

Д.т.н., професор Дичка І.А., магістрант Онай М.В.

**Національний технічний університету України
«Київський політехнічний інститут»**

СПОСІБ УЩІЛЬНЕННЯ АЛФАВІТНО-ЦИФРОВИХ ДАНИХ ПРИ ЇХ ПОДАННІ У ГРАФІЧНО-КОДОВАНОМУ ВИГЛЯДІ

Вступ

Графічне кодування даних є одним з видів автоматичної ідентифікації. Для подання даних у графічно-кодованому вигляді використовуються зазвичай двокольорові (чорно-білі) елементарні графічні форми (елементи) – круг, еліпс, прямокутник (штрих), квадрат тощо. Об'єднуючи ці форми певним чином у групи та розміщуючи їх на площині з дотриманням відповідних правил і структури отримують так звані графічні коди (ГК). Графічні коди бувають двох видів – лінійні та двовимірні [1, 2].

Постановка задачі

В сучасних системах автоматичної ідентифікації вимагається, щоб ГК подавав кількасот алфавітно-цифрових символів. Забезпечити виконання цієї вимоги можна за рахунок підвищення інформаційної щільності даних графічного коду за його незмінних геометричних розмірів. Досягти цього можна розробивши ефективний спосіб ущільнення алфавітно-цифрових даних, що підлягають поданню у графічно-кодованій формі.

Таким чином, актуальною є задача розробки методики подання алфавітно-цифрових даних у вигляді двовимірних графічних кодів з підвищеними показниками інформаційної щільності.

Ущільнення даних

У графічно-кодованому вигляді подають будь-які послідовності алфавітно-цифрових символів. Вважатимемо, що один алфавітно-цифровий символ передає 8 біт інформації.

Розглянемо структурний спосіб підвищення інформаційної щільності графічних кодів, що ґрунтується на поділі алфавіту на набори символів.

Алфавіт, що використовується у вхідних потоках даних, поділимо на 3 набори (табл. 1) – набір літер кирилиці (“К”), набір латинських літер (“Л”) та набір спеціальних символів (“СС”). Наприкінці кожного набору додамо символи-перемикачі, що дозволяють переходити з одного набору символів до іншого. Зокрема, ЗСС – зсув з поточного набору до набору спеціальних символів (для кодування одного символа), ПЛ – перехід до набору латинських символів з поточного набору, ПК – перехід до набору символів кирилиці з поточного набору, ПСС – перехід до набору спеціальних символів з поточного набору.

Таблиця 1. Поділ алфавіту на набори символів

Значення символу	Набір символів					
	Набір «К»		Набір «Л»		Набір «СС»	
	Символ	Код ASCII	Символ	Код ASCII	Символ	Код ASCII
0	А	192	А	65	0	48
1	Б	193	В	66	1	49
...
33	пропуск	160	space	160	\	92
34	ЗСС		ЗСС		€	136
35	ПЛ		ПК		ПК	
36	ПСС		ПСС		ПЛ	

Поставимо за мету кодувати кожні два суміжні алфавітно-цифрові символи (двосимвольну пару) у вхідному потоці даних одним ГК-знаком.

Нехай кожний з наборів алфавіту налічує n_s символів (для нашого прикладу n_s дорівнює 37).

Утворимо числовий еквівалент двосимвольної пари, який обчислимо за формулою: $d = an_s + b$, де a – числове значення лівого (старшого) символу пари; b – правого (молодшого) символу пари. Величину d називають кодословом [3].

Таким чином, символній парі буде поставлено у відповідність число – кодослово. Вхідному алфавітно-цифровому потоку даних (вхідному повідомленню) буде відповідати послідовність чисел – кодослів, а кожне кодослово (d) буде відображатись на носії відповідним графічно-кодовим знаком, що складається з L модулів ($L = 14$ на рис. 1). Модуль – це безрозмірна величина, якій кратні всі розміри графічного коду.

Номери	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Ширина	1	1	1	1	1	1	4				4			

Рис. 1. Приклад графічно-кодового знаку

Перетворення вхідного повідомлення слід здійснювати за два проходи.

Вважатимемо початковим набором символів набір «К». Аналізуємо повідомлення посимвольно зліва направо.

