

WALERY SUSŁOW

ADAM SŁOWIK

MICHAŁ STATKIEWICZ

# CHCĘ ZOSTAĆ INFORMATYKIEM



- Poznaj teoretyczne podstawy informatyki
- Naucz się rozwiązywać praktyczne problemy z wykorzystaniem metod informatycznych
- Rozpocznij karierę profesjonalnego programisty

Wszelkie prawa zastrzeżone. Nieautoryzowane rozpowszechnianie całości lub fragmentu niniejszej publikacji w jakiegokolwiek postaci jest zabronione. Wykonywanie kopii metodą kserograficzną, fotograficzną, a także kopiowanie książki na nośniku filmowym, magnetycznym lub innym powoduje naruszenie praw autorskich niniejszej publikacji.

Wszystkie znaki występujące w tekście są zastrzeżonymi znakami firmowymi bądź towarowymi ich właścicieli.

Autor oraz Wydawnictwo HELION dołożyli wszelkich starań, by zawarte w tej książce informacje były kompletne i rzetelne. Nie biorą jednak żadnej odpowiedzialności ani za ich wykorzystanie, ani za związane z tym ewentualne naruszenie praw patentowych lub autorskich. Autor oraz Wydawnictwo HELION nie ponoszą również żadnej odpowiedzialności za ewentualne szkody wynikłe z wykorzystania informacji zawartych w książce.

Redaktor prowadzący: Tomasz Waryszak

Projekt okładki: Studio Gravite / Olsztyn  
Obarek, Pokoński, Pazdrijowski, Zaprucki

Fotografia na okładce została wykorzystana za zgodą Shutterstock.

Wydawnictwo HELION  
ul. Kościuszki 1c, 44-100 GLIWICE  
tel. 32 231 22 19, 32 230 98 63  
e-mail: [helion@helion.pl](mailto:helion@helion.pl)  
WWW: <http://helion.pl> (księgarnia internetowa, katalog książek)

Drogi Czytelniku!

Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres

<http://helion.pl/user/opinie/chcezi>

Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Dodatkowe materiały do książki można znaleźć pod adresem:

<ftp://ftp.helion.pl/przyklady/chcezi.zip>

ISBN: 978-83-246-8525-7

Copyright © Helion 2014

Printed in Poland.

- [Kup książkę](#)
- [Poleć książkę](#)
- [Oceń książkę](#)

- [Księgarnia internetowa](#)
- [Lubię to! » Nasza społeczność](#)

# Spis treści

<b>O autorach</b> .....	<b>13</b>
<b>Słowo wstępne</b> .....	<b>15</b>
<b>Wprowadzenie</b> .....	<b>17</b>
<b>Rozdział 1. Wymiarowanie informacji</b> .....	<b>25</b>
Wstęp .....	25
Zadania .....	28
1.1. Trwałość hasła .....	28
1.2. Umowa na Internet dla telewizora .....	29
1.3. Wbudowana pamięć skanera .....	29
1.4. Pamięć wideo w monitorze .....	29
1.5. Czytnik e-booków .....	29
1.6. Język angielski dla informatyków .....	30
1.7. Kamera wideo .....	30
1.8. Genom ludzki .....	30
1.9. Wyniki kartkówki .....	31
1.10. Tryb High Color .....	31
1.11. Bardzo duży plik wideo .....	31
1.12. Jakość obrazków w gazetce studenckiej .....	32
1.13. Zapisujemy fale dźwiękowe .....	32
1.14. Co to jest digitalizacja? .....	32
1.15. Biblioteka na dysku komputera .....	32
1.16. Skanowane zdjęcie .....	33
1.17. Obliczamy kolor .....	33
Rozwiązania .....	33
1.1. Trwałość hasła .....	33
1.2. Umowa na Internet dla telewizora .....	33
1.3. Wbudowana pamięć skanera .....	34
1.4. Pamięć wideo w monitorze .....	35
1.5. Czytnik e-booków .....	35
1.6. Język angielski dla informatyków .....	36

