theta) = (a_1 = b'(1, 1); \dots; a_\xi = b'(p, q); \dots; a_\theta = b'(n, n)), \quad (4)$$

де a_ξ — ξ -й елемент вектору $A(\theta)$, $\xi = \overline{1, \theta}$.

Отриманий в результаті лінеаризації трансформанти B'_τ «зигзагоподібний» вектор $A(\theta)$ має такі властивості:

1) при проходженні по ньому зліва направо ймовірність зустріти ненульовий елемент a_ξ (тобто $a_\xi \neq 0$) тим нижча, чим більше нульових елементів a_ξ (тобто $a_\xi \neq 0$) було вже зустрічно [17].

2) елементи a_ξ вектору $A(\theta)$ відсортовані за критерієм просторової частоти $f(a_\xi)$. Це означає, що першому елементу вектору $A(\theta)$ (тобто елементу a_1) відповідає найнижча частота $f(a_1)$. Зі збільшенням індексу ξ елемента a_ξ підвищується його частота $f(a_\xi)$.

Таким чином у результаті лінеаризації трансформанти B'_τ з використанням алгоритму зигзагсканування формується вектор $A(\theta)$ елементи a_ξ якого відібрані за критерієм підвищення просторової частоти $f(a_\xi)$. Таке представлення даних є вигідним для застосування наступного методу зовнішньої реструктуризації даних — **алгоритму групового кодування** (алгоритму RLE), який застосовується на етапі кодування даних.

Слід зазначити, що проаналізовані вище методи зовнішньої реструктуризації даних мають ряд суттєвих недоліків, а саме:

1) збільшується час на обробку даних інформаційного ресурсу;

2) втрата якості відновленого зображення на етапі квантування значень компонент трансформанти;

3) необхідність в спільному використанні декількох методів зовнішньої реструктуризації даних для більш вигідного подання кодованих даних;

4) необхідність у використанні на заключному етапі підготовки до кодування технології позиціонування. Так без використання впорядкованої лінеаризації подальше застосування групового кодування було б неефективним.

Аналіз методів зовнішньої реструктуризації, які використовуються на етапі кодування даних інформаційного ресурсу

Далі пропонується проаналізувати застосування методів зовнішньої реструктуризації даних, які використовуються у форматі JPEG на етапі їх кодування. Зазначені методи реалізуються наступним чином:

1) кодування окремих елементів за спільною ознакою — групове кодування AC-коефіцієнтів;

2) кодування окремих елементів за динамічними діапазонами — кодування DC- і AC-коефіцієнтів за діапазонами значень;

3) трансформація (перетворення) кодованих даних — різницевий код, який застосовується при кодуванні DC-коефіцієнтів;

4) спільне використання декількох варіантів реструктуризації. Так при кодуванні AC-коефіцієнтів застосовується кодування за динамічними діапазонами та групове кодування.

Недоліками застосування методів зовнішньої реструктуризації даних, які застосовуються в алгоритмі JPEG на етапі кодування, є наступні:

– при груповому кодуванні знижується завадостійкість кодової послідовності. Так зміна яскравості внаслідок перешкоди призводить до зміни яскравості всієї послідовності (штрихи уздовж рядків) або до «розсмикування» рядків у випадку зміни довжини серії внаслідок перешкоди [18];

– метод групового кодування неефективний при кодуванні напівтонових зображень, так як не скорочує довжину кодової послідовності, а збільшує її;

– необхідність розрізняти довжину серії від значення кодованого елемента вектору;

– кодування за динамічними діапазонами підвищує завадостійкість кодової послідовності за рахунок використання фіксованих кодових

таблиць, але при цьому ефективність статистичного кодування, з позиції скорочення довжини для подання інформації, залежить від ступеня насиченості зображення. Так для яскравих зображень довжина кодової послідовності зростає за рахунок зростання значень компонент трансформанти. Даний недолік може бути усунений при спільному використанні двох варіантів зовнішньої реструктуризації даних — кодування за динамічними діапазонами та групового кодування;

– спільне використання декількох методів — кодування за динамічними діапазонами та групового кодування дозволяє підвищити ефективність статистичного кодування з позиції скорочення довжини для представлення інформації та підвищити завадостійкість кодової послідовності, але не дозволяє забезпечити захист даних від несанкціонованого доступу.

Для усунення недоліків методів зовнішньої реструктуризації пропонується принципово новий підхід — внутрішня реструктуризація елементів u_ξ повідомлення $U(\theta)$.

Метод внутрішньої реструктуризації даних інформаційного ресурсу реалізується за рахунок

обліку закономірностей у внутрішній структурі елементів u_ξ повідомлення $U(\theta)$ [19–21]. Даний тип реструктуризації передбачає своє використання або спільно з зовнішньою реструктуризацією, або самостійно.

