

Odkąd w latach 60. XX wieku zaczęto komputeryzować banki, ważniejsze od nazwisk, imion i adresów klientów stały się numery rachunków bankowych. Ch. Jonscher w książce „Życie okablowane” (Warszawa 2001) przywołuje historię pewnego klienta banku w stanie Maine w USA, który na znak protestu kartkę na Boże Narodzenie do prezesa banku, zamiast imieniem i nazwiskiem, podpisał numerem konta. Rewolucja informatyczna wprowadziła zupełnie nowe znaczenia słowa *cyfrowy*. Ten termin kojarzy się nam zwykle z czymś nowoczesnym, a jednocześnie traktujemy go z respektem, bo zrozumienie jego istoty wydaje się nam niedostępne (a więc i zbędne?). Oby binarna matematyka komputerowa znalazła swe miejsce w szkole. Choćby w formie zadań z konkursu 😊 Radosnych Świąt Narodzenia Pańskiego i dobrego roku 2018 dla wszystkich nauczycieli i sympatyków konkursu KOALA!

Profesor Tim Bell z Uniwersytetu Canterbury w Nowej Zelandii, informatyk i dydaktyk, autor znanych na całym świecie podręczników akademickich z dziedziny kodowania i kompresji danych przygotował wykład z myślą o nauczycielach, którzy 3 listopada 2017 r. spotkali się na konferencji „Informatyka bez komputera” na Wydziale Matematyki i Informatyki UAM. Wykład skończył wyjątkowymi słowami: „Chciałbym podziękować Państwu za to, że jesteście gotowi podjąć wysiłek dokształcania się, żeby później pomóc uczniom zrozumieć cyfrowy świat, w którym żyją.”

Chcielibyśmy mieć w tym procesie swój mały wkład. Dlatego w drugim numerze biuletynu zamieszczamy zadania poprzednich edycji konkursu KOALA związane z tematem kodowania (cyfrowej reprezentacji informacji). Poza tym drukujemy artykuły popularnonaukowe, scenariusz lekcji oraz krzyżówkę.

Zadania pierwszego etapu V jubileuszowej edycji konkursu KOALA udostępnimy w czwartek 4 stycznia na stronie <http://koala.vlo.poznan.pl/>. Zapraszamy do udziału!

Paweł Perekieta

nauczyciel w V Liceum Ogólnokształcącym im. Klauzyny Potockiej w Poznaniu



Wydanie biuletynu KOALA było możliwe dzięki środkom finansowym pozyskanym przez Poznańską Fundację Matematyczną.

projekt: Potęga Matematyki 2017

finansowanie: Miasto Poznań

Kodowanie niejedno ma imię

Paweł Perekieta, V Liceum Ogólnokształcące w Poznaniu

Termin *kodowanie* stał się w ostatnich latach słowem-kluczem na określenie działań popularyzujących naukę programowania. Naukowe znaczenie terminu kodowanie w informatyce i telekomunikacji jest szersze: kodowanie to przekształcanie informacji w ciąg znaków lub symboli. Celem kodowania jest znalezienie sposobu reprezentowania (wymiarowania) informacji dla jej zapisu na jakimś nośniku (pamięci) lub przesłania do odbiorcy.

Widzimy wokół nas, że od dłuższego czasu to, co było *analogowe*, jest wypierane przez *cyfrowe*: cyfrowa fotografia, telefonia, telewizja itd. Co to właściwie oznacza?

Na przestrzeni kilku tysięcy lat ludzie wynaleźli mnóstwo sposobów rejestrowania, przechowywania, przetwarzania i przesyłania informacji. Różne niebanalne pomysły powstawały niezależnie w różnych częściach świata, a potem były ulepszone przez wieki. Wszystkie te metody w istocie można podzielić na dwie grupy: analogowe i cyfrowe (dyskretne). Współczesna informatyka to dziś prawie wyłącznie domena technik cyfrowych. Jesteśmy otoczeni sygnałami, które są ze swej natury analogowe i sami takie wytwarzamy (np. nasze serce i mózg). Technika cyfrowa dostarcza sprawnych narzędzi do ich rejestrowania i przetwarzania.

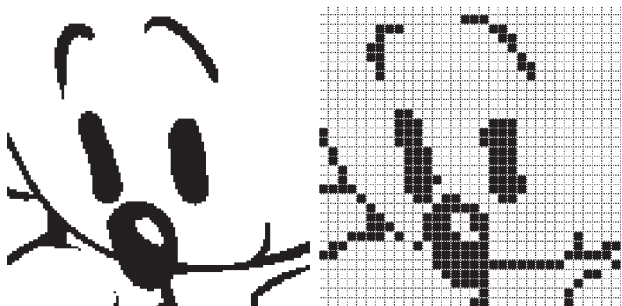
Dla danego cyfrowego sposobu kodowania (reprezentowania danych) musi być wybrany i ustalony pewien alfabet, zawierający co najmniej dwa symbole. Wszystkie dane zakodowane cyfrowo mają postać skończonych ciągów złożonych wyłącznie z symboli przyjętego alfabetu. Człowiek określa reguły kodowania, czyli przekształcania danych na postać cyfrową, reguły (algorytmy) przetwarzania danych (te algorytmy działają w istocie na liczbach) i interpretacji wyników (dekodowania).

	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I/J	K
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

Tablica kodowa Polibiusza, np. ABC = 11 12 13

Termin „cyfrowy” nie oznacza wcale, że mamy do czynienia koniecznie z cyframi: 0, 1, 2, 3... itd. Istotą jest to, dane składają się z ciągu podstawowych symboli alfabetu, które tworzą zbiór nieciągły, skokowy, dyskretny. Technicznie najdogodniejszy jest alfabet binarny. O tworzących go symbolach mówi się: 0 i 1.

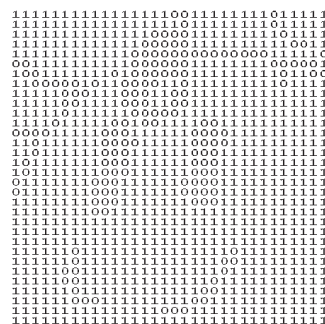
Jeśli chcemy np. wysłać faksem cyfrową kopię starej (analogowej) fotografii w odcieniach szarości lub kopię czarno-białego rysunku, to musimy dokonać konwersji analogowo-cyfrowej, zwanej dyskretyzacją. Obraz jest skanowany, linia po linii. Efekt konwersji, czyli postać cyfrowa obrazu, składa się z pewnej skończonej liczby elementów zwanych pikselami lub pelami (ang. picture element).