1.7. Kamera wideo .....	36
1.8. Genom ludzki .....	36
1.9. Wyniki kartkówki .....	37
1.10. Tryb High Color .....	37
1.11. Bardzo duży plik wideo .....	37
1.12. Jakość obrazków w gazecie studenckiej .....	38
1.13. Zapisujemy fale dźwiękowe .....	38
1.14. Co to jest digitalizacja? .....	39
1.15. Biblioteka na dysku komputera .....	39
1.16. Skanowane zdjęcie .....	40
1.17. Obliczamy kolor .....	40
Podsumowanie .....	40
<b>Rozdział 2. Matematyka a informatyka .....</b>	<b>43</b>
Wstęp .....	43
Zadania .....	46
2.1. Naturalne liczby szesnastkowe i ósemkowe .....	46
2.2. Obliczenia geometryczne .....	46
2.3. Siostra i liczby pierwsze .....	46
2.4. Jaka to funkcja? .....	47
2.5. Sumujemy szereg .....	48
2.6. Rysujemy model piramidy 3D .....	48
2.7. Kwadrat prawie jak fraktal .....	48
2.8. Programujemy liczby pierwsze .....	49
2.9. Ciąg arytmetyczny .....	49
2.10. Liczby zaprzyjaźnione .....	49
2.11. Badamy liczby razem z Piotrkim .....	50
2.12. Jak zauważyć tendencję? .....	50
2.13. Rozwiązanie nieliniowego równania .....	50
Rozwiązania .....	51
2.1. Naturalne liczby szesnastkowe i ósemkowe .....	51
2.2. Obliczenia geometryczne .....	51
2.3. Siostra i liczby pierwsze .....	52
2.4. Jaka to funkcja? .....	52
2.5. Sumujemy szereg .....	53
2.6. Rysujemy model piramidy 3D .....	54
2.7. Kwadrat prawie jak fraktal .....	54
2.8. Programujemy liczby pierwsze .....	55
2.9. Ciąg arytmetyczny .....	56
2.10. Liczby zaprzyjaźnione .....	56
2.11. Badamy liczby razem z Piotrkim .....	56
2.12. Jak zauważyć tendencję? .....	59
2.13. Rozwiązanie nieliniowego równania .....	60
Podsumowanie .....	60

<b>Rozdział 3. Logika a informatyka .....</b>	<b>63</b>
Wstęp .....	63
Zadania .....	66
3.1. Logika zapytań w wyszukiwarkach .....	66
3.2. Poszukiwanie prawdziwości w danych .....	67
3.3. Logika w rozgrywkach piłkarskich .....	67
3.4. Szachy, logika i informatyka .....	67
3.5. Turniej tenisa a funkcje boolowskie .....	68
3.6. Tablica prawdy .....	68
Rozwiązania .....	69
3.1. Logika zapytań w wyszukiwarkach .....	69
3.2. Poszukiwanie prawdziwości w danych .....	70
3.3. Logika w rozgrywkach piłkarskich .....	70
3.4. Szachy, logika i informatyka .....	70
3.5. Turniej tenisa a funkcje boolowskie .....	71
3.6. Tablica prawdy .....	72
Podsumowanie .....	72
<b>Rozdział 4. Algorytmika .....</b>	<b>75</b>
Wstęp .....	75
Zadania .....	79
4.1. Badamy systemy liczbowe .....	79
4.2. Porządkujemy liczby .....	79
4.3. Jaki system liczbowy? .....	79
4.4. Analiza algorytmu .....	80
4.5. Rozwiązanie równania metodą Newtona .....	80
4.6. Struktura danych dla koszykarzy .....	81
4.7. Uzupełniamy tablicę danych .....	81
4.8. Jak liczyć litery? .....	82
4.9. Jakie znaki zawiera wyraz (frazę)? .....	82
4.10. Upraszczamy sumator jednobitowy .....	82
4.11. Mówić wspak .....	83
4.12. Notacja polska .....	83
4.13. Anagramy .....	84
4.14. Najmniejsza wartość w zakresie 1...99 .....	84
4.15. Największa wartość dodatnia .....	84
4.16. Bardzo duża tablica .....	85
4.17. Sterujemy tarczą z diod .....	85
4.18. Justyna organizatorem studniówki .....	86
4.19. Kalkulator dla socjologów .....	86
4.20. Dzielimy arkusz .....	86
Rozwiązania .....	87
4.1. Badamy systemy liczbowe .....	87
4.2. Porządkujemy liczby .....	87

4.3. Jaki system liczbowy? .....	87
4.4. Analiza algorytmu .....	88
4.5. Rozwiązanie równania metodą Newtona .....	89
4.6. Struktura danych dla koszykarzy .....	89
4.7. Uzupełniamy tablicę danych .....	90
4.8. Jak liczyć litery? .....	90
4.9. Jakie znaki zawiera wyraz (frazę)? .....	91
4.10. Upraszczamy sumator jednobitowy .....	92
4.11. Mówić wspak .....	93
4.12. Notacja polska .....	93
4.13. Anagramy .....	94
4.14. Najmniejsza wartość w zakresie 1...99 .....	95
4.15. Największa wartość dodatnia .....	95
4.16. Bardzo duża tablica .....	96
4.17. Sterujemy tarczą z diod .....	97
4.18. Justyna organizatorem studniówki .....	98
4.19. Kalkulator dla socjologów .....	99
4.20. Dzielimy arkusz .....	101
Podsumowanie .....	101
<b>Rozdział 5. Programowanie komputerów .....</b>	<b>103</b>
Wstęp .....	103
Zadania .....	108
5.1. Czytaj kod ze zrozumieniem .....	108
5.2. Program szacujący czas podróży .....	109
5.3. Napraw zegar internetowy .....	109
5.4. Program formatuje obrazki .....	110
5.5. Kodołamacz .....	111
5.6. Zabawy z rombem .....	111
5.7. Programujemy arkusze dla speców od wzornictwa .....	112
5.8. Formaty arkuszy .....	113
5.9. Kod dla Ewy .....	113
5.10. Schemat kampusu .....	113
5.11. Kwadratura koła czy kołowość kwadratu? .....	113
5.12. Co w komórce piszczy? .....	114
5.13. Wypełniamy tablicę .....	115
5.14. Nowoczesna legitymacja studencka .....	116
5.15. Błędy logiczne w kodzie .....	116
5.16. Macierze magiczne .....	117
5.17. Oszczędzanie pamięci .....	117
5.18. Popraw kod .....	118
5.19. Zaokrąglenie w Pascalu .....	118
5.20. Kodowanie tabeli w HTML-u .....	118
5.21. Kwadratowa spirala .....	119

