

DOI: 10.18372/2310-5461.40.13264

УДК 621.327:681.5(045)

**В. В. Бараннік**, д-р техн. наук, проф.Харківський національний університет Повітряних Сил імені Івана Кожедуба  
orcid.org/0000-0002-2848-4524  
e-mail: vvbar.off@gmail.com;**І. М. Тулиця**Харківський національний університет Повітряних Сил імені Івана Кожедуба  
orcid.org/0000-0001-6806-4914  
e-mail: ivan20081982@gmail.com;**В. В. Бараннік**Харківський національний університет радіоелектроніки  
orcid.org/0000-0003-3516-5553  
e-mail: valera462000@gmail.com;**А. Д. Сорокун**Національний авіаційний університет  
orcid.org/0000-0001-8469-641X  
e-mail: sorokun@gmail.com

## ТЕХНОЛОГІЯ КЛАСТЕРИЗАЦІЇ ДАНИХ ІНФОРМАЦІЙНОГО РЕСУРСУ ЗА КІЛЬКІСНОЮ ОЗНАКОЮ

### Вступ

У сучасних алгоритмах стиснення продовжує активно застосовуватися статистичний підхід.

Використання методів статистичного кодування даних тісно пов'язане з поняттям «реструктуризація».

Це пов'язано з тим, що на ефективність методів статистичного кодування роблять значущий вплив наступні фактори: характер закону розподілу ймовірностей появи елементів в повідомленні (рівномірний, нерівномірний); позиціонування кодових конструкцій [1–3].

Останнім часом активно використовується концепція зовнішньої реструктуризації даних — трансформація вихідних даних у більш зручну форму з метою підвищення ефективності представлення кодованих даних [4; 5].

Методи зовнішньої реструктуризації, залежно від етапів обробки, на яких відбувається трансформація даних, поділяються на: методи, які можуть використовуватися на етапі підготовки вихідних даних до кодування — коли всі елементи кодованих даних піддаються трансформації, наприклад дискретно-косинусне перетворення; на етапі кодування — коли вихідні дані кодуються за динамічними діапазонами або окремі елементи кодуються за спільною ознакою, наприклад групове кодування (загальна ознака — довжина серії) [6–9].

### Аналіз останніх досліджень і публікацій

Проаналізувавши останні наукові публікації можна сказати, що використання зовнішньої реструктуризації даних має ряд недоліків [10–13]:

1) перетворення (трансформація) вихідних даних потребує додаткових витрат часу, тобто збільшується час обробки даних. Наприклад, застосування дискретно-косинусного перетворення (прямого і відповідно зворотного) потребує додаткових витрат часу, пов'язаних з математичною реалізацією і призводить до втрати якості відновленого зображення через застосування квантування значень компонент трансформанти;

2) об'єднання (трансформація) окремих елементів в групи за спільною ознакою — групове кодування, на фоні скорочення часових витрат та підвищення ефективності статистичного кодування з позиції скорочення довжини на подання інформації, знижує завадостійкість кодової послідовності;

3) кодування за діапазонами підвищує завадостійкість кодової послідовності, але за рахунок використання фіксованих кодових таблиць знижує ефективність статистичного кодування, з позиції скорочення довжини на подання інформації. Довжина кодової послідовності залежить від ступеня насиченості зображення.

Так для сильнаосичених зображень довжина кодової послідовності зростає за рахунок зростання значень компонент трансформанти;

4) спільне використання декількох варіантів зовнішньої реструктуризації даних, наприклад кодування за діапазонами та групового кодування, дозволяє підвищити ефективність статистичного кодування, з позиції скорочення довжини на подання інформації та підвищити поміхостійкість кодової послідовності, але не дозволяє забезпечити захист даних від несанкціонованого доступу;

5) відсутність універсального перетворення даних, яке б гарантувало отримання більш кращої форми представлення даних. Так, у форматі JPEG після дискретно-косинусного перетворення, з метою зменшення діапазону (розрядності) отриманих значень компонент трансформанти, застосовується квантування. В свою чергу, кодування AC- і DC-коефіцієнтів також передбачає спільне використання декількох варіантів зовнішньої реструктуризації даних — групове кодування та кодування за динамічними діапазонами;

6) суцільна префіксність для кодових конструкцій, якими кодуються компоненти трансформанти, дозволяє при несанкціонованому доступі визначити структуру кодових конструкцій, оскільки кодова послідовність, якою кодують квантовані компоненти трансформанти, завжди починається з коду, який присвоюється DC-коефіцієнту і закінчується маркерним роздільником.