Na rysunkach przedstawiono kwadratowy kadr postaci koali: po lewej obraz w linii składa się ze 128 pikseli, po prawej obraz ma cztery razy mniejszą rozdzielczość, bo składa z tylko 32 pikseli (co można sprawdzić, bo dodano „siatkę”). Jeśli uda nam się zmniejszyć niedokładność konwersji tak, że będzie mniejsza od granicy rozdzielczości (np. na ekranie monitora to ok. 100 pikseli na cal dla gołego, nieuzbrojonego oka; na wydruku – ok. 300 punktów na cal), to odbiorca nie będzie w stanie odróżnić cyfrowej kopii od oryginału.

W cyfrowej reprezentacji obraz staje się ciągiem liczb. W przypadku obrazu czarno-białego uzyskujemy tzw. mapę bitową, co oznacza, że każdy element obrazu możemy kodować cyfrą dwójkową, czyli bitem (ang. binary digit): 0 lub 1. Po zdekodowaniu oznaczać to będzie: zaciemnienie lub jego brak (w przypadku wydruku) albo wygaszenie lub zapalenie piksela (na ekranie). W przypadku obrazu w odcieniach szarości będzie to ciąg liczb z zakresu 0 – 255, określających poziom szarości, przy czym na każdy piksel przypada osiem bitów (jeden bajt).

Fizyczna realizacja zapisu binarnego na różnych etapach przetwarzania danych może być różna: dwa napięcia prądu w czasie konwersji analogowo-cyfrowej, modulacja częstotliwości (niski i wysoki ton) w miedzianym kablu telefonicznym.



Kod binarny mapy bitowej 32 x 32 kadru koali (format BMP)

W praktyce informatycy (tym bardziej graficy komputerowi) nie posługują się na co dzień kodem binarnym, gdyż zapis jest nieekonomiczny, to znaczy zajmuje dużo miejsca i jest narażony na błędy. Każdą czwórkę bitów (półbajt) zapisuje się zwykle z użyciem kodu szesnastkowego, np. liczbę binarną 0111 zapisuje się jako 7, a 1110 zapisuje się jako E.

Obraz cyfrowy może być poddany kompresji (upakowaniu bitów). Na przykład podczas faksowania korzysta się z kompresji, co zmniejsza objętość danych ok. 7 razy. Podczas transmisji na duże odległości, np. z sondy kosmicznej, konieczne jest dodanie dodatkowych bitów korekcyjnych, celem umożliwienia wykrycia i naprawienia przekłamań (zamiany 0 na 1 i odwrotnie) pojawiających się podczas transmisji.

To już tematy na inny artykuł. Lub zadanie konkursowe...



Z pogranicza matematyki i informatyki. Claude E. Shannon

„Claude Shannon był tym, który jako pierwszy uchylił drzwi do cyfrowego świata.”
(Jerzy Mieścicki)

Claude Elwood Shannon był cudownym dzieckiem. W 1936 r. w wieku 20 lat uzyskał na uniwersytecie w Michigan od razu dwa dyplomy ukończenia studiów licencjackich, na kierunkach inżynierii elektrycznej oraz matematyki. Już w 1937 r. ukończył studia w Massachusetts Institute of Technology (MIT). W pracy dyplomowej wykazał, że przy projektowaniu elektrycznych układów przekaźnikowych i przełączających można wykorzystać reguły algebry logicznej opisanej sto lat wcześniej przez George'a Boole'a, nieznaney inżynierom. Ta praca została później przez wielu uznana za najważniejszą pracę magisterską XX w.

Shannon doktorat obronił w MIT w roku 1940. W swej pracy zajmował się algebraicznymi modelami w genetyce. W 1943 r. przez kilka miesięcy współpracował z Alanem Turingiem. Opracowali wspólnie system cyfrowego kodowania mowy (celem było szyfrowanie rozmów telefonicznych na najwyższym szczeblu polityki; projekt nie został jednak wdrożony). Claude Shannon pracował w laboratorium badawczym Bell firmy telekomunikacyjnej AT&T od 1941 r. do 1956 r., potem objął stanowisko profesora w MIT. Tam pracował aż do emerytury w 1978 r. Był nauczycielem pierwszej generacji badaczy sztucznej inteligencji.

W 1947 r. Shannon (i w 1948 r. wspólnie z Warrenem Weaverem) opublikował fundamentalne artykuły o matematycznej teorii informacji i komunikacji. Trudno przecenić znaczenie tego dzieła. Stanowiły zbiór praw, które bezpośrednio odnosiły się do efektywnej ekonomicznie i sprawnej transmisji (np. ile energii musimy włożyć w rozsyłanie głosu przez radio, aby mieć pewność, że będzie on zrozumiany mimo zakłóceń atmosferycznych lub elektrostatycznych pochodzących z innych źródeł). Shannon wykazał, że dzięki opracowaniu właściwego kodu można transmitować dowolne wiadomości z dowolnie dużą niezawodnością. Ogólna idea była wystarczająco prosta, by Shannon mógł ją przedstawić za pomocą gry „pięciu pytań” (uproszczona odmiana „20 pytań”, znanej w USA gry rozrywkowej): Gracz numer 1 myśli o pewnej literze alfabetu. Gracz nr 2 próbuje ją odgadnąć, zadając pytania w rodzaju „czy jest w alfabecie przed L?”. Shannon zwrócił uwagę na to, że wystarczy nie więcej niż pięć pytań, aby zlokalizować dowolną literę alfabetu angielskiego. Jeśli ciąg decyzji binarnych „tak” lub „nie” prowadzących do określenia właściwej litery zostałby przekształcony w sekwencję 0 i 1 (technicznie np. jako impulsów włączenia i wyłączenia), mielibyśmy do dyspozycji kod służący do transmitowania wiadomości dalekopisowych (np. telegramów). Podstawowy element w tym kodzie (decyzja binarna) stał się podłożem wprowadzonej przez Claude'a Shannona podstawowej miary informacji: *bit* (ang. binary information unit).

W 1949 i 1951 Shannon opublikował pionierskie artykuły o matematycznej teorii systemów kryptograficznych oraz o teorii przetwarzania języka naturalnego. W 1950 r. opisał kombinatoryczne zasady działania programu komputerowego, który byłby zdolny do gry w szachy. Claude Shannon był jednym ze współautorów twierdzenia matematycznego, rozwiązującego problem efektywnego próbkowania (dyskretyzacji) sygnału dźwiękowego. On podał również sposób na konwersję cyfrowo-analogową, czyli rekonstrukcję sygnału ciągłego z ciągu cyfrowych próbek.

Jako ciekawostkę dodajmy, że Claude Shannon potrafił jeździć na jednokołowym rowerze własnej konstrukcji (i to o osi nie umieszczonej centralnie), jednocześnie zonglując kilkoma palcami.

