УДК 681.223

І.А. Пилипенко
Івано-Франківський ПВНЗ «Галицька академія», м. Івано-Франківськ, Україна
inna_mail@yahoo.com

Метод стиснення даних за допомогою багаторівневих кодів Галуа

У даній роботі розглянуто багаторівневі коди Галуа та можливість їх застосування для стиснення та шифрування текстових даних, що мають явні переваги над іншими базисами в зв'язку з мінімальною надлишковістю даних.

Вступ

У даний час обсяги інформації, що підлягає зберіганню, мають стійку тенденцію зростання. Для зберігання таких обсягів інформації необхідне її попереднє стиснення з подальшою можливістю відновлення і обробки.

Характерною особливістю більшості типів даних є їх надлишковість. Для людини надлишковість даних часто пов'язана з якістю інформації, оскільки вона, як правило, покращує зрозумілість та сприйняття інформації. Однак коли мова йде про зберігання та передачу інформації засобами комп'ютерної техніки, то надлишковість відіграє негативну роль, оскільки вона приводить до зростання вартості зберігання та передачі інформації. Особливо актуальною є ця проблема у випадку необхідності обробки величезних обсягів інформації при незначних об'ємах носіїв даних. У зв'язку з цим постійно виникає проблема позбавлення надлишковості або стиснення даних.

Основний принцип, на якому базується стиснення даних, полягає в економічному описі повідомлення, згідно з яким можливе відновлення початкового його значення з похибкою, яка контролюється.

При стисненні даних деякими методами існує можливість одночасного шифрування або без додаткових обчислювальних затрат, або з низькими затратами. Оскільки в більшості випадків файли, які передаються комп'ютерними мережами або зберігаються в пам'яті, представлені в стиснутому вигляді, то було б доцільно надати можливість користувачам виконувати при стисненні файлів і їх шифрування [1].

Метою статті є представлення багаторівневих кодів Галуа з можливістю їх подальшого використання для стиснення та шифрування інформації з мінімальними втратами.

Для досягнення даної мети були поставлені **наступні задачі**:
– розглянути способи утворення послідовності Галуа;
– вивчити можливості кодів Галуа;
– представити приклад використання кодів для стиснення та шифрування даних.

1. Двійкові коди Галуа

Двійкові коди зазвичай представляються в базисі Радемахера, об'єм коду якого дорівнює $V = N\log_2 N$. Недоліком даного базису є надлишковість та низька завадозахищеність.

Для того щоб уникнути надлишковості кодів, використовують кодову послідовність Галуа, яка утворюється за допомогою базису Галуа. Об'єм даного коду дорівнює V = N. Як бачимо, він більш ефективний.

Коди Галуа утворюються згідно з наступною формулою:

$$G_{i+1} = G_i \oplus G_{i-n}, \quad (1)$$

де n – довжина ключа [2].

Наступний біт в базисі Галуа формується таким чином:

1. Вибираємо розмір ключа.
2. Перший і останній біти ключа додаємо за модулем 2.
3. Зсуваємось на одиницю і додаємо наступний біт ключа з новоутвореним бітом.
4. Формування послідовності Галуа відбувається доти, поки вона не буде дорівнювати: $2^n - 1$, де n – розмір ключа.

Розглянемо приклад формування послідовності Галуа при використанні ключа, довжина якого дорівнює 4 бітам. Послідовність Галуа буде мати наступний вигляд (рис. 1).

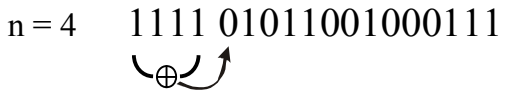

Рисунок 1 – Формування коду Галуа при n = 4

Перевагами базису Галуа є максимальне стиснення інформації, висока захищеність від помилок, а також уникнення надлишковості інформації.

2. Багаторівневі коди Галуа

Багаторівневими називаються послідовності з основами p > 2. Як основа можуть виступати прості числа: 3, 5, 7, 11, 13, 17, 19, 23…

Довжина послідовності визначається за формулою:

$$L = p^n - 1, \quad (2)$$

де p – основа послідовності, n – довжина ключа.

Сама послідовність буде утворюватися згідно з формулою:

$$G_{i+1} = (G_i \times a_{n-1} + G_{i-1} \times a_{n-2} + ... + G_{i-(n-1)} \times a_0) \mathrm{mod} P, \quad (3)$$

де $a_0, a_1, \ldots, a_n$ – ключ, G_i – і-й біт Галуа.