З метою усунення вищезазначених недоліків актуальним стає питання пошуку нових підходів реструктуризації даних для підвищення ефективності статистичного кодування з позиції підвищення захисту та скорочення довжини на подання інформації.

**Мета статті (постановка завдання)**

Мета — це проведення дослідження технології кластеризації даних інформаційного ресурсу за ознакою кількості серій одиниць.

**Виклад основного матеріалу**

Пропонується принципово новий підхід реструктуризації даних — внутрішня реструктуризація. Сутність даного підходу полягає у виявленні закономірностей у внутрішній структурі елемента повідомлення за кількісною ознакою [14–16]. Як кількісну ознаку  $\lambda_i$  пропонують використовувати ознаку кількості серій одиниць.

Технологія кластеризації елементів  $u_\xi$  повідомлення  $U(\theta)$  за ознакою  $\lambda_i$  кількості серій одиниць (КСО) полягає в такому — елементи  $u_\xi$ , з однаковим значенням ознаки  $\lambda_i$  КСО, об'єднуються у множини  $U(\lambda_i)$ .

Отже, якщо різні за значенням елементи  $u_i$  і  $u_\xi$  мають однакові значення ознаки  $\lambda_i$  КСО, тобто:

$$u_i \neq u_\xi, \lambda_i(u_j) = \lambda_i(u_\xi),$$

то вони формують множину  $U(\lambda_i)$ , яка задається таким виразом:

$$U(\lambda_i) = \{u_\xi | \lambda_i\}. \tag{1}$$

У результаті повідомлення  $U(\theta)$  розбивається на множини  $U(\lambda_i)$  таким чином, щоб кожен елемент  $u_\xi$  повідомлення  $U(\theta)$  змож належати одній і тільки одній множині розбиття, тобто множини не перетинаються, тобто:

$$\text{якщо } u_\xi \in U(\lambda_i), \text{ то } u_\xi \notin U(\lambda_j), \text{ де } i \neq j,$$

$$\lambda_i, \lambda_j \in \Lambda \text{ і } U(\lambda_j) \cap U(\lambda_i) = \emptyset.$$

Якщо елементи  $u_j$  і  $u_\xi$  мають різні значення ознаки  $\lambda_i$  КСО, то вони належать різним множинам, тобто:

$$\text{якщо } \lambda_j(u_i) \neq \lambda_i(u_\xi), \text{ то } u_i \in U(\lambda_j), u_\xi \in U(\lambda_i).$$

Таким чином, у результаті проведення кластеризації відбувається групування елементів  $u_\xi$  вихідного повідомлення  $U(\theta)$  з однаковими значеннями ознаки  $\lambda_i$  кількості серій одиниць у множини  $U(\lambda_i)$  (рис. 1):

$$U(\theta) \xrightarrow{f_{cl}} \{U(\lambda_1), \dots, U(\lambda_i), \dots, U(\lambda_n)\}, \tag{2}$$

де  $f_{cl}(u_\xi, \lambda_i)$  — функція кластеризації елементів  $u_\xi$  у множини  $U(\lambda_i)$  за значенням ознаки  $\lambda_i$  КСО.