Był jednym z najwybitniejszych uczonych XX w. Prof. Robert Fano, laureat nagrody im. Claude'a E. Shannona z 1976 r., mówił o wykładach Shannona: „Nie lubi nauczać. Nie lubi wykładać. Wszystkie jego wykłady są cudowne, każdy jest perełką. Brzmiały spontanicznie, jednak w rzeczywistości są bardzo starannie przygotowane.”

Zmarł w 2001 r., w wieku 84 lat. Niestety choroba Alzheimera w ostatnich latach życia nie pozwoliła mu na korzystanie z urzędzeń współczesnej teleinformatyki, pod które położył teoretyczne podwaliny.



Więcej informacji o genialnym twórcy teorii informacji, i wywodzącej się z niej teorii kodowania, można przeczytać m.in. w książkach: H. Rheingold „Narzędzia ułatwiające myślenie”, Warszawa 2003, J. Mieścicki „Wstęp do informatyki”, Legionowo 2013 oraz J. Gleick „Informacja. Bit. Wszechświat. Rewolucja”, Kraków 2012.

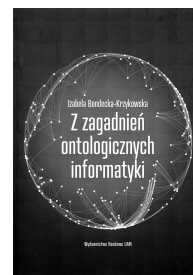
O teorii informacji

Istotą teorii informacji, sformułowanej przez Shannona, jest pojęcie niepewności. Im większa jest niepewność treści (znaczenia) sygnału, na przykład spowodowana zakłóceniami na drodze między nadawcą i odbiorcą, tym więcej bitów informacji trzeba do jej zapisu.

Shannon podał matematyczny dowód na to, że sygnał można tak zakodować (eliminując zakłócenia, czyli szum), by zminimalizować praktycznie do zera możliwość błędnego odczytania treści sygnału. Ograniczenie nałożone przez naturę wiąże się jedynie z koniecznością istnienia kanału komunikacyjnego. Miara zależności między złożonością kodu a stopniem pewności jest to, co Claude Shannon nazywa entropią.

Prawa opisane przez Shannona oznaczają bardzo wiele dla inżynierów, umożliwiły powstanie telewizji kolorowej i przekaz obrazu z Księżycą. Trzeba jednak podkreślić, że prawa Shannona są uniwersalne i wykraczają poza inżynierię elektryczną.

To, co Shannon nazywał kodowaniem źródła, dziś nazywa się kompresowaniem danych źródłowych (kto „zipował” plik, widział kodowanie danych na żywo). To, co nazywał kodowaniem kanału, dziś w praktyce realizuje się przez dodanie dodatkowych bitów danych, aby proces przesyłania (lub zapisu na nośniku pamięci) był bardziej odporny na zakłócenia występujące w kanale transmisyjnym.



Zainteresowanym filozofią informatyki (np. pojęciem informacji) polecamy książkę dr Izabeli Bondeckiej-Krzykowskiej „Z zagadnień ontologicznych informatyki”, wydanej w 2016 r. przez Wydawnictwo UAM.

Na s. 4 publikujemy scenariusz zajęć pt. Gra w 20 pytań, stworzony w ramach projektu Computer Science Unplugged, który jest próbą objaśnienia m.in. entropii.

Na stronie pl.khanacademy.org w dziale pt. Informatyka znaleźć można materiały wideo, o teorii informacji i kodowania.

Gra w 20 pytań (scenariusz lekcji)

Wprowadzenie

Ile informacji znajduje się w grubej książce? Czy więcej informacji znajduje się w encyklopedii o 1000 stronach, czy w powieści „Władcy Pierścieni” Tolkiena? Jeśli będziemy potrafili to oszacować, będziemy potrafili też oszacować, ile miejsca potrzeba do zapisania jakiegokolwiek informacji.

Czy potrafisz odczytać zapisane poniżej zdanie?

W tm zdn brkj smgłsk.

Bardzo możliwe, że tak. W samogłoskach nie ma zawartych zbyt wielu „informacji”. W ramach tych zajęć przedstawiona będzie metoda pomiaru „zawartości informacyjnej”.

Dyskusja wstępna

W jaki sposób moglibyśmy zmierzyć ilość informacji, która jest zawarta w jakiejś książce? Czy ważna jest np. liczba stron albo liczba wyrazów? Czy jedna książka może zawierać więcej informacji niż inna? Co w przypadku książki bardzo nudnej lub szczególnie interesującej? Czy książka, w której na 400 stronach powtarza się fraza „bla, bla, bla” ma mniej czy więcej informacji niż np. książka telefoniczna?

Usłyszenie czegoś, co jest już Ci znane – na przykład zdania „Przyszedłem dziś do szkoły pieszo.” od kolegi, który tak robi codziennie – nie wiąże się z otrzymaniem właściwie żadnej informacji, ponieważ nie jest to czymś mało prawdopodobnym (zaskakującym). Gdyby jednak kolega powiedział „Przyleciałem dziś do szkoły helikopterem.”, byłoby to zaskakujące, i dlatego wiązałoby się z przekazaniem większej ilości informacji.

W jaki sposób zmierzyć to, jak prawdopodobna jest informacja? Jednym ze sposobów jest sprawdzenie, czy łatwo odgadnąć daną informację. Twój kolega mówi: „Spróbuj zgadnąć, w jaki sposób dotarłem dziś do szkoły!”. Jeśli dotarł pieszo lub autobusem, prawdopodobnie od razu odgadniesz. Nie przyjdzie Ci na myśl, by wymienić helikopter czy statek kosmiczny.

Ilość informacji, jaka jest zawarta w danej wiadomości, mierzy się tym, jak łatwo lub trudno ją odgadnąć.

Gra w 20 pytań

Posłużymy się adaptacją dość znanej gry w 20 pytań: Wybrany uczestnik zajęć ustala w myślach liczbę (od 1 do 100 albo od 1 do 1000) a pozostali zadają pytania (aż do skutku), na które może on odpowiadać tylko „tak” lub „nie”.

Następnie zliczamy liczbę pytań, które zostały zadane. To będzie miara ilości informacji.

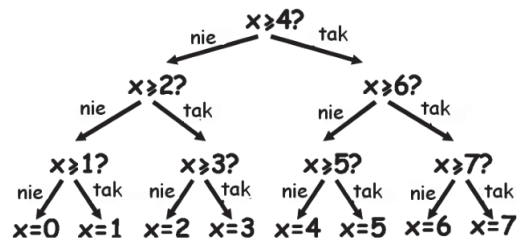
Dyskusja po grze

Pytamy uczestników o strategię, jakich używali. Które z nich były najlepsze? Do znalezienia liczby między 1 i 100 powinno wystarczyć siedem prób odgadnięcia, jeśli za każdym razem dzielimy przedział na pół. Dla przykładu:

- Czy to liczba mniejsza od 50? Tak.
- Czy to liczba mniejsza od 25? Nie.
- Czy to liczba mniejsza od 37? Nie.
- Czy to liczba mniejsza od 43? Tak.
- Czy to liczba mniejsza od 40? Nie.
- Czy to liczba mniejsza od 41? Nie.
- To musi być liczba 42! Tak!