Rozwiązania .....	119
5.1. Czytaj kod ze zrozumieniem .....	119
5.2. Program szacujący czas podróży .....	120
5.3. Napraw zegar internetowy .....	121
5.4. Program formatuje obrazki .....	121
5.5. Kodołamacz .....	122
5.6. Zabawy z rombem .....	123
5.7. Programujemy arkusze dla speców od wzornictwa .....	125
5.8. Formaty arkuszy .....	125
5.9. Kod dla Ewy .....	126
5.10. Schemat kampusu .....	127
5.11. Kwadratura koła czy kołowość kwadratu? .....	127
5.12. Co w komórce piszczy? .....	128
5.13. Wypełniamy tablicę .....	128
5.14. Nowoczesna legitymacja studencka .....	128
5.15. Błędy logiczne w kodzie .....	129
5.16. Macierze magiczne .....	130
5.17. Oszczędzanie pamięci .....	131
5.18. Popraw kod .....	131
5.19. Zaokrąglenie w Pascalu .....	131
5.20. Kodowanie tabeli w HTML-u .....	132
5.21. Kwadratowa spirala .....	133
Podsumowanie .....	133
<b>Rozdział 6. Sieci komputerowe .....</b>	<b>135</b>
Wstęp .....	135
Zadania .....	139
6.1. Książki w sieci .....	139
6.2. Aktualizacja Androida a przepustowość sieci .....	139
6.3. Tablica ogłoszeń .....	139
6.4. Tabliczka mnożenia .....	140
6.5. Puzzle z adresu IP .....	140
6.6. Czy to łącze jest szybkie? .....	140
6.7. Unicode .....	140
6.8. Mapa strony .....	141
Rozwiązania .....	141
6.1. Książki w sieci .....	141
6.2. Aktualizacja Androida a przepustowość sieci .....	141
6.3. Tablica ogłoszeń .....	142
6.4. Tabliczka mnożenia .....	142
6.5. Puzzle z adresu IP .....	143
6.6. Czy to łącze jest szybkie? .....	143
6.7. Unicode .....	144
6.8. Mapa strony .....	144
Podsumowanie .....	145

<b>Rozdział 7. Bezpieczeństwo informacji .....</b>	<b>149</b>
Wstęp .....	149
Zadania .....	152
7.1. Poufny SMS .....	152
7.2. Zakodowane terminy informatyczne .....	153
7.3. Szyfrowanie wiadomości telefonicznej .....	153
7.4. Zabezpieczone układy cyfrowe .....	153
7.5. Nieuważny szpieg .....	154
7.6. Ukryte informacje na stronie .....	154
7.7. Własny szyfr .....	155
7.8. Szyfrowanie od przedszkola .....	155
Rozwiązania .....	156
7.1. Poufny SMS .....	156
7.2. Zakodowane terminy informatyczne .....	156
7.3. Szyfrowanie wiadomości telefonicznej .....	157
7.4. Zabezpieczone układy cyfrowe .....	158
7.5. Nieuważny szpieg .....	158
7.6. Ukryte informacje na stronie .....	159
7.7. Własny szyfr .....	159
7.8. Szyfrowanie od przedszkola .....	160
Podsumowanie .....	162
<b>Rozdział 8. Gry z komputerem .....</b>	<b>165</b>
Wstęp .....	165
Zadania .....	167
8.1. Rozwiązujemy sudoku w czterech wariantach .....	167
8.2. Gra w statki .....	170
8.3. Chcę zostać milionerem — gram w lotto .....	170
8.4. Szukamy bomby — gra Saper .....	171
Rozwiązania .....	171
8.1. Rozwiązujemy sudoku .....	171
8.2. Gra w statki .....	174
8.3. Chcę zostać milionerem — gram w lotto .....	174
8.4. Szukamy bomby — gra Saper .....	175
Podsumowanie .....	177
<b>Rozdział 9. Rozwiązywanie problemów za pomocą komputera .....</b>	<b>181</b>
Wstęp .....	181
Zadania .....	184
9.1. Jak zakreskować pole prostokąta? .....	184
9.2. Jak policzyć panele podłogowe? .....	184
9.3. Jak pomóc geodetom? .....	185
9.4. Jaki jest rekord Krzyśka? .....	185
9.5. Jak zrozumieć kod przesyłki kurierskiej? .....	186