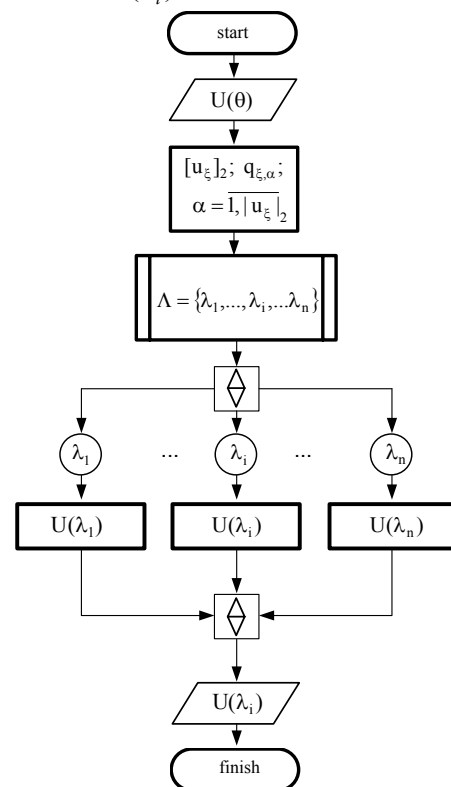


Рис. 1. Алгоритм кластеризації елементів  $u_\xi$  повідомлення  $U(\theta)$  за ознакою  $\lambda_i$  кількості серій одиниць

Набір різних значень ознаки  $\lambda_i$  КСО має такий вигляд:

$$\Lambda = \{\lambda_1, \dots, \lambda_i, \dots, \lambda_n\}, \quad (3)$$

де  $\Lambda$  — набір значень ознаки  $\lambda_i$  кількості серій одиниць, виявленої у внутрішній структурі елементів  $u_\xi$  повідомлення  $U(\theta)$ ;  $\lambda_i$ ,  $\lambda_n$  — значення  $i$ -ї та  $n$ -ї ознаки набору  $\Lambda$ .

Кількість різних комбінацій значень, які може приймати ознака  $\lambda_i$  КСО, задається таким виразом:

$$\eta_{\lambda_i} = \frac{|u_\xi|_2}{2} + 1. \quad (4)$$

де  $\eta_{\lambda_i}$  — кількість комбінацій значень, які може приймати ознака  $\lambda_i$  КСО;  $|u_\xi|_2$  — кількість двійкових розрядів  $q_{\xi,\alpha}$ ,  $\alpha = \overline{1, |u_\xi|_2}$ , якими задається елемент  $u_\xi$  повідомлення  $U(\theta)$ .

Таким чином, набір  $\Lambda$  можливих значень ознаки  $\lambda_i$  КСО, буде мати такий вигляд:

$$\Lambda = \left\{ \lambda_i \mid 0 \leq \lambda_i \leq \frac{|u_\xi|_2}{2} \right\}, \quad (5)$$

Максимальна потужність  $|\Lambda|$  ознаки  $\lambda_i$  КСО, залежить від довжини  $|u_\xi|_2$  послідовності  $[u_\xi]_2$  двійкових розрядів  $q_{\xi,\alpha}$ ,  $\alpha = \overline{1, |u_\xi|_2}$ , якими задається елемент  $u_\xi$  повідомлення  $U(\theta)$ , і задається таким виразом:

$$|\Lambda|_{\max} = \frac{|u_\xi|_2}{2} + 1, \quad (6)$$

де  $|\Lambda|_{\max}$  — максимальна потужність  $|\Lambda|$  ознаки  $\lambda_i$  кількості серій одиниць.

У свою чергу, кількість множин  $U(\lambda_i)$ , яка може формуватися в процесі кластеризації

елементів  $u_\xi$  повідомлення  $U(\theta)$ , обмежується максимальною  $|\Lambda|_{\max}$  потужністю  $|\Lambda|$  ознаки  $\lambda_i$  КСО та має такий вигляд:

$$N(U(\lambda_i)) = |\Lambda|_{\max}, \quad (7)$$

де  $N(U(\lambda_i))$  — кількість множин  $U(\lambda_i)$ , яка може формуватися в процесі кластеризації елементів  $u_\xi$  повідомлення  $U(\theta)$ .

Як було зазначено раніше, кількість різних комбінацій, які може приймати окремий елемент  $u_\xi$  повідомлення  $U(\theta)$ , задається так:

$$\eta_{|u_\xi|_2} = 2^{|u_\xi|_2}, \quad (8)$$

де  $\eta_{|u_\xi|_2}$  — кількість комбінацій, які може приймати елемент  $u_\xi$ .