Ciekawe jest to, że dziesięciokrotne rozszerzenie zakresu liczb (do 1000) nie wymaga dziesięciokrotnego zwiększenia nakładu pracy – liczba potrzebnych pytań wzrasta najwyżej o 3. Za każdym razem, gdy zakres liczb podwajamy, liczba pytań potrzebnych do znalezienia odpowiedzi zwiększa się tylko o 1.

Oto schemat zwany drzewem decyzyjnym, który może służyć odgadnięciu liczby z zakresu od 0 do 7:



Jakich pytań użyć do „odgadnięcia” liczby 5?

Liczby 0, 1, 2, ..., 7 z ostatniego rzędu zapisz binarnie.

Przyjrzyj się dokładnie drzewu. Niech „nie” = 0, a „tak” = 1. (Mamy zakodowany binarnie optymalny ciąg „zgadywać”).

O co w tym wszystkim chodzi?

Czasami można zadać takie pytanie, które eliminuje konieczność stawiania wielu innych. W takim wypadku zawartość informacyjna wiadomości (komunikatu) jest niska. Dla przykładu: informacja dotycząca wyniku rzutu monetą jest jednobitowa: orzeł lub reszka. Ale jeśli moneta okaże się być niesymetryczna i orzeł wypada średnio 9 razy na 10 rzutów, to wówczas miarą ilości informacji dotyczącej rzutu taką monetą jest mniej niż 1 bit. Musimy postawić odpowiednie pytanie dotyczące dwóch kolejnych rzutów. Należy użyć pytania: „Czy wynikami kolejnych dwóch rzutów monetą jest orzeł?”. W przypadku rzutu taką niesymetryczną monetą odpowiedź „tak” pojawi się w ok. 80% przypadków. W ok. 20% przypadków, kiedy pojawi się odpowiedź „nie”, potrzebne będzie zadanie dwóch dalszych pytań. Jednak średnio (statystycznie) liczba pytań będzie mniejsza niż jedno pytanie na jeden rzut monetą!

Shannon nazwał zawartość informacyjną wiadomości „entropią”. Entropia zależy nie tylko od liczby możliwych wyników – np. dwóch w przypadku rzutu monetą – ale także od prawdopodobieństwa ich wystąpienia. Mało prawdopodobne zdarzenia, czy zaskakujące informacje, wymagają postawienia o wiele większej liczby pytań prowadzących do odgadnięcia odpowiedzi, ponieważ zawierają więcej informacji, których wcześniej nie znaliśmy.

Entropia jest ważnym pojęciem dla informatyki. Niemożliwe jest skompresowanie (upakowanie) wiadomości tak, aby zajmowała mniej miejsca niż wartość jej entropii. Najlepsze systemy kompresji są, używając obrazowego porównania, odpowiednikami wyżej opisanej zgadywanki. Wówczas „zgadywanie” jest wykonywane przez program komputerowy. (Lista pytań może być później odtworzona.) Plik skompresowany to ciąg bitów, które są odpowiedziami w „zgadywance”. Tak długo jak odpowiedzi (bity) są przechowywane, możemy odtworzyć informację! Najlepsze systemy kompresji potrafią zmniejszyć rozmiar plików tekstowych do ok. ¼ ich rozmiaru oryginalnego.

Podobna metoda „odgadywania” może być zastosowana podczas projektowania interfejsu dla komunikacji człowiek – komputer. Aby przewidzieć kolejne czynności użytkowników, zwłaszcza tych szczególnych – osób niepełnosprawnych, dla których np. pisanie na klawiaturze stanowi dużą trudność. Komputer sugeruje wówczas np. najbardziej prawdopodobne (najczęściej używane) słowa, a użytkownicy tylko wskazują to właściwe. Dobrze zaprojektowany system wymaga średnio tylko dwóch pytań i odpowiedzi typu „tak” lub „nie”, w celu „odgadnięcia” kolejnego słowa.

Powyższy scenariusz znajduje się w zbiorze 20 scenariuszy pochodzących z projektu *Computer Science Unplugged* dostępnych na stronie <http://bezkomputera.wmi.amu.edu.pl/>

Binarna gra (I etap 2016/2017)

Bolek i Lolek grają w taką grę: Bolek wymyśla pięciobitowy kod binarny, tj. ciąg składający się z pięciu cyfr, z których każda jest albo 0, albo 1. Zadaniem Lolka jest zidentyfikować kod. Lolek podaje przykład kodu (np. 00110), a Bolek w odpowiedzi podaje informację o liczbie poprawnie wskazanych cyfr (np. dwa dla kodu 10011). Oto sześć pytań, jakie Lolek zadał i odpowiedzi, jakich udzielił Bolek:

- Czy pomyślałeś o 00000?
- Nie, ale dwa bity się zgadzają.
- Czy pomyślałeś o 01000?
- Nie, ale jeden bit się zgadza.
- Czy pomyślałeś o 00100?
- Nie, ale trzy bity się zgadzają.
- Czy pomyślałeś o 01110?
- Nie, ale trzy bity się zgadzają.
- Czy pomyślałeś o 11100?
- Nie, ale jeden bit się zgadza.
- Czy pomyślałeś o 00001?
- Nie, ale trzy bity się zgadzają.

Czy można zidentyfikować kod Bolka?



Rozwiązanie. Tak. Wykażemy, że kodem Bolka jest 00111. Z pierwszej odpowiedzi wnioskujemy, że w kodzie są dwa zera i trzy jedynki. Po drugiej odpowiedzi wiemy, że drugim (od lewej) bitem jest 0, bo po zaproponowaniu 1 na drugim miejscu liczba właściwych bitów zmalała o jeden w stosunku do poprzedniego pytania. Analogiczny argument prowadzi do wniosku po trzeciej odpowiedzi, że trzecim bitem jest jedynka. Na tym etapie wiemy, że ciąg ma postać ?01??. a na niezidentyfikowanych miejscach są w dwie jedynki i jedno zero. Teraz najwygodniej spojrzeć na pytanie piąte. Z odpowiedzi na to pytanie (i informacji, jakie już ma Lolek) wynika, że jedynym właściwym bitem jest jedynka na trzecim miejscu. A zatem kodem Bolka jest 00111.