9.6. Czy robot wyjdzie z labiryntu? .....	186
9.7. Jak zamontować router? .....	187
9.8. Jak zaprogramować linię produkcyjną? .....	187
9.9. Jak sprawdzić VIN? .....	188
9.10. Jak zaplanować trasę? .....	189
9.11. Jak obliczyć przekładnię zębatą? .....	190
9.12. Jak opisać plan miasta? .....	190
Rozwiązania .....	191
9.1. Jak zakreśkować pole prostokąta? .....	191
9.2. Jak policzyć panele podłogowe? .....	192
9.3. Jak pomóc geodetom? .....	193
9.4. Jaki jest rekord Krzyśka? .....	194
9.5. Jak zrozumieć kod przesyłki kurierskiej? .....	195
9.6. Czy robot wyjdzie z labiryntu? .....	195
9.7. Jak zamontować router? .....	196
9.8. Jak zaprogramować linię produkcyjną? .....	196
9.9. Jak sprawdzić VIN? .....	196
9.10. Jak zaplanować trasę? .....	198
9.11. Jak obliczyć przekładnię zębatą? .....	199
9.12. Jak opisać plan miasta? .....	201
Podsumowanie .....	203
<b>Zakończenie .....</b>	<b>205</b>
<b>Tabela kodów ASCII .....</b>	<b>209</b>



# Rozdział 1.

## Wymiarowanie informacji

### Wstęp

Jednym z najważniejszych zadań techniki jest przekazywanie wiadomości w przestrzeni i czasie. Nowoczesne środki komunikacji elektronicznej umożliwiają przekazywanie wiadomości jednocześnie z ich przetwarzaniem. Na przykład w komunikatorach w rodzaju *Gadu-Gadu*, *Skype* lub *Google Talk*, realizujących między innymi rozmowy głosowe, wykorzystywane są algorytmy zwiększające wyrazistość mowy i filtrujące niektóre rodzaje hałasu. Funkcjonalność takiego oprogramowania pozwala zawiesić rozmowę, nagrać treść przekazywanej wiadomości lub przesłać w trakcie rozmowy tekst, materiał wideo czy plik.

Termin **informacja** (łac. *informatio*) w kontekście kanałów informacyjnych oznacza wiadomość, uświadomienie sobie czegoś. Zatem nie chodzi tylko o przesyłanie danych, ale o odtwarzanie ich sensu. Rozpowszechnione definicje informacji noszą znamiona opisu czy objaśnienia i nie są tak ścisłe jak definicje pojęć fizycznych, bo „informacja” jest pojęciem podstawowym. Np. informacją nazywamy to wszystko, co można wykorzystać do sprawnego podejmowania decyzji. Całość zagadnień związanych z wydobywaniem, przechowywaniem, przekazywaniem, przetwarzaniem i prezentowaniem informacji nazywamy nauką o informacji. Stosowane podejście do tej nauki pierwotnie było rozwijane przez cybernetykę, która zajmowała się systemami sterowania oraz związanym z tym przetwarzaniem i przekazywaniem informacji. Obecnie teoria informacji jest składnikiem wiedzy bazowej nowoczesnej informatyki, ale także i bibliotekoznawstwa, a nawet ekonomii.

Obecnie dość sprawnie rozwijana jest **eksploracja danych** — dziedzina informatyki zwana z angielskiego *Data mining*. Ogólnie rzecz biorąc, *Data mining* (czasami nazywany też odkrywaniem wiedzy) to proces analizy danych z różnych perspektyw, który sprowadza dane do poziomu przydatnych informacji. Najczęściej chodzi o informacje, które mogą być wykorzystane do zwiększenia przychodów i obniżenia kosztów organizacji. Technicznie rzecz ujmując,

*data mining* to proces znajdowania korelacji lub wzorów wśród kilkudziesięciu pól w dużych relacyjnych bazach danych. Tak więc pojęcie *Data mining* określa sposób odkrywania nowego znaczenia (sensu) w dużych bazach danych poprzez ich przetwarzanie przy użyciu zaawansowanych możliwości algorytmów statystycznych i odkrywania wzorców informacyjnych. Oprogramowanie eksploracji danych jest jednym z wielu narzędzi analitycznych do analizowania danych. Pozwala ono użytkownikom na kategoryzowanie danych i podsumowywanie zidentyfikowanych relacji.