Можлива кількість комбінацій значень, які може приймати окремий елемент  $u_\xi$  повідомлення  $U(\theta)$  при значенні ознаки  $\lambda_i$  КСО, дорівнює [17–21]:

$$\eta_{|u_\xi|_2, \lambda_i} = \frac{(|u_\xi|_2 + 1)!}{(2\lambda_i)! (|u_\xi|_2 + 1 - 2\lambda_i)!}, \quad (9)$$

де  $\eta_{|u_\xi|_2, \lambda_i}$  — кількість комбінацій значень, які може приймати елемент  $u_\xi$  повідомлення  $U(\theta)$ , при значенні ознаки  $\lambda_i$ .

Для оцінки ефективності проведення кластеризації із застосуванням внутрішньої реструктуризації елементів  $u_\xi$  повідомлення  $U(\theta)$ , за значенням ознаки  $\lambda_i$  КСО, пропонується розглянути варіант розвитку подій, коли елемент  $u_\xi$  має довжину  $|u_\xi|_2$ , що дорівнює 8 біт, тобто:

$$|u_\xi|_2 = 8 \text{ біт.}$$

Результати проведення кластеризації наведені на рис. 2.

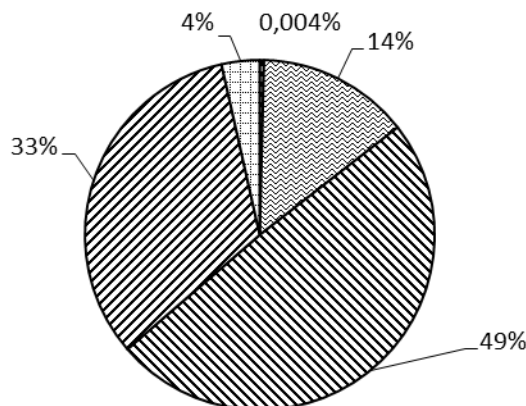


Рис. 2. Результати кластеризації комбінацій значень елемента повідомлення

На рис. 2 прийняті такі позначення:



— множина  $U(\lambda_1)$  комбінацій  $\eta_{|u_\xi|_2, \lambda_1}$  значень елемента  $u_\xi$  зі значенням ознаки  $\lambda_i$  рівним 0, тобто  $\lambda_1 = 0$ ;



— множина  $U(\lambda_2)$  комбінацій  $\eta_{|u_\xi|_2, \lambda_2}$  значень елемента  $u_\xi$  зі значенням ознаки  $\lambda_i$  рівним 1, тобто  $\lambda_2 = 1$ ;



— множина  $U(\lambda_3)$  комбінацій  $\eta_{|u_\xi|_2, \lambda_3}$  значень елемента  $u_\xi$  зі значенням ознаки  $\lambda_i$  рівним 2, тобто  $\lambda_3 = 2$ ;



— множина  $U(\lambda_4)$  комбінацій  $\eta_{|u_\xi|_2, \lambda_4}$  значень елемента  $u_\xi$  зі значенням ознаки  $\lambda_i$  рівним 3, тобто  $\lambda_4 = 3$ ;



— множина  $U(\lambda_5)$  комбінацій  $\eta_{|u_\xi|_2, \lambda_5}$  значень елемента  $u_\xi$  зі значенням ознаки  $\lambda_i$  рівним 4, тобто  $\lambda_5 = 4$ .

Аналізуючи отримані результати можна зробити висновок, що основна частина комбінацій  $\eta_{|u_\xi|_2, \lambda_i}$  значень елемента  $u_\xi$  припадає на множини  $U(\lambda_i)$  зі значеннями ознаки  $\lambda_i$  КСО, які дорівнюють 2 та 3 (49 та 33 % від можливої кількості комбінацій  $\eta_{|u_\xi|_2, \lambda_i}$ ), тобто множини  $U(\lambda_3)$  і  $U(\lambda_4)$ . Відповідно довжини множин  $U(\lambda_3)$  та  $U(\lambda_4)$  для досліджуваного прикладу дорівнюють:

$$|U(\lambda_3)| = 126, |U(\lambda_4)| = 84.$$

Таким чином, кластеризація можливих комбінацій  $\eta_{|u_\xi|_2, \lambda_i}$  значень елементів  $u_\xi$  повідомлення  $U(\theta)$  за значенням ознаки  $\lambda_i$  КСО дозволяє знизити потужність  $|U(\theta)|$  повідомлення  $U(\theta)$  мінімум удвічі (для множини  $U(\lambda_3)$ ).