Zepsuta pamięć (I etap 2015/2016)

Blok pamięci komputera, składający się z 4×4 komórek jest kontrolowany za pomocą ośmiu przełączników. Na przykład, jeśli włączone są przełączniki A i 1 , czytana jest komórka $A1$. Aktualna zawartość tego fragmentu pamięci jest taka:

	A	B	C	D
1	0	0	1	1
2	0	1	0	1
3	0	1	1	1
4	1	1	1	0

Jeśli zepsuje się przełącznik kolumny (mówimy wtedy, że kolumna ta jest zepsuta), odczyt każdej komórki w jakimś wierszu da wynik będący wartością komórki z zepsutej kolumny. Na przykład, gdy popsuty jest przełącznik A , odczyt każdej z komórek $A1$, $B1$, $C1$, $D1$ da wartość z komórki $A1$; zamiast wartości z komórek $A2$, $B2$, $C2$, $D2$, zostanie odczytana wartość z $A2$, itd. Podobny wpływ na odczyt ma zepsucie się przełącznika wiersza. Na przykład, gdy popsuty jest przełącznik 3, zamiast wartości z komórek $A1$, $A2$, $A3$, $A4$ zostanie odczytana wartość z komórki $A3$. Wiadomo, że jeden przełącznik jest zepsuty. Ponadto w efekcie odczytu komórki $A1$ otrzymano 0, w wyniku odczytu $B2$ otrzymano 1, a w przypadku $D1$ otrzymano 0. Który przełącznik jest zepsuty?

Rozwiązanie. Odczyt $A1$ daje 0, więc zepsuty nie może być żaden z przełączników: C , D , ani 4. Odczyt $B2$ daje 1, więc zepsuty nie może być żaden z przełączników: A , C , ani 1. Odczyt $D1$ daje 0, więc zepsuty nie może być żaden z przełączników: C , D , 1, 2, 3. Wnioskujemy, że zepsuty jest przełącznik B .

Blok pamięci (I etap 2015/2016)

Tabela przedstawia blok pamięci komputera. W każdym wierszu widzimy osiem bitów jednego bajtu pamięci. Każdy bajt powinien mieć parzystą liczbę jedynek, również na każdej pozycji (w każdej kolumnie bitów) liczba jedynek powinna być parzysta. Jednak jeden bit w bloku jest niewłaściwy. Który?

bajt 1	1	1	0	0	1	1	0	0
bajt 2	1	1	1	0	0	1	1	0
bajt 3	1	0	0	0	0	1	1	1
bajt 4	1	0	1	1	1	1	0	1

Rozwiązanie. Zmiana jednego bitu bloku pamięci zmienia parzystość liczby jedynek w wierszu i kolumnie, w którym zmieniany bit się znajduje. W podanej tabeli jedynym wierszem z nieparzystą liczbą jedynek jest wiersz drugi a jedyną kolumną z nieparzystą liczbą jedynek jest kolumna czwarta od lewej. Zatem, skoro w prawidłowym bloku pamięci wszystkie wiersze i kolumny bitów miały parzystą liczbę jedynek, niewłaściwy bit znajduje się w drugim wierszu i czwartej kolumnie. W tym miejscu zamiast 0 powinno być 1.

Komunikacja kosmiczna (finał 2016/ 2017)

W komunikacji z niektórymi sondami kosmicznymi stosowano następujące kodowanie korekcyjne: każdy bit źródłowej wiadomości przesyłany był pięciokrotnie, np. dla wiadomości 1011 wysyłane były cztery paczki: 11111 00000 11111 11111. Podczas transmisji możliwe były błędy: zakładamy, że prawdopodobieństwo zmiany bitu wynosiło 0,05 dla każdego bitu. Podczas odczytu za poprawny bit wiadomości uznawano ten, który w paczce bitów pojawiał się częściej. Na przykład paczka 10110 traktowana była jako bit 1, a paczka 10000 jako bit 0. Oblicz prawdopodobieństwo poprawnego odczytania wiadomości, która w wersji źródłowej składała się z ośmiu bitów. Jakie byłoby prawdopodobieństwo poprawnego odczytu, gdyby nie stosowano kodowania korekcyjnego?

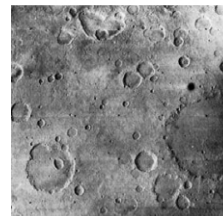
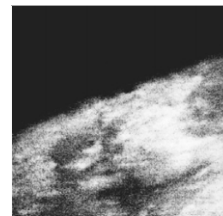
Rozwiązanie. Prawdopodobieństwo, że pojedynczy bit będzie odczytany poprawnie wynosi $1 - 0,05 = 0,95$. Gdyby nie stosowano żadnego kodowania korekcyjnego, prawdopodobieństwo odczytania wiadomości wynosiłoby $(0,95)^8 \approx 0,67$.

Po zastosowaniu kodowania, bit wiadomości będzie odczytany prawidłowo wtedy i tylko wtedy, gdy w odpowiadającej mu paczce będą co najwyżej dwa błędy. Prawdopodobieństwo zdarzenia liczymy, dzieląc na przypadki względem liczby przekłamanych bitów i uwzględniając ich położenie w paczce:

$$(0,95)^5 + \binom{5}{1}(0,95)^4 \cdot 0,05 + \binom{5}{2}(0,95)^3 \cdot (0,05)^2 = 0,998841875.$$

Prawdopodobieństwo poprawnego odczytania całej wiadomości, czyli wszystkich ośmiu bitów wynosi zatem $(0,998841875)^8 > 0,9907$.

Zdjęcia zostały wykonane przez sondy wysłane w kierunku Marsa. Pierwsze przez Mariner IV w 1964 r. Nie używano wówczas kodowania korekcyjnego. Drugie przez Mariner VI w 1969 r. Efekt zastosowania kodów Reeda-Mullera (bardziej wydajnych od kodu „2 z 5”) doskonale widać.



Zebry (II etap 2015/2016)

Zebry występują w wielu podgatunkach. Są na przykład takie, które mają paskowane tylko nogi. U jednego z nich zaobserwowano, że paski tworzą regularną strukturę (rodzaj naturalnego kodu paskowego podgatunku), odstrasżającą owady, którą można scharakteryzować tak:



- pierwszy pasek nad kopytem jest zawsze czarny,
- ostatni pasek też jest zawsze czarny,
- pasek (również pierwszy i ostatni) może mieć pojedynczą lub podwójną szerokość, czyli 4 cm lub 8 cm.

Ile jest niepowtarzalnych układów pasków, jeśli przyjąć, że paski pokrywają 52 cm nogi zebry?

Rozwiązanie 1. Zająć mogą cztery przypadki:

- Układ zaczyna się i kończy się paskami szerokości 4 cm.
- Układ zaczyna się paskiem 4 cm, a kończy się paskiem 8 cm.
- Układ zaczyna się paskiem 8 cm, a kończy się paskiem 4 cm.
- Układ zaczyna się i kończy się paskami szerokości 8 cm.

