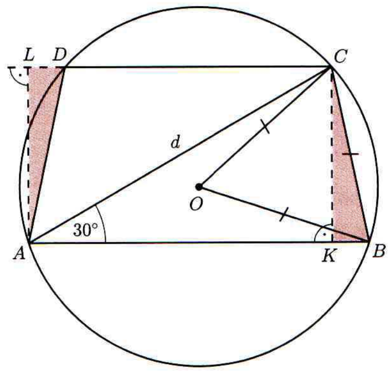
Kody na co dzień

Andrzej DĄBROWSKI

Rozwiązanie zadania M 1073.
Niech $ABCD$ będzie danym trapezem (o podstawach $AB \geqslant CD$), a punkt O – środkiem okręgu opisanego na tym trapezie. Wówczas trójkąt BOC jest równoboczny, a zatem

$$\sphericalangle BAC = \frac{1}{2}\sphericalangle BOC = 30^\circ.$$

Zadanie sprowadza się więc do wyznaczenia pola trapezu $ABCD$, znając długość przekątnej d oraz wiedząc, że $\sphericalangle BAC = 30^\circ$.



Niech K i L będą odpowiednio rzutami prostokątnymi punktów C i A na proste AB i CD. Wówczas pole trapezu $ABCD$ jest równe polu prostokąta $AKCL$, które wynosi $\frac{1}{4}\sqrt{3}\,d^2$.

Obserwowałem niedawno pracę kasjerki, która przesuwając szybkim ruchem towar pod czytnikiem, błyskawicznie podsumowała monstrualne zakupy mojego sąsiada. Uświadomiłem sobie, że jeszcze parę lat temu musiałaby ręcznie wpisywać nazwę towaru wraz z ceną, co oczywiście trwałoby o wiele dłużej. Człowiekiem, który dokonał gigantycznej rewolucji w obrocie towarowym na całym świecie był George J. Laurer.

George J. Laurer – potomek niemieckich emigrantów, właścicieli browaru w Binghampton w stanie Nowy Jork. Był pracownikiem koncernu IBM, gdzie wyróżnił się autorstwem wielu wynalazków.

W maju 1973 zaakceptowano jego projekt, nazwany UPC (*Universal Product Code*). Każdy produkt opisany był 11-cyfrowym kodem, zakończonym dodatkową cyfrą. Po pewnym czasie dopisano na początku kodu jeszcze jedną cyfrę – kod kraju. Dzięki temu kod UPC, nazwany w Europie EAN-13 (*European Article Number*), rozpoczął swój triumfalny pochód przez świat.

Historia zanotowała nawet pierwszą transakcję z użyciem kodu UPC i czytnika. Towarem, podobno wyciągniętym losowo z kosza 26 lipca 1974 roku, było opakowanie zawierające 10 sztuk owocowej gumy do żucia Wrigley's Juicy Fruit. Historyczne opakowanie gumy uwidoczniono na specjalnej planszy w Smithsonian Institution's National Museum of American History.

Szybko przetłumaczono kod cyfrowy na zapis paskowy, aby można było wykorzystać istniejące czytniki optyczne. Dodatkowa, ostatnia cyfra kodu została wprowadzona po to, aby system sam wykrywał najczęściej popełniane błędy.

Z badań statystycznych wynika, że około 80 procent popełnianych błędów to zamiana jednej cyfry na inną (np. częsta zamiana 3 i 8). Dziesięć procent błędów to z kolei przestawienie dwóch sąsiadujących cyfr (w Polsce nazywane czeskim błędem). Dobry system zabezpieczający kod winien zatem przede wszystkim wykrywać te najczęściej występujące błędy. Cyfra kontrolna, wprowadzona przez Laurera w kodzie EAN-13, została tak wybrana, aby liczba