Informację można klasyfikować według dziedzin wiedzy, np. techniczna, ekonomiczna, historyczna, ale i według fizycznego charakteru jej rejestrowania przez człowieka, np. wzrokowa, słuchowa, smakowa. Najbardziej odpowiednią dla zagadnień technicznych wydaje się klasyfikacja wg właściwości metrologicznych: informacja parametryczna, np. wyniki pomiarów, informacja topologiczna, np. formy przestrzenne, informacja abstrakcyjna, np. formuły matematyczne itd. Za przykład przydatnej w szkole informacji parametrycznej mogą posłużyć wynikowe oceny z kartkówki; umożliwią one rankingowanie uczniów według posiadanej wiedzy i umiejętności z danego przedmiotu.

W większości przypadków informacja nie jest dostępna bezpośrednio, dostępny jest tylko **sygnał** lub **nośnik** zawierający informację. Zawartość informacji w sygnale oznacza, że niektóre jego parametry zależą od informacji. W wyniku przekształceń sygnału wydobywa się informację, która może odwzorowywać realne stany obiektów i procesów. Dane i informacje są szczególnie narażone na zniekształcenia i błędy podczas transmisji. Pewnie kojarzysz grę „głuchy telefon”. Istota tej zabawy polega na tym, że przekazywana wiadomość jest stopniowo zniekształcana i jej końcowe brzmienie może być całkowicie odmienne od pierwotnego. W systemach komunikacji elektronicznej, aby uniknąć zniekształceń, stosowane są **kody korekcyjne**, korygujące błędy. Są to kody nadmiarowe, które umożliwiają wykrycie i automatyczne usunięcie błędów w przekazywanych wiadomościach. Jednym z najbardziej znanych i efektywnych kodów korekcyjnych jest **kod Hamminga**. Wykrywa on i koryguje przekłamanie pojedynczego bitu w jednostce kodowania, może również wykryć dwa jednocześnie przekłamanie bity, ale nie skoryguje ich.

Głównym problemem teorii informacji jest ustalanie miar ilości i jakości informacji. Znane miary informacji odpowiadają trzem podejściom do teorii informacji: strukturalnemu, statystycznemu i semantycznemu. Przy podejściu strukturalnym przedmiotem analizy jest dyskretny charakter informacji, a pomiar ilości informacji odbywa się z wykorzystaniem metod kombinatorycznych. Właśnie to podejście jest przydatne do wytłumaczenia wymaganej minimalnej długości hasła dostępu np. do serwisu internetowego lub konieczności jednoczesnego wykorzystania w składzie hasła liter, cyfr i znaków specjalnych. Dłuższe hasło zawiera więcej informacji, a hasło złożone z rozmaitych znaków zawiera więcej informacji niż dowolne znane słowo języka naturalnego o tej samej długości. A co za tym idzie, im więcej informacji dostarcza hasło, tym trudniejsze jest zadanie dla ewentualnego hakera, aby to hasło złamać. Przy statystycznym podejściu przedmiotem analizy są probabilistyczne własności zdarzeń, a miarą informacji jest **entropia** (mogłeś o niej słyszeć na zajęciach z fizyki). W ramach tego podejścia informacja powiązana z mniej prawdopodobnym komunikatem lub z bardzo rzadkim zdarzeniem jest większa niż w sytuacji z przewidywalnym komunikatem albo często i regularnie wystę-

pującym zdarzeniem. Na przykład prognoza nawałnicy jest bardziej informatywna (dostarcza więcej informacji) niż komunikat o niedługim nadejściu wiosny. Przy semantycznym podejściu przedmiotem analizy jest celowość, użyteczność i znaczenie pragmatyczne informacji. Przy takim podejściu bardziej informatywne są komunikaty, które odpowiadają celom komunikacji, są pożyteczne dla odbiorcy i mają dla niego znaczenie praktyczne. Na przykład merytoryczna odpowiedź dla osoby skupionej w danej chwili nad rozwiązaniem pilnego zadania jest bardziej informatywna od powiadomienia organizacyjnego lub od dowolnego komunikatu niezwiązanego z tym zadaniem.

