

ТЕХНОЛОГІЯ КОДУВАННЯ СУЧАСНИХ ШТРИХ-КОДІВ

Skubak O. M., Mokryntsev O. A. Advanced coding technology of barcodes. The systematic analysis of various modern barcodes used for labeling is considered and conducted. An linear and two dimensional barcode symbolics are offered. The paper reviewed and analyzed the features, capabilities and benefits of each of these barcodes. Analysis of the modern technology of bar codes leads to the conclusion that the promising trend is the use of barcodes UPC (Universal Product Code) and more modern, comfortable and most reliable for reading code ITF (Interleaved Two of Five). The issue of bar code development includes the following items: the development of modern types of devices for reading barcodes, coding technologies, algorithms barcode recognition; pretreatment, segmentation or allocation boundaries of barcodes lines, scanning, decoding and recognition partially damaged barcodes.

Keywords: barcode, UPC, Universal Product Code, ITF, Interleaved Two of Five, linear barcode, two-dimensional barcode

Скубак О. М., Мокринцев О. А. Технологія кодування сучасних штрих-кодів. Розглянуто та проведено системний аналіз різних сучасних штрих-кодів, що застосовуються для маркування товарів. Наведено лінійні та двовимірні символики штрих-кодів. Питання розвитку штрих-кодів включає в себе наступні проблеми: розробку приладів для зчитування штрих-кодів, технологій кодування, алгоритмів розпізнавання штрих-кодів; питання попередньої обробки, сегментації або виділення меж ліній штрих-кодів, сканування, декодування та розпізнавання пошкоджених штрих-кодів.

Ключові слова: штрих-код, UPC, універсальний товарний код, ITF, чергування два з п'яти, лінійний штрих-код, двовимірний штрих-код

Скубак А. Н., Мокринцев А. А. Технология кодирования современных штрих-кодов. Рассмотрено и проведено системный анализ различных современных штрих-кодов, которые применяются для маркирования товаров. Приведены линейные и двухразмерные символики штрих-кодов. В настоящее время сферу производства и оборота товаров невозможно представить без существования штрих-кодов. Вопрос развития штрих-кодов включает в себя проблемы: разработку приборов для считывания штрих-кодов, технологий кодирования, алгоритмов распознавания штрих-кодов; предварительной обработки, сегментизации или выделения границ линейный штрих-кодов, сканирование, декодирование и распознавание поврежденных штрих-кодов.

Ключевые слова: штрих-код, UPC, универсальный товарный код, ITF, чередование два с пяти, линейный штрих-код, двухразмерный штрих-код

Вступ та постановка задачі

Штрих-код представляє собою оптичне представлення даних, що наноситься на поверхню, маркування або упаковку виробів, що представляє можливість зчитування технічними засобами [1]. Розрізняють лінійні та двовимірні символики штрих-кодів.

Лінійними (одномірними) на відміну від двовимірних називаються штрих-коди, що читаються в одному напрямку (по горизонталі). Найбільш розповсюджені лінійні символики: EAN, UPC, Code39, Code128, Codabar, Interleaved 2 of 5. Лінійні символики дозволяють кодувати невеликий об'єм інформації (до 20–30 символів — зазвичай цифр) за допомогою нескладних штрих-кодів, що читаються недорогими сканерами. Приклад коду символики EAN13 представлено на Рис. 1.



Рис. 1. Лінійний штрих-код.

Пізніше були розроблені так звані двовимірні (2D) коди, які перетворилися в набір прямокутників, точок, шестикутників і інших геометричних фігур (Рис. 2). В залежності від конкретного типу, в середньому лінійні штрих-коди можуть містити інформацію про 20-25 символів, а 2D коди вміщують до 2000 символів.

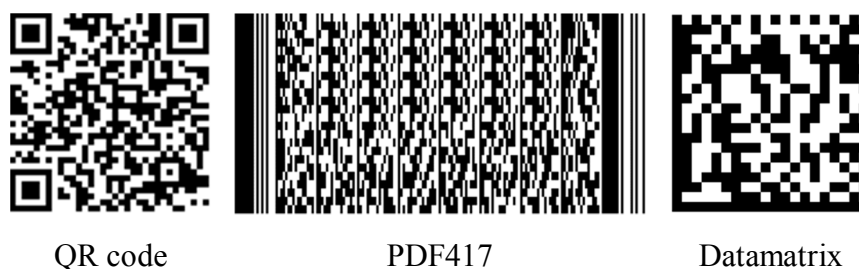


Рис. 2. Двовимірні штрих-коди.