$$a_{13} + 3a_{12} + a_{11} + \ldots + 3a_2 + a_1$$

była wielokrotnością liczby 10. Kod kreskowy *Delty* spełnia ten warunek: $9+3\cdot7+7+3\cdot0+1+3\cdot3+7+3\cdot3+0+3\cdot0+0+3\cdot0+7=70.$
Kod EAN-n zawierający n-cyfrowe słowo kodowe $a_n, a_{n-1}, \ldots, a_2, a_1$ jest szczególnym przykładem tzw. *kodu liniowego modulo p*. Ostatnia cyfra takiego kodu jest cyfrą kontrolną, a jej wartość jest ustalana tak, by wyrażenie

$$w_n a_n + w_{n-1}a_{n-1} + \ldots + w_2 a_2 + w_1 a_1$$

dawało resztę r z dzielenia przez p dla odpowiednio ustalonych współczynników wagowych $w_1, w_2, \ldots, w_n$. Kody liniowe modulo p spotykamy na każdym kroku.

Dla kodu EAN-13 mamy: $n = 13$, $p = 10$, $r = 0$. Współczynniki wagowe są tu równe 1 dla wyrazów nieparzystych i 3 dla parzystych.

Identyfikację czasopisma umożliwia kod ISSN (*International Standard Serial Number*) – ośmiocyfrowy kod liniowy modulo 11 z wagami $1, 2, 3, 4, 5, 6, 7, 8$ oraz $r = 0$. Cyfra kontrolna czasami ma wartość 10 – jest wówczas kodowana jako X. Kod ISSN jest zawsze tłumaczony na kod EAN-13 – pierwsze 3 cyfry z lewej (prefiks) to symbol czasopism 977, do nich dołączanych jest 7 cyfr początkowych słowa kodu ISSN uzupełnione na końcu dwoma zerami. Na przykład, kodem ISSN dla Delty jest 0137-3005 (sprawdź, że cyfra kontrolna jest poprawna). W systemie EAN-13 ma on postać 977 0137300 00 7 – tutaj ostatnia cyfra kontrolna jest wyznaczona według reguły kodowania EAN-13.

ISSN 0123-456X

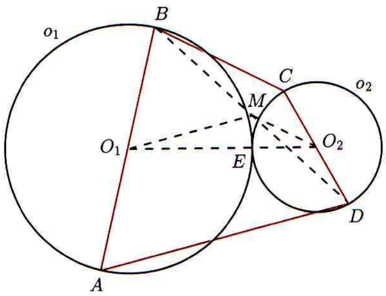
9 770123 456008

ISBN 83-01-01373-7

9 788301 013738

Odpowiednikiem kodu ISSN dla książek jest kod ISBN (*International Standard Book Number*). ISBN jest dziesięciocyfrowym kodem liniowym modulo 11 z wagami $1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ oraz $r = 0$. Cyfra kontrolna równa 10 jest kodowana jako X. ISBN zbudowany jest z trzech części. Pierwsza oznacza język albo kraj pochodzenia. Druga część oznacza wydawcę, a trzecia numer książki.

Rozwiązanie zadania M 1074.
Niech O_1 i O_2 będą odpowiednio środkami okręgów o_1 i o_2. Niech ponadto E będzie punktem styczności danych okręgów, a punkt M – środkiem odcinka BD.



Wówczas
$$AD + BC = 2O_1M + 2O_2M \geqslant$$
$$\geqslant 2O_1O_2 = 2O_1E + 2O_2E = AB + CD.$$

Kody kreskowe występujące na książkach to zakodowany numer ISBN w standardzie EAN-13. Prefiksem symbolizującym książki jest 978. Kolejne 9 cyfr stanowi początkowe cyfry kodu ISBN. Cyfra kontrolna jest wyznaczona według reguły kodowania EAN-13.

Książki polskie mają symbol 83, książki w języku angielskim symbol 0 lub 1, francuskim – 2, niemieckim – 3. Jeden z dłuższych kodów ma Andora – 99913.

Każdy obywatel Polski ma nadany kod PESEL. Jest to 11-cyfrowy kod liniowy modulo 10 z wagami $1, 3, 1, 9, 7, 3, 1, 9, 7, 3, 1$ oraz $r = 0$. Kodami liniowymi są również NIP (Numer Identyfikacji Podatkowej) oraz REGON (Rejestr Gospodarki Narodowej).