Суть внутрішньої реструктуризації полягає у виявленні закономірностей у внутрішній структурі елементів u_ξ повідомлення $U(\theta)$ за кількісною ознакою λ_i .

В якості внутрішньої структури елемента u_ξ повідомлення $U(\theta)$ пропонується використовувати його двійкове подання $[u_\xi]_2$.

Елемент u_ξ повідомлення $U(\theta)$ складається з послідовності $[u_\xi]_2$ двійкових розрядів $q_{\xi,\alpha}$, $\alpha = \overline{1, |u_\xi|_2}$.

Це задається наступним чином:

$$[u_\xi]_2 = \{q_{\xi,1}; \dots; q_{\xi,\alpha}; \dots; q_{\xi,|u_\xi|_2}\}, \quad (5)$$

де $q_{\xi,\alpha}$ — α -й розряд елемента u_ξ .

На рис. 2 наочно представлена внутрішня структура елемента u_ξ довжиною $[u_\xi]_2$.

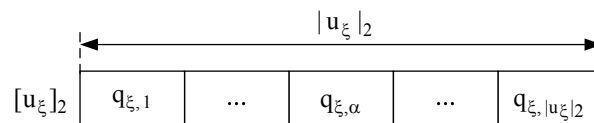


Рис. 2. Внутрішня структура елемента u_ξ

Кількість різних значень, які приймає ознака, тобто потужність ознаки λ_i , може бути n . Набір різних значень ознаки λ_i описується таким виразом:

$$\Lambda = \{\lambda_1, \dots, \lambda_i, \dots, \lambda_n\}, \quad (6)$$

де Λ — набір значень ознаки λ_i , виявленої у внутрішній структурі елементів u_ξ повідомлення $U(\theta)$; λ_i , λ_n — значення i -ї та n -ї ознаки набору Λ .

Метою використання методу внутрішньої реструктуризації даних є подальше групування (кластеризація) елементів u_ξ повідомлення $U(\theta)$, які мають однакові значення кількісної ознаки λ_i у множини $U(\lambda_i)$. Це задається наступним чином:

$$U(\theta) \xrightarrow{f_{cl}} \{U(\lambda_1), \dots, U(\lambda_i), \dots, U(\lambda_n)\}, \quad (7)$$

де f_{cl} — функція групування елементів u_ξ у множини $U(\lambda_i)$ за значенням ознаки λ_i ;

n — кількість множин $U(\lambda_i)$, які формуються в процесі кластеризації елементів u_ξ повідомлення $U(\theta)$.

Слід зазначити, що кластеризація даних дозволяє знизити потужність повідомлення $U(\theta)$ до рівня множини $U(\lambda_i)$ і тим самим забезпечує підвищення ефективності статистичного кодування з позиції скорочення довжини на подання інформації.

Висновки

Використання для більш вигідного представлення кодованих даних методу внутрішньої реструктуризації за кількісною ознакою має ряд переваг у порівнянні з методами зовнішньої реструктуризації, а саме:

- не потребує проведення трансформацій над елементами повідомлення;
- скорочується час, необхідний для обробки даних;

– дозволяє на етапі, що передує кодуванню, провести кластеризацію елементів з однаковими значеннями кількісної ознаки;

– дозволяє підвищити ефективність статистичного кодування з позиції скорочення довжини кодових конструкцій, які присвоюються елементам повідомлення зарахунок зменшення потужності кодованих даних;

– дозволяє підвищити ефективність статистичного кодування з позиції підвищення захисту зарахунок необхідності використання ідентифікаторів для однозначного декодування кодових конструкцій.

Як наслідок, пропонується використовувати метод внутрішньої реструктуризації даних інформаційного ресурсу у методах статистичного кодування для підвищення їх ефективності з позиції скорочення довжини на подання інформації та підвищення захисту інформаційного ресурсу.

Перспективи подальших досліджень

Визначений напрям спонукає на подальшу розробку концепції для формування кількісної ознаки для використання методу внутрішньої реструктуризації даних інформаційного ресурсу у методах статистичного кодування з метою підвищення ефективності статистичного кодування з позиції підвищення захисту та скорочення довжини на подання інформації.