Штрихове кодування винайшов молодий інженер Девид Коллінз [1]. Після закінчення в 1950-х роках інженерного факультету Массачусетського технологічного інституту він поступив працювати на Пенсільванську залізницю, де йому довелося зіткнутися з кропіткою рутинною роботою – сортуванням вагонів. Їх треба було перерахувати, оперативно з'ясувати номери, поратися по них в документації, визначити, куди кожен вагон повинен прослідкувати. Йому прийшла ідея освітлювати номери вагонів прожекторами і зчитувати їх за допомогою фотоелементів. Для спрощення розпізнавання, він записував номери не тільки звичайними цифрами, але й спеціальним кодом, що складається з червоних і синіх смуг, розташованих на стінці вагона в прямокутнику довжиною до півметра.

Випробування підтвердили: скануючий пристрій здатний правильно зчитувати коди навіть при швидкості руху вагона близько 100 км/год. Успіх підштовхнув його до подальшого вдосконалення системи. У 1968 році він використовував замість прожекторів, що витрачали багато енергії, жорстко сфокусований лазерний промінь. Розміри скануючої установки стали набагато менше, зменшилося і саме кодове маркування, що навело Коллінза на думку використовувати придуманий ним штриховий код не тільки на залізниці, але і в якості товарного коду. Так штрих-код отримав визнання і в супермаркетах.

У 1973 році в США було створено «Універсальний товарний код» (UPC – Universal Product Code), що пропагує використання штрих-кодів в промисловості і торгівлі. А з 1977 року в Західній Європі для ідентифікації споживчих товарів стала застосовуватися аналогічна система під назвою «Європейський артикул» (EAN – European Article Numbering). Саме ця установа і займається присвоєнням штрих-кодів. Сьогодні вона включає 98 організацій з 100 країн світу. Виробник може отримати штрих-код для свого товару, попередньо зареєструвавшись у цієї асоціації.

Всього приблизно за сорок років технологія штрих-кодування зробила крок так далеко, що сьогодні вдається зчитувати код за допомогою світлової плями діаметром всього в чверть міліметра. Штриховий код дозволяє зчитувати інформацію про номер товару практично миттєво і з високою точністю.

Сьогодні широкий діапазон персональних переносних пристроїв загального призначення, таких як мобільні телефони, оснащений цифровими оптичними камерами. У поєднанні з новими послугами, це може докорінно змінити повсякденний досвід здійснення покупок, включаючи цінову перевірку в магазині, інформацію про продукт, доступ до відгуків про продукцію, визначення місцезнаходження подібних продуктів і послуг і цінове порівняння на місці. Розпізнавання штрих-кодів у зображеннях, зробленими такими пристроями, може бути особливо складним через обмеження інтегрованої камери і можливостей цифрової обробки. Вони часто оснащені низькоякісними лінзами та спрощеною системою обробки зображень в порівнянні з більш якісними цифровими фотоапаратами чи сканерами. Оптична система часто не призначена для того, щоб зняти

об'єкт близько до лінзи, що призводить до погано сфокусованих і освітлених зображень з ненасиченими кольорами. Ці обмеження часто означають що методи, що використовують звичайну порогову обробку не дають задовільних результатів при розпізнаванні. Алгоритм, який розробляється для широкого спектру переносних пристроїв, повинен прийняти ці обмеження до уваги.

Ставиться задача: розглянути та проаналізувати, яким чином зберігається інформація у деяких найбільш поширених штрих-кодах.

Найбільш поширені сучасні штрих-коди.

UPC або **Universal Product Code** (універсальний код товару) (Рис. 3.) – американський стандарт штрих-коду [2], призначений для відстеження товарів в магазинах. UPC був розроблений в 1973 році Джорджем Джозефом Лорером (George Joseph Lauger), інженером корпорації IBM. Штрих-код UPC, що містить 12 цифр, є попередником європейського вдосконаленого коду EAN-13, що кодує 13 цифр. Код UPC є окремим випадком коду EAN-13. Код UPC перетвориться в код EAN-13 дописуванням нуля перед дванадцятьма цифрами коду UPC. Саме з цієї причини товари США і Канади в європейському кодуванні починаються з нуля.