Kod PESEL składa się z 11 cyfr. Pierwszych sześć to data urodzenia w układzie rr-mm-dd. Jeżeli dana osoba nie urodziła się w XX wieku, to zaszyfrowane jest to w kodzie miesiąca. Dla osób urodzonych w XIX wieku oznaczenie miesiąca zwiększa się o 80 (styczeń – 81, luty – 82, …, grudzień – 92). Dla osób, które urodzą się w XXI wieku, liczba miesiąca będzie powiększona o 20; (styczeń – 21, luty – 22, …, grudzień – 32) itd. Kolejne cztery cyfry służą do rozróżnienia ludzi urodzonych tego samego dnia. Czwarta z nich dodatkowo oznacza płeć: nieparzysta – mężczyznę, parzysta – kobietę. Ostatnia cyfra kodu jest cyfrą kontrolną.

Wagi i podstawa modulo NIPu oraz REGONu kodów są tajne i należy zwrócić się do GUSu z prośbą o ich udostępnienie. Dociekliwi internauci doświadczalnie, próbując różne wagi i podstawy modulo dla poprawnych kodów (w końcu jest to układ równań z kilkoma niewiadomymi) już dawno odsłonili, zresztą nieskomplikowane, reguły tworzenia tych kodów.

Prawdopodobieństwo niewykrycia błędu zamiany w kodzie liniowym

Przypuśćmy, że j-ta cyfra w kodzie liniowym została zastąpiona przez cyfrę x i nie zostało to wykryte. Oznacza to, że wyrażenie
$$w_n a_n + \ldots + w_{j+1}a_{j+1} + w_j x + w_{j-1}a_{j-1} \ldots + w_1 a_1$$
ma resztę r przy dzieleniu przez p, taką samą jak wyrażenie
$$w_n a_n + \ldots + w_{j+1}a_{j+1} + w_j a_j + w_{j-1}a_{j-1} \ldots + w_1 a_1.$$
Wynika stąd, że przez p podzielna jest różnica napisanych wyżej liczb, tzn. liczba $w_j(x - a_j)$. Jeśli jednak w_j nie ma z p wspólnego dzielnika większego od 1 (tak jest w kodzie EAN, gdzie $w_j = 3$ albo $w_j = 1$, a $p = 10$), to przez p dzielić się musi liczba $(x - a_j)$. Ale to nie jest możliwe, bo liczba $|x - a_j|$, jako różnica dwóch cyfr, jest mniejsza od liczby 10. Widzimy więc, że kody typu EAN zawsze wykrywają błąd zamiany cyfry. Podobnie uzasadniamy, że kod PESEL oraz kody ISSN i ISBN wykrywają taki błąd.

W związku z wejściem Polski do Unii Europejskiej, byliśmy świadkami ożywionej dyskusji o systemie IACS (Integrated Animal Counting System). W systemie IACS zwierzęta są zaopatrzone w kolczyki z informacjami: symbol kraju (PL), znak graficzny Agencji Restrukturyzacji i Modernizacji Rolnictwa, kod cyfrowy składający się z numeru w serii (dwie cyfry), numeru zwierzęcia (9 cyfr) i cyfry kontrolnej. Kolczyk ma tzw. część żeńską i męską. Część żeńska zawiera wszystkie informacje wraz z kodem kreskowym części cyfrowej (na rysunku kod cyfrowy jest postaci 00 501093711 7).

część żeńska część męska

Zdecydowano, po wielu perturbacjach, aby cyfra kontrolna kodu wyznaczona była w standardzie EAN-12. (ARiMR w przetargu na dostawę kolczyków zapomniała dopisać, w jaki sposób ma być liczona cyfra kontrolna. Trzech zwycięzców przetargu w różny sposób, niektórzy bardzo skomplikowanie, kodowali informacje, co uniemożliwiało stworzenie jednolitej bazy danych. Błąd ten poprawiono dopiero po pewnym czasie.)