ЛІТЕРАТУРА

1. Сэломон Д. Сжатие данных, изображений и звука: пер. с англ. В. В. Чепыжова. М.: Техносфера, 2004. 368 с.
2. Кудряшов Б. Д. Теория информации. СПб: Питер, 2009. 320 с.
3. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М.: Техносфера, 2005. 1073 с.
4. Miano J. Compressed image file formats: JPEG, PNG, GIF, XBM, BMP / by John Miano, 1999. 264 p.
5. Мандель И. Д. Кластерный анализ. М.: Финансы и Статистика, 1988. 543 с.
6. Pratt W. K., Chen W. H., Welch L. R. Slant transform image coding. Proc. Computer Processing in communications. New York: Polytechnic Press, 1969. P. 63-84.
7. Jain A., Murty M., Flynn P. Data clustering: A review. *ACM Computing Surveys*. 1999. Vol. 31, no. 3. Pp. 264–323.
8. Miano J. Formats and image compression algorithms in action. К.: Triumph, 2013. 336 p.
9. Ding Z., Chen H., Gua Y., Peng Q. GPU accelerated interactive space-time video matting. *In Computer Graphics International*. 2010. P. 163-168.
10. Lee S. Y., Yoon J. C. Temporally coherent video matting. *Graphical Models* 72. 2010. P. 25–33.
11. Воронцов К. В. Алгоритмы кластеризации и многомерного шкалирования. Курс лекций. МГУ, 2007. 145 с.
12. Lazarovych I., Melnychuk S., Kozlenko M. Optimization of entropy estimation computing algorithm for random signals in digital communication devices. *Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, 14th International Conference, 2018. P. 1073–1078.
13. Tso B., Mather P.M. Classification methods for remotely sensed data. US, CRC Press, 2009, 349 p.
14. Grundmann M., Kwatra V., Han M., Essa I. Efficient hierarchical graph based video segmentation. *IEEE CVPR*. 2010. P. 85-91.
15. Zhang Y., Negahdaripour S., Li Q. Error-resilient coding for underwater video transmission. *OCEANS 2016 MTS/IEEE Monterey*, Monterey. CA. 2016. Pp. 1–7.
16. Musienko A., Ganjaric J. Technology of coding of digital aerial photographs taking into account classes of a semantic saturation of blocks in system of air monitoring. *Engineer of XXI Century: VII Inter University Conference of Students, PhD Students and Young Scientists (08 December 2016 at the University of Bielsko-Biala (ATH) / Bielsko-Biala)*, Poland, 2016. pp. 215-220.
17. Barannik V., Sidchenko S., Tupitsya I., Stasev S. Synthesis of combined crypto-compressed systems for providing safety video information in infocommunications. *EWDTs: 2015 IEEE East-West Design & Test Symposium*, Batumi, Georgia, 2015, pp. 1-4. DOI:10.1109/EWDTs.2015.7493145
18. Barannik V., Tupitsya I., Sidchenko S., Tarnopolov R. The method of crypto-semantic presentation of images based on the floating scheme in the basis of the upper boundaries. *Problems of Infocommunications Science and Technology: 2nd International Scientific-Practical Conference PIC S and T 2015 (13 – 15 October 2015, Kharkiv)*, Ukraine. pp. 248-250, DOI:/10.1109/Infocommst.2015.7357326
19. Barannik V., Tupitsya I., Shulgin S., Sidchenko S., Larin V. The application for internal restructuring the data in the entropy coding process to enhance the information resource security. *EWDTs: 2016 IEEE East-West Design & Test Symposium*, Yerevan, Armenia, 2016. pp. 561-565. DOI:10.1109/EWDTs.2016.7807749
20. Бараннік В. В., Тупиця І. М., Бараннік В. В., Сорокун А. Д. Технологія кластеризації даних інформаційного ресурсу за кількісною ознакою ресурса. *Наукоємні технології*. 2018. Вип. 4(40). С. 398-404. DOI: 10.18372/2310-5461.40.13264
21. Barannik V., Tupitsya I., Dodukh O., Barannik V., Parkhomenko M. The Method of Clustering Information Resource Data on the Sign of the Number of Series of Units as a Tool to improve the Statistical Coding Efficiency. *CADSM: 2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (February 26 – March 2 2019, Polyana-Svalyava (Zakarpatya))*, Ukraine. Pp. 3/32–3/36.

Тупиця І. М.

МЕТОДОЛОГІЯ РЕСТРУКТУРИЗАЦІЇ ДАНИХ ІНФОРМАЦІЙНОГО РЕСУРСУ ДЛЯ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ СТАТИСТИЧНОГО КОДУВАННЯ

У сучасних алгоритмах кодування для більш вигідного представлення кодованих даних активно використовується «реструктуризація» даних. Під цим поняттям розуміється трансформація вихідних даних в більш зручну форму з метою підвищення ефективності представлення кодованих даних.