Аналізуючи технологію кластеризації елементів  $u_\xi$  повідомлення  $U(\theta)$  за ознакою  $\lambda_i$  кількості серій одиниць можна зробити такі висновки:

1. У процесі кластеризації повідомлення  $U(\theta)$  розбивається на множини  $U(\lambda_i)$  таким чином, що елемент  $u_\xi$  повідомлення  $U(\theta)$  може належати тільки одній множині  $U(\lambda_i)$  розбиття, тобто множини не перетинаються.

2. Кількість  $N(U(\lambda_i))$  множин  $U(\lambda_i)$ , які утворюються в процесі кластеризації залежить

від кількості двійкових розрядів  $q_{\xi, \alpha}, \alpha = \overline{1, |u_\xi|_2}$ , якою задається елемент  $u_\xi$ , тобто від його довжини  $|u_\xi|_2$ . Відповідно чим більша довжина  $|u_\xi|_2$  елемента  $u_\xi$ , тим більше множин  $U(\lambda_i)$  може сформуватися в процесі кластеризації, тобто:

$$\begin{aligned} &\text{якщо } |u_\xi|_2 > |u_j|_2, \text{ то} \\ &|\Lambda(\lambda_i, u_\xi)|_{\max} > |\Lambda(\lambda_i, u_j)|_{\max} \Rightarrow \\ &\Rightarrow N(U(\lambda_i), u_\xi)_{\max} > N(U(\lambda_i), u_j)_{\max}, \end{aligned}$$

де  $u_\xi, u_j, \xi$ -й і  $j$ -й — елементи повідомлення  $U(\theta)$ ;  $|u_\xi|_2, |u_j|_2$  — довжина  $\xi$ -го та  $j$ -го елементів повідомлення  $U(\theta)$ ;  $\Lambda(\lambda_i, u_\xi), \Lambda(\lambda_i, u_j)$  — набір  $\Lambda$  ознаки  $\lambda_i$  для  $\xi$ -го та  $j$ -го елементів повідомлення  $U(\theta)$ ;  $|\Lambda(\lambda_i, u_\xi)|_{\max}, |\Lambda(\lambda_i, u_j)|_{\max}$  — максимальна потужність набору  $\Lambda$  ознаки  $\lambda_i$  для  $\xi$ -го та  $j$ -го елементів повідомлення  $U(\theta)$ ;  $N(U(\lambda_i), u_\xi)_{\max}, N(U(\lambda_i), u_j)_{\max}$  — максимальна кількість множин  $U(\lambda_i)$ , які можуть формуватися в процесі кластеризації елементів  $u_\xi$  та  $u_j$ ;

3. Кластеризація елементів  $u_\xi$  повідомлення  $U(\theta)$  буде неефективною для випадка, коли набір  $\Lambda$  ознаки  $\lambda_i$  обмежується одним значенням, тобто:

$$\Lambda = \{\lambda_i | i = 1\}.$$

Це можливо в тому випадку, якщо в результаті кластеризації всі елементи  $u_\xi$  повідомлення  $U(\theta)$  будуть належати тільки одній множині  $U(\lambda_i)$ , тобто:

$$\text{якщо } \forall u_\xi \in U(\lambda_i), \text{ де } i = 1,$$

то потужність  $|U(\theta)|$  повідомлення  $U(\theta)$  буде дорівнювати потужності  $|U(\lambda_i)|$  множини  $U(\lambda_i)$ , тобто:

$$|U(\theta)| = |U(\lambda_i)|,$$

де  $|U(\lambda_i)|$  — довжина множини  $U(\lambda_i)$ , тобто кількість елементів у множині;  $U(\theta)$  — довжина вхідного повідомлення  $U(\theta)$ .