Prawdopodobieństwo niewykrycia błędu przestawienia w kodzie liniowym

Przypuśćmy, że j-ta cyfra w kodzie liniowym została zamieniona miejscami z $j+1$-szą cyfrą i nie zostało to wykryte. Oznacza to, że wyrażenie
$$w_n a_n + \ldots + w_{j+1}a_j + w_j a_{j+1} + w_{j-1}a_{j-1} \ldots + w_1 a_1$$
ma resztę r przy dzieleniu przez p, taką samą jak wyrażenie
$$w_n a_n + \ldots + w_{j+1}a_{j+1} + w_j a_j + w_{j-1}a_{j-1} \ldots + w_1 a_1.$$
Wynika stąd, że przez p podzielna jest różnica napisanych wyżej liczb, tzn. liczba $(w_j - w_{j+1})(a_{j+1} - a_j)$. Dla kodów ISSN i ISBN jest to niemożliwe, gdyż któryś z czynników iloczynów musiałby być podzielny przez 11, ale obie różnice są co do wartości bezwzględnej mniejsze od 11. Tak więc kody ISSN i ISBN wykrywają błąd przestawienia. Dla kodów EAN już tak nie jest. Warunek podzielności przez p iloczynu $(w_j - w_{j+1})(a_{j+1} - a_j)$ jest dla tego kodu równoważny podzielności przez 10 iloczynu $2(a_{j+1} - a_j)$. Oznacza to, że kody EAN nie wykrywają zamiany sąsiednich cyfr różniących się o 5. Prawdopodobieństwo takiego zdarzenia wynosi 0,1.

Prawdopodobieństwo wyboru określonego miejsca na zamianę wynosi $1/(n-1)$, prawdopodobieństwo tego, że na tym miejscu będą różne cyfry wynosi 0,9, a prawdopodobieństwo warunkowe tego, że różne cyfry na wybranym miejscu różnią się o 5, jest równe $1/9$. Ze wzoru na prawdopodobieństwo całkowite otrzymamy wynik 0,1.

Polecamy Czytelnikom sprawdzenie, jak to jest w kodzie PESEL.

Karty kredytowe

Karty kredytowe (VISA, VISA Electron, Master Card, American Express itp.) są kodowane w sposób bardzo zbliżony do kodu liniowego. Metoda kodowania, zwana wzorem Luhna, została opracowana przez grono matematyków w 1960 r. W metodzie tej używa się współczynników wagowych, równych 1 dla wyrazów nieparzystych i 2 dla parzystych. Karta kredytowa jest poprawnie zakodowana, jeśli suma

$$w_n \circ a_n + \ldots + w_2 \circ a_2 + w_1 \circ a_1$$

jest podzielna przez 10, przy czym w powyższym wzorze znak $\circ$ oznacza iloczyn zmodyfikowany, określony następująco: $x \circ y$ jest równe sumie cyfr zwykłego iloczynu xy. Na przykład zmodyfikowany iloczyn 2 i 5 ($2 \circ 5$) jest równy 1 (bo $2 \cdot 5 = 10$, $1 + 0 = 1$). Podobnie $1 \circ 5 = 5$, $2 \circ 9 = 9$.

Czy pierwszych 16 cyfr rozwinięcia liczby π: 3141592653589793 może być numerem karty kredytowej? Ciąg zmodyfikowanych iloczynów ma postać 6181194613189793. Nie jest to poprawny numer karty kredytowej, gdyż ostatnia cyfra kontrolna powinna mieć wartość 6 (dlaczego?).

Wzór Luhna pozwala bezbłędnie wykryć błąd zamiany. Przypuśćmy, że błąd w pozycji j może nie być wykryty. Wtedy sumy

$$w_n \circ a_n + \ldots + w_j \circ a_j + \ldots + w_1 \circ a_1$$

i

$$w_n \circ a_n + \ldots + w_j \circ x + \ldots + w_1 \circ a_1$$

są podzielne przez 10 oraz $x \neq a_j$. Z tych warunków wynika, że różnica

$$w_j \circ a_j - w_j \circ x$$

jest podzielna przez 10, a tym samym

$$w_j \circ x = w_j \circ a_j.$$

Łatwo sprawdzić, że nie istnieją cyfry $x \neq a_j$ spełniające ten warunek, co stanowi sprzeczność.