У статті розглядаються питання, пов'язані з розробкою нового підходу до реструктуризації даних з метою підвищення ефективності статистичного кодування з позиції підвищення захисту та скорочення довжини на подання інформації. Досліджуються існуючі методи реструктуризації даних інформаційного ресурсу, які використовуються з метою більш вигідного представлення кодованих даних. Аналізуються недоліки методів зовнішньої реструктуризації даних, які активно використовуються у сучасних алгоритмах кодування інформації. Розроблено принципово новий підхід до реструктуризації даних інформаційного ресурсу – внутрішня реструктуризація, яка полягає у виявленні закономірностей у внутрішній структурі елементів повідомлення за кількісною ознакою. Аналізуються вимоги, що пред'являються до кількісної ознаки. Проводиться порівняльний аналіз існуючих методів реструктуризації даних. Для підвищення ефективності статистичного кодування з позиції скорочення довжини на подання інформації та підвищення захисту інформаційного ресурсу пропонується використовувати метод внутрішньої реструктуризації даних за кількісною ознакою.

Визначений напрям спонукає на подальшу розробку концепції для формування кількісної ознаки для використання методу внутрішньої реструктуризації даних інформаційного ресурсу у методах статистичного кодування з метою підвищення ефективності статистичного кодування з позиції підвищення захисту та скорочення довжини на подання інформації.

Ключові слова: реструктуризація; кількісна ознака; кодування.

Tupitsya I. M.

METHODOLOGY FOR RESTRUCTURING INFORMATION RESOURCE DATA TO IMPROVE THE EFFICIENCY OF STATISTICAL CODING

In modern coding algorithms, "data restructurings" are actively used for a more favorable representation of coded data. Under this notion is the transformation of the source data into a more convenient form in order to increase the efficiency of the representation of coded data.

The article discusses issues related to the development of a new approach to data restructuring in order to improve the efficiency of statistical coding from the standpoint of increasing protection and reducing the length of information presentation. Existing data restructuring methods of the information resource, which are used to better present the coded data, are investigated. The disadvantages of external data restructuring methods that are actively used in modern information coding algorithms are analyzed. A fundamentally new approach to the restructuring of information resource data has been developed - internal restructuring, which is to identify patterns in the internal structure of message elements by a quantitative attribute. The requirements for the quantitative trait are analyzed. A comparative analysis of existing data restructuring methods is carried out. To improve the efficiency of statistical coding from the standpoint of reducing the length of the presentation of information and improve the protection of the information resource, it is proposed to use the method of internal restructuring of data by quantitative attribute.

The determined direction induces the further development of the concept for the formation of a quantitative trait for using the method of internal restructuring of the data resource in statistical coding methods in order to increase the efficiency of statistical coding from the point of increasing security and reducing the length of the presentation of information.

Keywords: restructuring; quantitative attribute; coding.

Тупиця І. М.

МЕТОДОЛОГІЯ РЕСТРУКТУРИЗАЦІЇ ДАНИХ ІНФОРМАЦІЙНОГО РЕСУРСУ ДЛЯ ПОВИЩЕННЯ ЕФЕКТИВНОСТІ СТАТИСТИЧНОГО КОДУВАННЯ

В сучасних алгоритмах кодування для більш вигідного представлення кодованих даних активно використовується «реструктуризація» даних. Під цим поняттям розуміється трансформація вихідних даних в більш зручну форму з метою підвищення ефективності представлення кодованих даних.

В статті розглядаються питання, пов'язані з розробкою нового підходу до реструктуризації даних з метою підвищення ефективності статистичного кодування з позиції підвищення захисту та скорочення довжини на подання інформації.

Исследуются существующие методы реструктуризации данных информационного ресурса, которые используются с целью более выгодного представления кодированных данных. Анализируются недостатки методов внешней реструктуризации данных, которые активно используются в современных алгоритмах кодирования информации. Разработан принципиально новый подход к реструктуризации данных информационного ресурса — внутренняя реструктуризация, которая заключается в выявлении закономерностей во внутренней структуре элементов сообщения по количественному признаку. Анализируются требования, предъявляемые к количественному признаку. Проводится сравнительный анализ существующих методов реструктуризации данных. Для повышения эффективности статистического кодирования с позиции сокращения длины на представление информации и повышения защиты информационного ресурса предлагается использовать метод внутренней реструктуризации данных по количественному признаку.

Определенное направление побуждает на дальнейшую разработку концепции для формирования количественного признака для использования метода внутренней реструктуризации данных информационного ресурса в методах статистического кодирования с целью повышения эффективности статистического кодирования с позиции повышения защиты и сокращения длины на представление информации.

Ключевые слова: реструктуризация; количественный признак; кодирование.

Стаття надійшла до редакції 25.04.2019 р.
Прийнято до друку 12.06.2019 р.