### Висновки

Розглянуто технологію кластеризації даних інформаційного ресурсу, інструментом якої виступає внутрішня реструктуризація за

кількісною ознакою, що дозволяє досягти більш вигідного представлення кодованих даних. Як кількісну ознаку запропоновано використовувати ознаку кількості серій одиниць.

Використання для кластеризації даних методу внутрішньої реструктуризації за ознакою кількості серій одиниць має ряд переваг в порівнянні з зовнішньою, а саме:

- не потребує проведення трансформацій над елементами повідомлення, скорочуючи при цьому час, необхідний для обробки даних;

- дозволяє на етапі, що передує кодуванню, провести кластеризацію елементів, які мають однакові значення кількісної ознаки;

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

- дозволяє підвищити ефективність статистичного кодування з позиції підвищення захисту за рахунок необхідності використання ідентифікаторів для однозначного декодування кодових конструкцій.

Як наслідок, пропонується використовувати досліджувану технологію кластеризації 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 використання зазначеної технології для вдосконалення існуючих методів статистичного кодування даних інформаційного ресурсу.

### ЛІТЕРАТУРА

1. Сэломон Д. Сжатие данных, изображений и звука: пер. с англ. В. В. Чепыжова. — М. : Техносфера, 2004. С. 368.

2. Кудряшов Б. Д. Теория информации. СПб : Питер, 2009. С. 320.

3. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М. : Техносфера, 2005. P.1073.

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.

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. J. Miano. 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.

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), 14<sup>th</sup> 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. and Li Q. Error-resilient coding for underwater video transmission, *OCEANS 2016 MTS/IEEE Monterey*, Monterey, CA, 2016. P. 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, University of Bielsko-Biala (ATH)), Bielsko-Biala, Poland*, 2016. P. 215–220.

17. Barannik V. V., Alimpiev A. N., Sidchenko S. A. The method of cryptocompression presentation of videoinformation resources in a generalized structurally positioned space. *Telecommunications and Radio Engineering [English translation of Elektrosvyaz and Radiotekhnika]*, Vol. 76, 2017 Issue 6 journal-article. P. p. 521–534, DOI: 10.1615/TelecomRadEng.v76.i6.60

18. Barannik V. V., Ryabukha Yu. N., Tverdokhlebo V. V., Baranmk D. V. Methodological basis for constructing a method for compressing of transformants bit representation, based on non-equilibrium positional encoding. *2<sup>nd</sup> IEEE International Conference on Advanced Information and Communication Technologies, AICT 2017, Proceedings, Lviv*, 2017. — P. 188–192. DOI: 10.1109 / AIACT.2017.8020096.

19. Barannik V., Tupitsya I., Shulgin S., Sidchenko S. and Larin V. The application for internal restructuring the data in the entropy coding process to enhance the information resource security.

2016 *IEEE East-West Design & Test Symposium (EWDTS)*, Yerevan, Armenia, 2016. — P. 561–565. DOI:10.1109/EWDTS.2016.7807749

**20. Barannik V., Sidchenko S., Tupitsya I. and Stasev S.,** Synthesis of combined crypto-compressed systems for providing safety video information in info-communications. *2015 IEEE East-West Design & Test Symposium (EWDTS)*, Batumi, Georgia, 2015. P. 1–4. DOI:10.1109/EWDTS.2015.7493145

**21. 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. *2<sup>nd</sup> International Scientific-Practical Conference Problems of Info-communications Science and Technology, PIC S and T 2015* (13–15 October 2015, Kharkiv, Ukraine), P. 248–250, DOI: Org/10.1109/Infocommst.2015.7357326

**Бараннік В. В., Тулиця І. М., Бараннік В. В., Сорокун А. Д.  
ТЕХНОЛОГІЯ КЛАСТЕРИЗАЦІЇ ДАНИХ ІНФОРМАЦІЙНОГО РЕСУРСУ ЗА КІЛЬКІСНОЮ  
ОЗНАКОЮ**