Багаторівнева послідовність в базисі Галуа формується таким чином:

– вибираємо розмір ключа;
– кожен біт ключа множимо на попередні біти Галуа і додаємо отримані значення за модулем P;
– зсуваємось на одиницю і повторюємо процедуру з бітами ключа та з новоутвореним бітом;
– формування послідовності Галуа відбувається доти, поки вона не буде дорівнювати: $p^n - 1$, де n – розмір ключа, p – основа послідовності.

Пошук ключів для формування багаторівневих послідовностей є доволі трудомістким процесом. Зазвичай це відбувається методом перебору.

Розглянемо приклад формування послідовності Галуа з основою три з використанням ключа, довжина якого дорівнює 3 бітам. Довжина цієї послідовності дорівнює $L = 3^3 - 1 = 26$. Шляхом перебору знайдено ключі: 201, 210, 212, 221. Для фор-

мування послідовності скористаємося ключем 201. Перші три біти візьмемо довільно. Нехай це будуть числа 012. Четвертий біт за формулою (3) буде утворюватися наступним чином:

$G_4 = (G_3 \times a_2 + G_2 \times a_1 + G_1 \times a_0) \bmod P = (2 \times 1 + 1 \times 0 + 0 \times 2) \bmod 3 = 2 \bmod 3 = 2.$

В загальному вигляді приклад формування багаторівневої послідовності Галуа з основою 3 представлено на рис. 2.

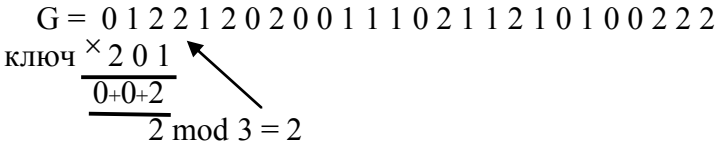

Рисунок 2 – Формування багаторівневих кодів Галуа з основою 3

Як і двійкові коди Галуа, багаторівневі послідовності можна використовувати для стиснення інформації, захисту від помилок та уникнення надлишковості.

3. Метод кодування даних багаторівневими кодами Галуа

Розглянемо можливість застосування багаторівневих кодів Галуа для стиснення алфавітно-цифрових даних. Для цього скористаємось двійковою кодовою таблицею Галуа [3], в якій запишемо послідовність, що представлена на рис. 2. Задана послідовність більш за все підходить для стиснення інформації. Розділимо всі символи на чотири групи: латинські літери, кирилиця, числа, знаки пунктуації (табл. 1).

Таблиця 1 – Багаторівнева кодова таблиця Галуа з поділом на групи

0	1	2	2	1	2	0	2	0	0	1	1	1	0	2	1	1	2	1	0	1	0	0	2	2	2	0	1
		x	t	h	e	s	i	n	d	o	r	a	m	p	l	y	c	u	b	j	f	g	w	k	v	.	_
		q	z	д	к	й	м	в	р	н	я	т	у	ш	ц	л	г	ч	з	п	и	ь	б	ф	щ	.	_
		ю	ж	ъ	ы	э	є	ї	ё	=	+	–	0	1	2	3	4	5	6	7	8	9	*	/	^	,	;
		:	?	!	“	№	%	(	)	~	`	“	#	$	&	|	\	<	>	{	}	[	]	_			

Алгоритм стиснення інформації багаторівневими кодами Галуа з використанням наведеної вище таблиці буде виглядати наступним чином:

1. Зчитуємо перший символ даних.

2. Якщо перший символ, який ми кодуємо, знаходиться в першій групі, записуємо число «0» і переходимо до п. 3; якщо в другій групі – «1» і переходимо до п. 3; якщо в третій або четвертій групі – «2» і переходимо до п. 4.

3. Якщо символ є маленькою літерою – записуємо число «0»; великою літерою – «1»; і переходимо до п. 5.

4. Якщо символ знаходиться в третій групі – записуємо число «0», в четвертій групі – «1».

5. Записуємо три числа – цифру, що знаходиться над літерою в кодовій таблиці і дві попередні цифри.

6. Зчитуємо наступний символ.

7. Якщо символ знаходиться в таблиці після попереднього, записуємо цифру, що знаходиться над ним в кодовій таблиці; переходимо до п. 6.

8. Якщо символ знаходиться в таблиці в тій самій групі, що і попередній – збільшуємо на 1 цифру, що знаходиться в таблиці після попереднього символу; переходимо до п. 5.