Teoria informacji wyróżnia **dyskretne** źródła informacji i dyskretne wiadomości. Wiadomości dyskretne składają się z przeliczalnych zbiorów elementów, czasem są to sekwencje czasowe. Zbiór elementów takich wiadomości nazywamy alfabetem. Liczba elementów alfabetu (znaków, liter) to jego objętość. W przeciwieństwie do wiadomości dyskretnych, w teorii występują wiadomości ciągłe, tworzą je zmienne wielkości fizyczne, np. dane pogodowe. W praktyce tego rodzaju sygnały poddawane są dyskretyzacji (**próbkowaniu** i **kwantowaniu**) w celu sprowadzenia ich do postaci wiadomości dyskretnych. Na przykład utwory zapisane w plikach muzycznych przechowują dźwięk w postaci cyfrowej. Powstaje on w procesie dyskretyzacji, kiedy sygnał analogowy z mikrofonu zostaje poddany próbkowaniu — zamiast zapisu ciągłego zostają zapisane próbki (chwilowe wartości sygnału) zmierzone w ustalonych odstępach czasu. Aby nie dało się usłyszeć przekłamań (aby możliwe było odtworzenie dźwięku bez zniekształceń), częstotliwość próbkowania dźwięku musi być większa od podwojonej największej wartości częstotliwości, jaka występuje w próbkowanym sygnale. Jeśli założymy, że częstotliwości sygnału audio zawierają się w przedziale od 20 Hz do 20 kHz, wówczas częstotliwość próbkowania musi być większa niż 40 kHz. Z tego też względu między innymi przyjęty standard nagrywania płyt CD-Audio to 44 100 próbek na sekundę (częstotliwość próbkowania 44,1 kHz). Próbkowany dźwięk jednocześnie zostaje poddany kwantowaniu (kwantyzacji) — stratnej konwersji chwilowych wartości sygnału na postać cyfrową. Dla CD-Audio każdej wartości zostaje przypisany jeden z kodów dwójkowych dostępnych przy szesnastobitowym kodowaniu od 0 do 65 535. Dodatkowo cyfrowy dźwięk (podobnie jak cyfrowe fotografie i wideo) często poddawany jest dodatkowej kompresji stratnej, przy użyciu której zmniejsza się objętość pliku, a słuchacz praktycznie nie zauważa pogorszenia jakości dźwięku. Tak działa między innymi format MP3, który wykorzystuje model **psychoakustyczny**. Zgodnie z tym modelem z zapisu zostają usunięte te informacje o dźwięku, które są niezauważalne lub mało istotne dla człowieka, w sensie jakości odbioru słuchowego.

Jeszcze jeden aspekt informacji zawsze powinien być brany pod uwagę w praktyce. Chodzi tutaj mianowicie o doskonałość, pewność, precyzyjność posiadanych informacji. W tym sensie informacja niedoskonała tylko zmniejsza niepewność sytuacji, w jakiej znajduje się odbiorca informacji, ale nie eliminuje tej niepewności. Może to utrudniać np. podejmowanie decyzji biznesowych. Realnie zawsze należy dopuszczać pewną nieścisłość informacyjną komunikatu, rozmytość zawartego w nim sensu. Właśnie taką informacją my, ludzie, się posługujemy, tzn. są dla nas jasne pojęcia „około dziesięciu”, „za chwilę”, „wszyscy to wiedzą”, mimo że są one nieprecyzyjne. Informatycy używają tak zwanej logiki rozmytej, aby te nieprecyzyjności modelować komputerowo i tworzyć systemy informatyczne o wyższym poziomie inteligencji.

W niniejszym rozdziale zebrane są zadania przewidujące zastosowanie miar informacji w sytuacjach związanych z podejmowaniem decyzji technicznych lub organizacyjnych. Ich rozwiązanie nie wymaga głębokiej wiedzy naukowej z zakresu teorii informacji. Generalnie potrzebna jest wiedza z zakresu podstaw technologii informacyjnych i zdrowy rozsądek oraz logika. Rozwiązując te zadania, proszę pamiętać, że informatyk to specjalista od przetwarzania informacji — aby mógł panować nad informacją, musi nauczyć się ją mierzyć. Aby poznać podstawową wiedzę w zakresie teorii informacji i zagadnień pokrewnych, można zajrzeć do następujących książek:

1. Henryk Górecki, *Teoria informacji*, Wydawnictwo Wyższej Szkoły Informatyki w Łodzi, Łódź 2006.
2. Józef Oleński, *Ekonomika informacji. Podstawy*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2001.
3. Paweł Przybyłowicz, *Wstęp do teorii informacji i kodowania*, Centrum Modelowania Matematycznego Sigma, 14 maja 2008, <http://www.scribd.com/doc/49766365/wstep-do-teorii-informacji>.
4. Gleick James, *Informacja. Bit, wszechświat, rewolucja*, Społeczny Instytut Wydawniczy Znak, Kraków 2012.
5. Magdalena Karciarz, Maciej Dutko, *Informacja w Internecie*, Wydawnictwo Naukowe PWN, Warszawa 2010.

## Zadania

### 1.1. Trwałość hasła

Po rozpoczęciu studiów Michał chciał założyć sobie konto na serwerze pocztowym uczelni. Wpisane przez niego do formularza rejestracyjnego dziewięcioliterowe hasło składało się tylko z małych liter (klawiatura o układzie klawiszy typu **QWERTY**). Chciał on dowiedzieć się, ile maksymalnie czasu zajęłoby złamanie jego hasła, ale nikt ze studentów pierwszego roku nie potrafił dać mu dokładnej odpowiedzi. Pomóż Michałowi, oszacuj czas złamania jego hasła metodą **brute force**<sup>1</sup> przy założeniu, że haker dysponuje komputerem z procesorem 3 GHz, a do sprawdzenia jednego potencjalnego hasła komputer potrzebuje 100 taktów zegara systemowego.