Pomiędzy skrajnymi paskami czarnymi znajdują się układy pasków, pokrywające odpowiednio 44, 40, 40 i 36 cm, w których pierwszy i ostatni pasek są białe. Po odrzuceniu pasków skrajnych mamy do rozwiązania zadanie analogiczne do zadania pierwotnego, ale o mniejszym rozmiarze danych (kolor pasków skrajnych jest bez znaczenia dla istoty zadania). Rozumując w ten sam sposób dla układów pasków długości $4n$ cm, wnioskujemy, że dla $n \geq 5$, liczba L_n układów pokrywających $4n$ cm nogi zebry jest następującą sumą:

$$L_n = L_{n-2} + 2L_{n-3} + L_{n-4}.$$

Dla mniejszych n przeliczamy wszystkie układy pasków bezpośrednio i otrzymujemy: $L_1 = 1$, $L_2 = 1$, $L_3 = 1$, $L_4 = 3$. Kolejne wartości aż do szukanej w zadaniu liczby L_{13} wyznaczamy z powyższego wzoru, a wyniki ilustruje poniższa tabelka.

n	1	2	3	4	5	6	7	8	9	10	11	12	13
$L(n)$	1	1	1	3	4	6	11	17	27	45	72	116	189

Odpowiedzią jest zatem 189.

Rozwiązanie 2. Pasków może być co najwyżej 13, gdyż $52/4 = 13$. Pasków jest nieparzysta liczba, bo układ zaczyna się i kończy paskiem czarnym. Może być ich zatem 11 (w tym dwa podwójne), 9 (w tym cztery podwójne), 7 (w tym sześć podwójnych), a mniej już nie, bo co najwyżej 5 pasków zajmuje co najwyżej 40 cm. Jeśli mamy m pasków i k z nich ma podwójną grubość, to układów mamy $\binom{m}{k}$. Podsumowując, wszystkich układów jest

$$\binom{13}{0} + \binom{11}{2} + \binom{9}{4} + \binom{7}{6} = \frac{13!}{0!13!} + \frac{11!}{2!9!} + \frac{9!}{4!5!} + \frac{7!}{6!1!} \\ = 1 + 55 + 126 + 7 = 189.$$



Rozwiązania zadań zredagowała Małgorzata Bednarska-Bzdęga.

Wiadomość z kosmosu (II etap 2015/2016)

Wiadomości, jakie chcemy przesyłać ze statku kosmicznego na Ziemię są liczbami, w których nie występuje zero (mogą wystąpić cyfry 1, 2, ..., 9). Niektóre cyfry mogłyby zostać zgubione podczas transmisji, więc dla bezpieczeństwa wiadomość jest wysyłana tak: po każdej cyfrze d wstawiamy $d - 1$ zer. Na przykład, gdybyśmy chcieli przekazać wiadomość 42312, wysłalibyśmy 400020300120. Załóżmy, że wysłaliśmy ze statku wiadomość, a Ziemia otrzymała 0040300102400. Jaka jest najmniejsza liczba cyfr (w tym zer), jakie mogły zostać w czasie transmisji zgubione?

Rozwiązanie. Uzasadnimy, że zgubiono co najmniej sześć cyfr. Jeśli blok zer po jakiejś niezerowej cyfrze w ciągu otrzymanym na Ziemi jest za krótki, to należy dodać odpowiednią liczbę 0. Zgubiono zatem przynajmniej dwa 0 występujące w oryginalnej wiadomości po pierwszej cyfrze 4, przynajmniej jedno 0 występujące po drugiej czwórce i przynajmniej jedno występujące po cyfrze 2. Widzimy też, że w wiadomości na Ziemi pojawia się blok 10. Po cyfrze 1 statek nie nadaje 0, więc w wiadomości ze statku między 1 a 0 występowała na pewno cyfra większa niż 1. Zauważmy ponadto, że statek na początku nadaje cyfrę niezerową, a Ziemia otrzymała na początku 0, zatem przynajmniej jedna cyfra z początku została zgubiona. Podsumowując, brakuje przynajmniej sześciu cyfr.

Należy jeszcze wykazać, że istnieje komunikat, z którego po usunięciu sześciu cyfr powstaje ciąg otrzymany przez Ziemię. Statek mógł wysłać na przykład 3004000300120204000. Po wykreśleniu pogrubionych cyfr otrzymujemy wiadomość z zadania.

Sklejanie DNA (finał gimnazjalny 2016/2017)

Treść zadania znajduje się na następnej stronie.

Rozwiązanie. Uzasadnimy, że najkrótszy łańcuch liczy 11 cząstek. Zauważmy, że odcinki *ACTA* i *TACG* i *ACGA* są podłańcuchami innych, tj. zawierają się w odcinku *GACTA* lub *ATACG*. Wystarczy więc znaleźć najkrótszy łańcuch dla pozostałych pięciu odcinków DNA. Oto przykład łańcucha o 11 cząstkach, zawierającego wszystkie odcinki z zadania: *GACTATACGAC*. Należy jeszcze wykazać, że krótszego łańcuch nie ma. W tym celu można przeanalizować przypadki, np. zacząć od *ATACGA* i rozważyć dwa przypadki zazębiania się *ATACGA* i *GACTA* (gdyby się nie zazębiały, to od razu byłaby długość 11). Ta metoda szybko prowadzi do wniosku, że mniej niż 11 cząstek być nie może. Zaprezentujemy niżej metodę nieco inną, przydatną w bardziej skomplikowanych przypadkach.

Mamy pięć odcinków i wiemy, że żaden nie zawiera się w innym. Załóżmy, że mamy łańcuch L zawierający wszystkie te odcinki. Zaznaczmy początki (czyli lewe końce) wszystkich pięciu odcinków na łańcuchu L . Oczywiście każdy odcinek ma inny początek, bo żaden odcinek nie zawiera się w innym. Numerujemy odcinki zgodnie z porządkiem ich początków na L . Zauważmy, że dzięki tej numeracji żaden i -ty odcinek nie zawiera się w sumie odcinków od 1 do $i - 1$, bo w przeciwnym wypadku odcinek i -ty zawierałby się w całości w odcinku $i - 1$. Dlatego mamy 4 „zakładki” (odcinki w L łączą się „na zakładkę”). Łatwo sprawdzić, oglądając pięć odcinków z zadania, że każda zakładka ma długość co najwyżej 3. Długość łańcucha L jest zatem nie mniejsza niż: suma długości wszystkich odcinków pomniejszona o $4 \cdot 3$, czyli wynosi co najmniej $23 - 12 = 11$.