9. Якщо символ знаходиться в таблиці в іншій групі, ніж попередній – збільшуємо на 2 цифру, що знаходиться в таблиці після попереднього символу.

10. Якщо група, до якої відноситься символ, є наступною після групи попереднього символу – записуємо цифру 0; якщо вона знаходиться через одну групу – записуємо цифру 1; якщо вона знаходиться через дві групи – записуємо цифру 2; переходимо до п. 5.

11. Кожні 5 цифр, що утворилися при кодуванні, переведемо в десяткове число, яке, в свою чергу, через ASCII-код запишемо у вигляді символа. Цей символ заносимо в новий файл.

Розглянутий алгоритм ефективний для інформації, в якій часто зустрічаються однакові символьні поєднання (такою властивістю, наприклад, володіє текстова інформація).

Даний метод крім стиснення можна використовувати і для шифрування даних. Ефект використання кодів Галуа при криптографічному перетворенні полягає в тому, що забезпечується маскування природної частотної статистики початкової мови повідомлення, оскільки конкретний знак з алфавіту А може бути перетворений на декілька різних знаків шифрувального алфавіту В.

Спробуємо даний метод для кодування символів з різних груп. Так, слово Windows98 буде зашифровано наступним чином:

Таблиця 2

Вихідний текст	W	$\overline{G}$	i	n	d	o	$\overline{G}$	w	$\overline{G}$	s	$\overline{G}$	Зміна групи	9	$\overline{G}$	8
Кодування тексту	01002	0	202	0	0	1	2	002	0	120	1	1	101	1	210

Згрупуємо отриману послідовність по п’ять символів, перетворимо її в десяткові числа та представимо у вигляді символів. В результаті всіх цих перетворень отримаємо:

Таблиця 3

Закодований текст	01002	02020	01200	20120	11101	12100
ASCII-код	29	60	45	177	118	144
Зашифрований текст	↔	<	–	▒	v	P

Дане слово займає 9 символів. Після упаковки його за допомогою багаторівневої послідовності Галуа воно займатиме 6 символів. Отже, розмір слова зменшився в 1,5 рази. Крім того, задане повідомлення було надійно зашифроване іншими символами.

Висновки

У роботі описані двійкові та багаторівневі послідовності Галуа, наведено метод стиснення алфавітно-цифрових даних за допомогою багаторівневої послідовності Галуа.

Даний метод є більш ефективний для одноалфавітних текстових даних з максимальною кількістю символів, розташованих поруч в кодовій таблиці.

Послідовності в базисі Галуа мають явні переваги над іншими базисами в зв'язку з мінімальною надлишковістю даних. Завдяки цьому використання кодів Галуа при розробці системи стиснення інформації, одержаної в результаті експериментальних досліджень, є найперспективнішим.

Література

1. Орищенко В.И. Сжатие данных в системах сбора и передачи информации // Орищенко В.И., Санников В.Г., Свириденко В.А. ; [под ред. В.А. Свириденко]. – М. : Радио и связь, 1995. – 184 с.
2. Пилипенко І.А. Проблеми розвитку методів стиснення масивів даних на основі рандомізації та теоретико-числового базису Галуа / І.А. Пилипенко, Н.Я. Возна, Я.М. Николайчук // Міжнародний науково-технічний журнал «Оптико-електронні інформаційно-енергетичні технології» / Вінницький національний технічний університет. – Вінниця, 2006. – С. 40-47.
3. Пилипенко І.А. Архіватори знань про одновимірні та двовимірні об'єкти діагностування / І.А. Пилипенко // Штучний інтелект. – 2008. – № 4. – С. 418-423.

И.А. Пилипенко

Метод сжатия данных с помощью многоуровневых кодов Галуа

В данной работе рассмотрены многоуровневые коды Галуа и возможность их применения для сжатия и шифрования текстовых данных, которые имеют явные преимущества над другими базисами в связи с минимальной избыточностью данных.

I.A. Pylypenko

Metod of the Data Compression by Multi-Levels Galoua Codes

Modern development status of science and technique is characterized by all greater growth of information streams, which it is often necessary to process and transmit. Consequently, before placing information in archives or transmission on communication channels with the purpose of reduction of their size or disk space, there is the necessity of data compression. One of effective methods of this problem decision is abbreviation of natural surplus. In the given work are considered multi-levels Galoua codes and possibility of their application for text data compression.

Стаття надійшла до редакції 26.05.2009.