*У статті розглянуті питання, пов'язані з розробкою та дослідженням технології кластеризації даних інформаційного ресурсу з використанням методу внутрішньої реструктуризації за кількісною ознакою - ознакою кількості серій одиниць для більш вигідного представлення кодованих даних. Проаналізовано недоліки існуючих методів реструктуризації даних — методів зовнішньої реструктуризації. З метою усунення недоліків методів зовнішньої реструктуризації даних інформаційного ресурсу пропонується принципово новий підхід — внутрішня реструктуризація. Інструментом кластеризації виступає внутрішня реструктуризація даних, яка полягає у виявленні закономірностей у внутрішній структурі елементів повідомлення за кількісною ознакою. В якості кількісної ознаки використовується ознака кількості серій одиниць. Аналізуються вимоги, які пред'являються до технології кластеризації даних інформаційного ресурсу. Проводиться порівняльний аналіз існуючих методів реструктуризації даних. Для підвищення ефективності статистичного кодування з позиції підвищення захисту інформаційного ресурсу і скорочення довжини на подання інформації пропонується використовувати технологію кластеризації даних інформаційного ресурсу за ознакою кількості серій одиниць.*

**Ключові слова:** кластеризація; внутрішня реструктуризація; ознака кількості серій одиниць.

**Barannik V. V., Tupitsya I. M., Barannik V. V., Sorokun A. D.  
THE CLUSTERING TECHNOLOGY OF INFORMATION RESOURCE DATA ON A  
QUANTITATIVE ATTRIBUTE**

*In modern algorithms, the compressed prodovuzhe actively zastosovuvatsya statistical pidhid. Victory methods of the statistical code danich are determined by understanding of the restructuring. The article deals with issues related to the development and research of information resource clustering technology using an internal restructuring method based on a quantitative attribute — a sign of the number of series of units for a more advantageous representation of the encoded data. The shortcomings of the existing data restructuring methods — external restructuring methods are analyzed. By means of statistics on the availability of technology for clustering resources for information on the quality of gray odinits. To eliminate the shortcomings of the methods of external restructuring of information resource data, a fundamentally new approach is proposed — internal restructuring. The clustering tool is internal data restructuring, which consists in identifying patterns in the internal structure of message elements by a quantitative attribute. As a quantitative trait, the trait of the number of series of units is used. The requirements that are imposed on the information resource clustering technology are analyzed. A comparative analysis of existing data restructuring methods is carried out. To improve the efficiency of statistical coding from the standpoint of increasing the protection of an information resource and reducing the length of information presentation, it is proposed to use the information resource clustering technology based on the number of series of units. The perspective of those who are directly involved in the display of technology is meant for technological comprehension of practical methods of statistical methods.*

**Keywords:** clustering; internal restructuring; a sign of the number of series units.

**Баранник В. В., Тупица И. М., Баранник В. В., Сорокун А. Д.  
ТЕХНОЛОГИЯ КЛАСТЕРИЗАЦИИ ДАННЫХ ИНФОРМАЦИОННОГО РЕСУРСА ПО  
КОЛИЧЕСТВЕННОМУ ПРИЗНАКУ**

*В статье рассмотрены вопросы, связанные с разработкой и исследованием технологии кластеризации данных информационного ресурса с использованием метода внутренней реструктуризации по количественному признаку — признаку количества серий единиц для более выгодного представления кодируемых данных. Анализируются недостатки существующих методов реструктуризации данных — методов внешней реструктуризации. Для устранения недостатков методов внешней реструктуризации данных информационного ресурса предлагается принципиально новый подход — внутренняя реструктуризация. Инструментом кластеризации является внутренняя реструктуризация данных, которая заключается в выявлении закономерностей во внутренней структуре элементов сообщения по количественному признаку. В качестве количественного признака используется признак количества серий единиц. Анализируются требования, которые предъявляются к технологии кластеризации данных информационного ресурса. Проводится сравнительный анализ существующих методов реструктуризации данных. Для повышения эффективности статистического кодирования с позиции повышения защиты информационного ресурса и сокращения длины на представление информации предлагается использовать технологию кластеризации данных информационного ресурса по признаку количества серий единиц.*

**Ключевые слова:** кластеризация; внутренняя реструктуризация; признак количества серий единиц.

Стаття надійшла до редакції 30.09.2018 р.

Прийнято до друку 11.12.2018 р.