Sklejanie DNA (finał gimnazjalny 2016/2017)

Ludzka nić DNA liczy ok. 3 mld elementów, czyli cząstek kwasu dezoksyrybonukleinowego. Cząstka może mieć jedną z czterech wartości: G, T, C lub A. Odczytanie tak długiego łańcucha jest bardzo kłopotliwe. Ponieważ łatwo jest odczytać krótkie łańcuchy, więc jedną z metod odcyfrowania DNA jest podzielenie łańcucha, a ściślej wielu kopii tego samego łańcucha w roztworze, na mniejsze odcinki. Podziały dla poszczególnych kopii łańcucha mogą być różne i np. jedna kopia łańcucha ACTACAG może zostać podzielona na odcinki: A, CT, AC, AG, a inna kopia na odcinki: AC, TA, CAG. Wszystkie tak powstałe odcinki współlistniają w tym samym roztworze.

Później wczytuje się informacje o odcinkach do pamięci komputera, który zajmuje się odtworzeniem prawdopodobnego wyglądu wyjściowego łańcucha. Komputer szuka takiego łańcucha, który zawiera wszystkie odcinki z roztworu, a jednocześnie jego długość jest jak najmniejsza.

Na przykład, najkrótszym łańcuchem zawierającym wszystkie odcinki: AC, GTA, CC, TAC jest przykładowo GTACC. Kolejność odczytu jest ważna i łańcuch CCATG nie byłby dobrym rozwiązaniem.

Jaka jest najmniejsza możliwa długość łańcucha, zawierającego wszystkie odcinki: ACTA, CTAT, CGAC, ATACGA, ACGA, TACG, GACTA, TATA? Prócz znalezienia przykładowego najkrótszego łańcucha należy uzasadnić, że nie ma krótszego.

Bioinformatyka, czyli biologia wspierana informatyką i matematyką dyskretną

Andrzej P. Urbański, Instytut Informatyki, Politechnika Poznańska

Bioinformatyka jest jedną z najmłodszych nauk, której burzliwy rozwój został wymuszony przez postęp w dziedzinie nauk biologicznych, a umożliwiły go dokonane przełomowe osiągnięcia i wdrożenia w samej informatyce. [cyt. Błażewicz2011]

Wielu badaczy, mówiąc o bioinformatyce, ma na myśli głównie aspekty związane z biologią na poziomie molekularnym (DNA, RNA, białko). Stymulującym to odkryciem było podanie w 1953 r. przez Watsona i Cricka (razem ze współtwórcami Wilkinsem i Franklin nagrodzonymi w 1962 r. Noblem) modelu podwójnej helisy łańcucha DNA, przechowującego (kodującego) informację genetyczną we wszystkich organizmach żywych, co barwnie przedstawia jeden z autorów w swej powieści [Watson1995].

Zadanie „Sklejanie DNA” to przykład problemu biologicznego polegającego na odczytaniu łańcucha DNA (tzw. sekwencjonowaniu DNA) jakiego w naturalnej wielkości nie da się przeprowadzić bez pomocy komputera i odpowiednich programów komputerowych.

W szerszym ujęciu problem przedstawiony w zadaniu polega na odczytaniu łańcucha o długości około 3 miliardów nukleotydów tworzonych przez jedną z czterech zasad: adeninę („A”), guaninę („G”), cytozynę („C”) i tyminę („T”). Metoda zarysowana w zadaniu to jeden z komputerowo wspomaganych sposobów zwany „Sekwencjonowaniem przez hybrydyzację”. Szkic nie uwzględnia na przykład błędów jakie mogą powstać w trakcie odczytywania sekwencji polegających na pojawianiu się odczytów, którym nie odpowiadają żadne rzeczywiste sekwencje (tzw. błędy dodatnie) lub nieodczytywaniu sekwencji, które powinny być w roztworze (tzw. błędy ujemne). Sekwencjonowanie DNA ma dziś ogromne znaczenie w kryminalistyce, sądownictwie, rolnictwie, archeologii, farmaceutyce i medycynie.



Podwójna helisa DNA [Błażewicz2011]

Sekwencja DNA stanowi w pewnym sensie program działania mechanizmów w żywych komórkach, w szczególności pewne jej rejony określają jakie białko zostanie wytworzone a inne, w uproszczeniu, kiedy i gdzie (proces regulacji genów). Białka kodowane są za pomocą kodu genetycznego, w którym trójki zasad kodują pojedynczy aminokwas. Pojedyncze białko jest sekwencją aminokwasów, których standardowo wyróżniamy 20. Większość aminokwasów może być zakodowana w DNA na kilka sposobów (kilka różnych trójek zasad koduje ten sam aminokwas). Sekwencje aminokwasów to inaczej łańcuchy polipeptydowe i mogą one przyjmować złożone struktury przestrzenne. Ustalenie biologicznie aktywnych struktur przestrzennych białek odpowiadających danej sekwencji aminokwasowej jest jednym z ciekawszych problemów z jakim boryka się biologia molekularna wspierana przez tzw. „bioinformatykę strukturalną”. Tutaj z po-

mocą przychodzą zaawansowane metody informatyczne tzw. uczenia maszynowego. Otóż znane z doświadczeń biologicznych pary: sekwencja aminokwasów i struktura białkowa są podawane na wejście programu uczącego się, który po pełnym cyklu uczenia powinien umieć prawidłowo wskazać strukturę dla zadanej sekwencji lub sekwencję dla zadanej struktury (np. przy projektowaniu leków). W chwili obecnej istnieją ogromne bazy danych gromadzące te odpowiedniki i używane w maszynowym uczeniu. Pomimo tego ciągle jeszcze nam daleko do perfekcyjnego prognozowania struktur białek.

Ogólnie rzecz biorąc, absolwent kierunku bioinformatyka wyższej uczelni powinien reprezentować wysoki poziom wiedzy biologicznej jak też wiedzy informatycznej, w tym talentów programistycznych i posługiwania się systemami informatycznymi. W Poznaniu makrokierunek bioinformatyka jest prowadzony we współpracy Wydziału Biologii Uniwersytetu Adama Mickiewicza i Wydziału Informatyki Politechniki Poznańskiej. Na terenie Politechniki Poznańskiej działa Europejskie Centrum Bioinformatyki i Genomiki (ECBiG) jako unikalna na terenie Wielkopolski jednostka badawczo-rozwojowa, powstała na bazie konsorcjum zawiązanego pomiędzy Politechniką Poznańską a Instytutem Chemii Bioorganicznej PAN w Poznaniu.