Код UPC – практично симетричний лінійний штрих-код. Симетричність і висока перешкодозахищеність обумовлена недостатньо розвинутою цифровою технікою часів створення цих кодів. Код складається з 2 груп цифр, по 6 чисел в кожній групі – лівої і правої. Групи цифр обрамляються так званими захисними, або огорожувальними, штрих-шаблонами (Guard Patterns). Ці шаблони містять штрихи одиначної ширини, які служать для синхронізації сканера штрих-коду. Наявність саме трьох таких полів обумовлено в першу чергу можливим нанесенням штрих-коду на закруглену поверхню.

І якщо зараз це не є значною проблемою, то в часи створення цього коду сканеру потрібно було знати ширину одиначного штриха на початку, середині і наприкінці коду. Ліві і праві захисні шаблони складаються з 3 штрихів одиначної ширини – двох темних і одного світлого між ними. Середній захисний шаблон складається з 5 штрихів – трьох світлих і двох темних. Решта – цифри [3].



Рис. 3. Штрих- код UPS.

Кожна цифра лівої чи правої групи кодується за допомогою чотирьох штрихів: двох світлих і двох темних (Табл. 1). Кожен штрих може мати відносну ширину в одну, дві, три або чотири одиниці.

Табл. 1

Тиха зона	Старт	Ліва група цифр									Середина	Права група цифр									Фініш	Тиха зона		
		0	1	2	3	4	5	6	7	8	9		0	1	2	3	4	5	6	7	8	9		

Загальна ширина штрихів для однієї цифри завжди становить сім одиниць. Бітова комбінація для кожної цифри розроблена таким чином, щоб цифри, наскільки це можливо, відрізнялися один від одного. Максимальна довжина темного або світлого ділянки не може перевищувати чотирьох одиниць. Загальна ширина всього коду завжди дорівнює 95 одиницям. У будь-якому коді 29 світлих і 30 темних штрихів. Всі ці технічні рішення дуже важливі для надійності і простоти сканування цього коду.

Перша цифра коду, так званий префікс, має деяке логічне значення, але не настільки важлива з технічної точки зору. Остання цифра - контрольне число, служить для виявлення можливої помилки при читанні коду сканером або ручного введення цифр коду з клавіатури.

Напрямок читання комбінації штрихів значення не має, код спеціально розроблявся так, щоб він однаково просто зчитувався як у прямому, так і зворотному напрямку (якщо товар перегорнули). Також не має значення те, яке виконання має весь штрих-код – фотографічно позитивне або негативне. Тобто штрих-код, нанесений світлими смужками по темному фону читається так само, як і темними смужками по світлому фону. Мало того, кольори штрихів і фону не обов'язково повинні бути білими і чорними, можливі й інші кольорові комбінації.

При проектуванні структури коду, в умовах ще не надто розвиненої електроніки, було важливим зробити його якомога простішим для зчитування сканером щоб спростити апаратну частину самого сканера та здешевити його. У числі складнощів була проблема ймовірного зчитування коду в зворотному напрямку, тобто зчитування коду на товарі, який касир підніс до сканера «догори ногами». Тому дуже важливо було, щоб чергування смуг було однаковим в обох напрямках – спочатку темний штрих, потім білий, потім знову темний і так далі. Ще було б бажано, щоб положення захисних шаблонів було завжди на одному і тому ж місці.

Рішення було знайдено. Можна звернути увагу, що код виглядає дуже симетрично, тобто кількість штрихів праворуч і ліворуч від центру завжди однакові, а ширина правої і лівої частини штрих-коду однакові. Тобто механізм зчитування штрих-коду завжди однаковий, як цей код не поверни.

Що стосується однакової послідовності чергування світлих і темних штрихів при прямому і зворотному читанні, то розробники домоглися цього тим, що кодування правої і лівої груп цифр трохи відрізняється – праві символи мають фотографічно негативний накреслення щодо лівих. Тобто шаблони штрихів для однієї і тієї ж цифри ідентичні, але позитивні чи негативні. Інакше кажучи, відмінність тільки в тому, що якщо для лівої частини коду це світлий штрих, то для правої – темний.