Czy system kodowania wykryje przestawienie dwóch kolejnych, różnych cyfr w numerze karty kredytowej? Numer karty, dla którego system nie wykryłby błędu przy zamianie j-tej i $j{+}1$-szej cyfry, musiałby spełniać warunek: liczby

$$w_n \circ a_n + \ldots + w_j \circ a_j + w_{j+1} a_{j+1} + \ldots + w_1 \circ a_1$$

i

$$w_n \circ a_n + \ldots + w_j \circ a_{j+1} + w_{j+1} a_j + \ldots + w_1 \circ a_1$$

byłyby podzielne przez 10. Różnica tych liczb, równa

$$w_j \circ a_j + w_{j+1} a_{j+1} - w_j \circ a_{j+1} - w_{j+1} a_j$$

powinna też być podzielna przez 10. Z dokładnością do znaku różnica ta wynosi

$$2 \circ a_j - a_j - (2 \circ a_{j+1} - a_{j+1}).$$

Liczba ta jest podzielna przez 10 wyłącznie wtedy, gdy a_j oraz a_{j+1} są liczbami 0 lub 9. Prawdopodobieństwo tego błędu, obliczone podobnie jak dla kodu EAN, wynosi 0,018.

Zadania

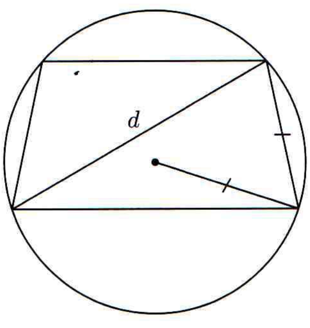
Redaguje Mikołaj KORZYŃSKI

F 627. Obliczyć opór zastępczy między dowolnymi dwoma spośród n kontaktów, połączonych każdy z każdym opornikami o oporze R.
Rozwiązanie na str. 16

F 628. Wewnątrz naładowanej równomiernie sfery o ładunku całkowitym Q i promieniu R znajduje się metalowa kula o promieniu $\frac{R}{3}$. Jaki musi być ładunek elektryczny q metalowej kuli, aby przy braku innych oddziaływań sfera pozostawała nieruchoma?
Rozwiązanie na str. 16

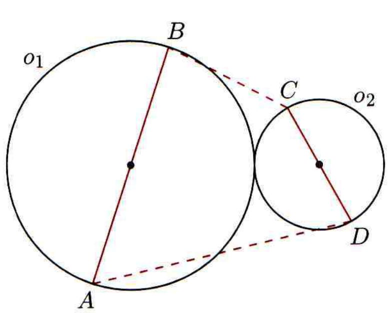
Redaguje Waldemar POMPE

M 1072. W pewnym turnieju każda drużyna rozegrała z każdą dokładnie jeden mecz i nie zanotowano remisu. Udowodnić, że jeżeli pewne dwie drużyny wygrały taką samą liczbę meczów, to istnieją takie trzy drużyny A, B, C, że A wygrała z B, B wygrała z C oraz C wygrała z A.
Rozwiązanie na str. 3



Rys. 1

M 1073. Promień okręgu opisanego na trapezie równoramiennym jest równy długości ramienia tego trapezu. Znając długość d przekątnej trapezu, wyznaczyć pole trapezu (rys. 1).
Rozwiązanie na str. 4



Rys. 2

M 1074. Okręgi o_1 i o_2 są styczne zewnętrznie. Odcinek AB jest średnicą okręgu o_1; odcinek CD jest średnicą okręgu o_2 (rys. 2). Wykazać, że

$$AD + BC \geqslant AB + CD.$$

Rozwiązanie na str. 5