Literatura:

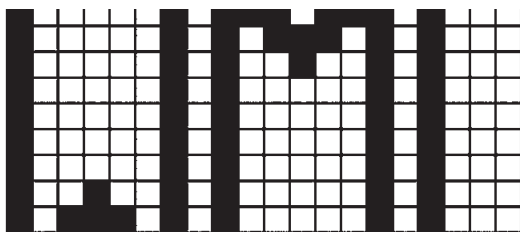
[Błażewicz2011] J. Błażewicz, „Bioinformatyka i jej perspektywy”, wykład inauguracyjny, 2011, Politechnika Poznańska, http://www2.cs.put.poznan.pl/wp-content/uploads/2011/11/wykklad_inauguracyjny_2011.pdf

[Watson1995] J. D. Watson „Podwójna helisa. Historia odkrycia struktury DNA”, Warszawa 1995.

Autor dziękuje panu Maciejowi Miłostanowi za konsultację treści. Rysunek publikujemy za zgodą prof. dr. hab. inż. Jacka Błażewicza.



(C) Hanna Kuik

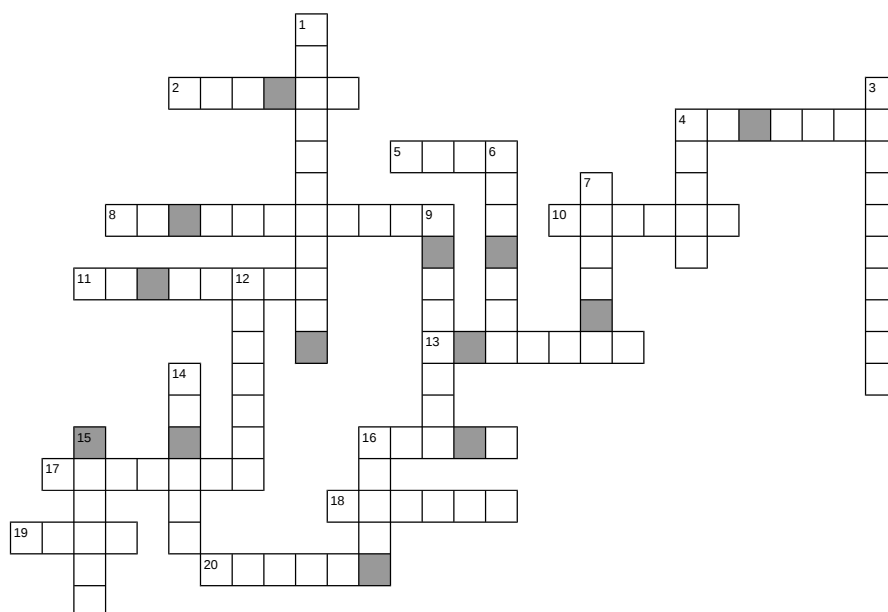


W tym miejscu chcielibyśmy podziękować Hubertowi, Tymonowi i Wojtkowi ze Szkoły Społecznej nr 4 w Poznaniu, którzy zgodzili się wziąć udział w pokazowej lekcji informatyki w czasie konferencji „Informatyka bez komputera”, która 3 listopada odbyła się w Poznaniu. Dziękujemy również pani Agnieszce Kukli, nauczycielce chłopców. W czasie zajęć powtórzono m.in. doświadczenie Shannona wyznaczania średniej ilości informacji odpowiadającej literze w języku angielskim.

Zachęcamy do zapoznania się z nagraniami wideo wykładów, dostępnymi na stronie <http://bezkomputera.wmi.amu.edu.pl/>:

- prof. Tima Bella z Nowej Zelandii, w którym znajdują Państwo cenne przykłady metodyczne, ilustrujące istotę cyfrowego zapisu informacji (kodowania) oraz algorytmów przetwarzających takie cyfrowe dane (np. w grafice komputerowej), w powiązaniu z zagadnieniami projektowania efektywnego interfejsu oraz programowania;
- prof. Michał Armoni z Izraela, wygłoszonego dla pracowników i studentów WMI UAM, w którym przedstawiła poglądowo kluczowe zagadnienia dydaktyki informatyki: rolę abstrakcji, redukcji i dowodów w rozwiązywaniu problemów informatycznych.

Krzyżówka konkursowa



Poziomo

2. Polski informatyk, jego imienia nagrodę przyznaje się młodym informatykom. 4. W XIX wieku opracował dotykowy alfabet do zapisu wyrazów, znaków matematycznych i nut. 5. Amerykański matematyk i informatyk, autor słynnego algorytmu kwantowego. 8. Wybitny polski logik i filozof, powojenny rektor Uniwersytetu im. Adama Mickiewicza w Poznaniu. 10. Laureat nagrody Nobla, rozwiązał tajemnicę kodowania białek. 11. Jeden z trzech, którzy złamali kod Enigmy. 13. Niemiecki matematyk, w XVIII wieku napisał artykuł o systemie dwójkowym. 16. Laureat nagrody Nobla, rozwiązał tajemnicę kodowania białek. 17. Twórca popularnej metody kompresji danych, opartej na kodach prefiksowych. 18. Opracowała polską wersję alfabetu Braille'a. 19. Amerykański wynalazca, jeden z twórców kodu Morse'a. 20. Amerykanka, pionierka informatyki, jako pierwsza użyła określenia „I'm debugging” w odniesieniu do komputera.

Pionowo

1. Twórca słynnej notacji polskiej w logice. 3. Fiński matematyk, jego nazwiskiem nazwano nagrodę przyznaną za osiągnięcia matematyczne w informatyce teoretycznej. 4. Angielski pisarz i filozof z przełomu XV i XVI wieku, twórca sposobu szyfrowania, nazwanego później jego nazwiskiem. 6. Jeden z trzech, którzy złamali kod Enigmy. 7. Polski logik z XX wieku, twórca definicji prawdy w logice. 9. Jeden z trzech, którzy złamali kod Enigmy. 12. Amerykański matematyk, autor dzieła fundamentalnego dla teorii informacji i kodowania. 14. Jeden z twórców algorytmu kryptograficznego RSA. 15. Brytyjski matematyk i kryptolog, uważany za jednego z twórców informatyki. 16. Podobno w I wieku p.n.e. szyfrował wiadomości za pomocą szyfru nazwanego później jego imieniem.

Rozwiązaniem krzyżówki jest słowo, które składa się z liter na szarym tle. Litery trzeba ustawić w odpowiedniej kolejności.

Na rozwiązania czekamy do 5 stycznia 2018 roku. Do wygrania upominki.

Rozwiązania należy przesyłać na adres: koala@vlo.poznan.pl. W temacie prosimy napisać „krzyżówka”.

Biuletyn KOALA został przygotowany przez zespół: Małgorzata Bednarska-Bzdęga, Paweł Perekieta oraz Andrzej P. Urbański.

Ślad komputerowy w systemie \LaTeX przygotował Paweł Perekieta. Rysunki wykonała Hanna Kuik.

Kontakt z redakcją: koala@vlo.poznan.pl