Проблема розпізнавання прямого або зворотного зчитування точно так само легко вирішується (Табл. 2). Наприклад, якщо сканер зчитує цифру з товщиною штрихів 3-2-1-1, то він розуміє, що це цифра «нуль» і це її пряме зчитування, а якщо він зчитує штрихи товщиною 1-1-2-3, то він розуміє, що це той же «нуль», але зчитаний у зворотному напрямку.

Табл. 2

Цифра	Ліва група	Права група	Товщина ліній
0	0001101	1110010	3-2-1-1
1	0011001	1100110	2-2-2-1
2	0010011	1101100	2-1-2-2
3	0111101	1000010	1-4-1-1
4	0100011	1011100	1-1-3-2
5	0110001	1001110	1-2-3-1
6	0101111	1010000	1-1-1-4
7	0111011	1000100	1-3-1-2
8	0110111	1001000	1-2-1-3
9	0001011	1110100	3-1-1-2

Зчитуючи числа, закодовані дзеркально щодо звичайної кодування, сканер розуміє, що весь штрих-код зчитується в зворотному напрямку, отже, і всю отриману послідовність з 12 цифр потрібно передати комп'ютеру в зворотному порядку.

При зчитуванні коду правильність зчитування перевіряється таким способом:

- підсумовуються всі парні цифри і множаться на 3;
- підсумовуються всі непарні цифри включаючи контрольну цифру;
- ці суми складаються і залишається остання цифра від результату.

Технічно цифри обробляються послідовно, за один прохід, з множенням кожної цифри на 1 або 3, в залежності від парності позиції, додаванням до суми і взяттям залишку по модулю 10 від поточної суми. Іншими словами десятки відразу відкидаються, що спрощує механізм обчислення.

Якщо результат дорівнює нулю, то приймається рішення, що код сканований правильно, якщо будь-яка інша цифра, то код однозначно зчитаний невірною.

Більш сучасним, зручним, а головне надійним для зчитування кодом є **ITF** [4, 5], або чергування 2 з 5 (від Interleaved Two of Five) (Рис. 4), спосіб потокового кодування цифр за допомогою двох типів ширин ліній, що утворюють групи по 5 чорних ліній з 5 пробілами між ними. Цей код застосовується комерційно, наприклад у маркуванні 135 плівок, для штрих-кодів ITF-14, на картонних оптових упаковках деяких продуктів (в той час як продукти всередині марковані UPC або EAN) та в багатьох інших місцях.



Рис. 4. Штрих-код ITF.

В штрих-коді **ITF** цифри кодуються парами. Перша цифра закодована в п'яти чорних лініях, у той час як друга цифра закодована в п'яти білих проміжках якими чергуються ці чорні лінії. Два з кожних п'яти барів або місць широкі (отже точно 2 з 5). Цифри кодуються наступним чином (Табл. 3):

Табл. 3

Цифра	Лінія або проміжок, ширина	Мнемоніка (див. вагу)
0	<i>n</i> <i>n</i> <i>W</i> <i>W</i> <i>n</i>	4+7=11, замість 0
1	<i>W</i> <i>n</i> <i>n</i> <i>n</i> <i>W</i>	1+0=1
2	<i>n</i> <i>W</i> <i>n</i> <i>n</i> <i>W</i>	2+0=2
3	<i>W</i> <i>W</i> <i>n</i> <i>n</i> <i>n</i>	1+2=3
4	<i>n</i> <i>n</i> <i>W</i> <i>n</i> <i>W</i>	4+0=4
5	<i>W</i> <i>n</i> <i>W</i> <i>n</i> <i>n</i>	1+4=5
6	<i>n</i> <i>W</i> <i>W</i> <i>n</i> <i>n</i>	2+4=6
7	<i>n</i> <i>n</i> <i>n</i> <i>W</i> <i>W</i>	7+0=7
8	<i>W</i> <i>n</i> <i>n</i> <i>W</i> <i>n</i>	1+7=8
9	<i>n</i> <i>W</i> <i>n</i> <i>W</i> <i>n</i>	2+7=9
Вага	1 2 4 7 0	