Під час першого проходу повідомлення розбиваємо на групи суміжних символів, до яких входять лише символи з одного набору. Перед черговою групою вставляємо відповідний символ-перемикач. Якщо поточною групою є група символів з набору «К» або «Л», а наступною є група до складу якої входить лише один символ з набору «СС», то перед цим символом вставляють символ зсувач ЗСС.

Під час другого проходу отриману послідовність символів розбивають зліва направо на символні пари, і для кожної з них обчислюють значення кодослова.

Наприклад, вхідному повідомленню «ДИНАМО (КУІV)» відповідатимуть 7 кодослів: 157, 592, 572, 1256, 1046, 896, 806. ГК-позначка для цього прикладу матиме вигляд як на рис. 2.



Рис. 2. ГК-позначка

Оцінимо ступінь ущільнення вхідних алфавітно-цифрових даних.

При перетворенні повідомлення під час першого проходу його довжина може зрости на p символів-перемикачів, а після другого проходу воно буде перетворено на $\lfloor \frac{L+p}{2} \rfloor$ кодослів, кожному з яких відповідатиме ГК-знак завдовжки L модулів. Тоді коефіцієнт ущільнення U вхідного повідомлення становитиме:

$$U = \frac{8k}{\lfloor \frac{L+p}{2} \rfloor} = \frac{16k}{L+p}, \quad (1)$$

де $L \geq 8, p < k$.

Ущільнення має місце, якщо $U > 1$, тобто коли

$$\frac{16k}{L + p} > 1. \quad (2)$$

Визначимо, при якій частоті k/p появи символів-перемикачів у вхідному повідомленні досягається ущільнення.

Вираз (2) перетворимо до такого вигляду:

$$\frac{k}{p} > \frac{L}{16 - L}.$$

Наприклад, якщо $L = 14$, то $k/p > 7$ або $p < k/7$. Отже, при $L = 14$ ущільнення досягається тоді, коли символ-перемикач зустрічається в середньому не частіше, ніж після кожних 8-ми символів вхідного повідомлення.

Набори символів, які представлені в табл. 1 містять 94 різні алфавітно-цифрові символи, але у вхідних повідомленнях можуть зустрічатися будь-які символи таблиці розширеного ASCII. Тому для кодування довільних вхідних наборів будемо використовувати 3 режими: текстовий режим, ASCII режим та цифровий режим.

Текстовий режим призначений для кодування підпоследовностей, утворених з символів, що входять до табл. 1. Режим ASCII – для підпоследовностей, утворених з символів ASCII, що не входять до табл. 1. Цифровий режим – для підпоследовностей, що складаються з десяткових цифр. Для переходів з одного режиму до іншого використовують спеціальні службові кодослова.

Дослідивши ASCII режим можна дійти висновку, що максимальне ущільнення досягається, коли п'ятьом суміжним символам ASCII ставляться у відповідність чотири кодослова. Аналогічно, дослідивши цифровий режим доходимо висновку, що найбільшого ущільнення последовностей десяткових чисел можна досягти, якщо трьом суміжним десятковим цифрам ставити у відповідність одне кодослово.

Висновки

Ущільнення алфавітно-цифрових даних досягається за рахунок використання кількох режимів кодування вхідних повідомлень, поділу алфавіту на набори символів в основному режимі – текстовому, та застосуванні числового перетворення групи символів одного алфавіту (алфавіту вхідних повідомлень) в інший алфавіт – символіку графічного коду. Це перетворення має характер перетворення числа з однієї системи числення в іншу. У прикладі, що розглядався, маємо перетворення з системи числення за основою 37 (текстовий режим), 256 (режим ASCII), 10

(цифровий режим) у систему числення за основою 1373. Таке перетворення забезпечує ущільнення даних в середньому на 10%.

В подальшому слід дослідити ступінь ущільнення даних для випадку матричних графічних кодів.

Література

1. *Арманд В. А., Железнов В. В.* Штриховые коды в системах обработки информации. – М.: Радио и связь, 1989. – 92 с.
2. Two-Dimensional Codes. – Data Capture Institute, 1995. – 30 p.
3. *Ted Williams.* Data Matrix is – Lazerlight Systems, Inc, 1990. – 22 p.