---

<sup>1</sup> Algorytm siłowy, który opiera się na sukcesywnym sprawdzaniu wszystkich możliwych kombinacji dostępnych znaków.

## 1.2. Umowa na Internet dla telewizora

Rodzice Tomek kupili telewizor LED Full HD z dekodowaniem MPEG-4. Telewizor ten posiada technologię **Smart TV** (możliwość przyłączenia Internetu). Tomek, student elektroniki, wie, że można oglądać na nim programy telewizyjne nadawane w Internecie w trzech różnych rozdzielczościach: **Standard** (720 na 576 pikseli), **HD Ready** (1366 na 768 pikseli) oraz **Full HD** (1920 na 1080 pikseli), i chce udowodnić rodzicom, że potrzebują zwiększenia przepustowości posiadanego czteromegabitowego łącza internetowego.

Pomóż Tomkowi, oszacuj minimalną przepustowość łącza internetowego w Mb/s, niezbędną dla każdej rozdzielczości, aby obraz był odtwarzany płynnie (minimum 30 klatek na sekundę). Przy obliczeniach należy przyjąć, że kolor każdego piksela jest kodowany 24 bitami oraz że przesyłane dane są w formacie MPEG-4 z poziomem kompresji 1:60 w stosunku do obrazu nieskompresowanego.

## 1.3. Wbudowana pamięć skanera

Skaner ręczny ma wbudowaną pamięć **flash**. Jeżeli do skanowania użyjemy palety barw z  $2^{32}$  kolorów, to w danej pamięci można розміścić 7 obrazków i jeszcze zostanie wolnych 2,25 MB. Jeżeli użyjemy zredukowanej palety barw składającej się z  $2^{24}$  kolorów, to wówczas w tej samej pamięci zmieści się dokładnie 10 obrazków. Wewnętrzny format przechowywania zeskanowanych obrazków przewiduje stały rozmiar pliku, w którym oprócz grafiki znajduje się 256 kB informacji — są to **metadane** opisujące obraz cyfrowy. Oblicz, ile wynosi objętość wbudowanej pamięci danego skanera.

## 1.4. Pamięć wideo w monitorze

Monitor pewnej firmy, przeznaczony do montowania w obrabiarkach CNC (z układem sterowania numerycznego) posiada kolorowy ekran wyświetlający 256 kolorów. Rozmiar wyświetlanego obrazu wynosi 7,5 cm na 12,5 cm. Dołączona do monitora pamięć wideo ma objętość  $1,35 \cdot 10^4$  bajtów. Z jaką rozdzielczością obraz jest wyświetlany na ekranie tego monitora? (Odpowiedź proszę podać w punktach na cal. Należy przyjąć, że 1 cal wynosi 2,5 cm). Odpowiedz dodatkowo, czy ta rozdzielczość jest zbliżona do rozdzielczości wykorzystywanych przez Ciebie ekranów?

## 1.5. Czytnik e-booków

Tomek, student drugiego roku, chcąc zapoznać się z nowym gadżetem, pożyczył od kolegi w akademiku czytnik e-booków wyprodukowany przez firmę Amazon. W czytniku była zainstalowana pamięć 1 GB. Tomek kilkakrotnie zapełniał pamięć czytnika maksymalnie. W tabeli 1.1 przedstawiono 5 przypadków zapełnienia pamięci czytnika e-booków, łącznie z liczbą książek, jaką udało się w nim Tomkowi zmieścić.

**Tabela 1.1.** Wyniki pełnego wykorzystania pamięci 1 GB przy kolejnych próbach wgrania dużej ilości książek do czytnika

Lp.	Liczba książek
1	1020
2	789
3	654
4	943
5	654

Oblicz w megabajtach średni rozmiar książki w wersji elektronicznej oraz podaj odchylenie standardowe dla wyznaczonego średniego rozmiaru książki.

## 1.6. Język angielski dla informatyków

Informatycy muszą znać język angielski, bo większość popularnych technologii i narzędzi informatycznych jest opracowywana z użyciem właśnie tego języka. Studenci pierwszego roku politechniki mają do wyboru jeden z trzech języków obcych: angielski, niemiecki i francuski. W jednej z grup kierunku informatyka 12 studentów nie wybrało języka angielskiego. Żartując na temat tego faktu, pewien wykładowca przedmiotu teoria informacji powiedział, że komunikat o tym, że przypadkowo wybrany student z tej grupy wybrał język angielski, dostarcza mu  $\log_2 3$  bitów informacji, a że inny przypadkowo wybrany student wybrał język francuski — dokładnie 1 bit informacji. Czy możesz powiedzieć, ilu studentów w tej grupie wybrało język niemiecki?