В Табл. 3 "*n*" – вузький інтервал (лінія або проміжок) і "*W*" – широкий (у 2.0 — 3.0 рази). Широкі лінії формуються парами з послідовними вагами 1, 2, 4, 7, та 0, де коду 0 відповідає значення 11. Перед кодованою інформацією присутній шаблон початку коду, що складається з *nnnn* (вузька лінія – вузький проміжок – вузька лінія – вузький проміжок), після всіх символів має бути шаблон кінця послідовності, що складається з *Wnn* (широка лінія – вузький проміжок – вузька лінія). Оскільки цифри задаються парами, може бути закодовано

тільки парне число цифр. Як правило, до непарного числа цифр просто додається "0" як перша цифра або використовуються п'ять вузьких проміжків (у цьому випадку слід зазначити що ця остання пара символів візуально буде більш вузькою ніж решта пар).

Висновок

Робота містить системний аналіз різних сучасних штрих-кодів, що застосовуються для маркування товарів. Наведено лінійні та двовимірні символи штрих-кодів, розглянуто та проаналізовано особливості, можливості та переваги кожного з даних штрих-кодів.

Аналіз розвитку технології сучасних штрих-кодів приводить до висновку, що перспективним напрямком являється застосування наступних штрих-кодів:

– **UPC** або **Universal Product Code** (універсальний код товару) — американський стандарт штрих-коду, призначений для відстеження товарів в магазинах;

– більш сучасний, зручний, а головне надійний для зчитування код **ITF**, або чергування 2 з 5 (від Interleaved Two of Five), в основі якого лежить спосіб потокового кодування цифр за допомогою двох типів ширин ліній, що утворюють групи по 5 чорних ліній з 5 пробілами між ними.

Питання розвитку штрих-кодів включає в себе наступні проблеми: розробку сучасних типів приладів для зчитування штрих-кодів, технологій кодування, алгоритмів розпізнавання штрих-кодів; питання попередньої обробки, сегментації або виділення меж ліній штрих-кодів, сканування, декодування та розпізнавання частково пошкоджених штрих-кодів.

Література

1. Pavlidis T. Fundamentals of bar code information theory / T. Pavlidis, J. Swartz, Y. P. Wang // Computer. – Apr. 1990. – Vol. 23. – No. 4. – PP. 74-86.
2. Ouaviani E. A common image processing framework for 2d barcode reading / E. Ouaviani, A. Pavan, M. Bottazzi, E. Brunelli, F. Caselli, M. Guerrero // Proc. Seventh Int. Conf on Image Processing and Its Applications (Conf Publ.). – July 1999. – Vol. 2. – No. 465. – PP. 652-655.
3. Joseph E. Bar code waveform recognition using peak locations / E. Joseph, T. Pavlidis // IEEE Trans. Pattern Analysis and Machine Intelligence. – June 1994. – Vol. 16. – No. 6. – PP. 630-640,
4. Waveform recognition with application to bar codes // Proc. IEEE Int. Conf on Systems, Man, and Cybernetics. Decision Aiding for Complex Systems, Conference Proceedings. – 1991, –Vol. 1. – PP. 129-134.
5. Hahn H. I. Implementation of algorithm to decode two-dimensional barcode PDF-417 / H. I. Hahn, J. K. Joung // Proc. IEEE 6th Int. Conf on Signal Processing. – Aug. 2002. – Vol. 2. – PP. 1791-1794.

Автори статті

Скубак Олександр Миколайович – кандидат технічних наук, доцент кафедри прикладного програмування, Державний університет телекомунікацій, м. Київ. Тел. +380 (67) 403 99 90. E-mail: skubaksp@rambler.ru

Мокринцев Олексій Анатолійович – аспірант кафедри прикладного програмування, Державний університет телекомунікацій, м. Київ. Тел +380 (99) 428 03 71.

Authors of the article

Skubak Oleksandr Mykolayovych – candidate of sciences (technic), assistant professor at applied programming department, State University of Telecommunications. Tel.: +380 (67) 403 99 90. E-mail: skubaksp@rambler.ru.

Mokryntsev Oleksiy Anatoliyovych – postgraduate student at applied programming department, State University of Telecommunications. Tel.: +380 (99) 428 03 71.

Дата надходження в редакцію: 11.01.2016 р.

Рецензент: д.т.н., проф. Барабаш О. В.