## 1.7. Kamera wideo

Janek nakręcił film ze studniówki swoją kamerą JVC Everio (kamera na karty SD). Po zgraniu pliku wideo na dysk komputera film zajmował 5 134 567 367 bajtów. Kolega, który przyszedł do Janka po ten film, przyniósł ze sobą trzy pendrive'y o pojemności 512 MB, 1 GB oraz 2 GB. Przed podziałem filmu na mniejsze pliki Janek skompresował materiał wideo, uzyskując plik o wielkości 85% pliku oryginalnego. Czy uda się Jankowi tak podzielić skompresowany film, aby kolega mógł go zabrać do domu na tych trzech pendrive'ach? Ile zostanie wolnego miejsca lub ile miejsca zabraknie na pendrive'ach?

## 1.8. Genom ludzki

Pewnie niejednokrotnie słyszałeś o naukowych osiągnięciach współczesnej genetyki. Ciekawostką jest, że sekwencję każdego DNA można przedstawić jako słowo zapisane w alfabecie, który zawiera tylko cztery litery {A, C, G, T} pochodzące od nazw zasad azotowych będących składnikami DNA: adeniny (A), cytozyny (C), guaniny (G) i tyminy (T). Zapis genomu ludz-



kiego, który jest reprezentowany przez całą sekwencję DNA, składa się z około 3 miliardów takich liter. Jaka, Twoim zdaniem, powinna być minimalna pojemność nośnika pamięci, aby można było na nim zapisać tekst kodu genetycznego pojedynczej osoby bez kompresji?

## 1.9. Wyniki kartkówki

Grupa 32 studentów napisała kartkówkę z podstaw informatyki. Nauczyciel ogłosił wyniki kartkówki na swojej stronie internetowej w sposób następujący: „Niestety, część grupy, ze starostą na czele, nie zaliczyła kartkówki. Zawiadamiam, że poprzednie zdanie zawiera tylko 4 bity informacji”. Ilu studentów zaliczyło kartkówkę?

## 1.10. Tryb High Color

Pewnie wiesz, że *High Color* to metoda zapisywania informacji obrazkowej w pamięci komputera i połączona z nią technika wyświetlania barw na ekranie graficznym. Przy szesnastobitowej głębi kolorów umożliwia ona wyświetlanie 65 536 różnych kolorów. Oszacuj, ile megabajtów pamięci wideo potrzeba do realizacji tego trybu graficznego przy rozdzielczości 1024x768 pikseli. Jak, Twoim zdaniem, pamięć ta powinna być podzielona pomiędzy kolory czerwony, zielony i niebieski?

### Komentarz:

1. Ostatnie pytanie nie wymaga precyzyjnej odpowiedzi, a jedynie sformułowania idei rozwiązania technicznego.
2. **Podpowiedź:** oko człowieka jest bardziej czułe na kolor zielony, niż na czerwony i niebieski.

## 1.11. Bardzo duży plik wideo

Piotrek uczestniczy w zajęciach z technologii multimedialnych. W wyniku obróbki wideo otrzymał plik o wielkości 123 456 789 000 B. Na ile mniejszych plików musi go podzielić, aby można go było zarchiwizować na dyskach o następujących systemach plikowych:

- a) dysk A — FAT16,
- b) dysk B — FAT32,
- c) dysk C — NTFS,
- d) dysk D — APFS?

## 1.12. Jakość obrazków w gazetce studenckiej

Podczas pracy nad reklamą dla gazetki studenckiej obrazek o rozmiarze 3 na 5 cali narysowano w edytorze graficznym i zapisano na dysku w formacie mapy bitowej z wykorzystaniem palety 256 kolorów. W jakiej rozdzielczości w punktach na cal (ang. *dot per inch* — dpi) należy drukować dany obrazek, by osiągnąć najlepszą jakość wydruku, jeśli rozmiar pliku graficznego wyniósł  $432 \cdot 10^5$  bitów?

## 1.13. Zapisujemy fale dźwiękowe

W Laboratorium Cyfrowego Przetwarzania Sygnałów należało dokonać pomiarów fali dźwiękowej o czasie trwania równym 30 minut. Pomiar ten związany był z pracami nad nową metodą służącą do automatycznego rozpoznawania dźwięku mówionego (ang. *automatic speech recognition*). W tym celu dźwięk, który miał podlegać dalszej analizie, musiał zostać nagrany na dysk twardy komputera przy użyciu mikrofonu i wyspecjalizowanej karty dźwiękowej. W trakcie badań przyjęto, że częstotliwość próbkowania powinna wynosić 44,1 kHz, oraz określono, że pojedyncza próbka powinna być kodowana przy użyciu 24 bitów. Wyznacz, ile miejsca na dysku twardym komputera będą zajmować dane opisujące badany dźwięk.

## 1.14. Co to jest digitalizacja?

Digitalizacją będziemy nazywać zamianę danych analogowych na postać cyfrową. Zakładamy, że zakres zmienności danych analogowych dzieli się na równe przedziały  $\beta$ .

Jaka jest najmniejsza liczba bitów, aby możliwe było zakodowanie w ciągu binarnym wartości analogowych z zakresu zmienności  $\langle -5; 8 \rangle$ , przy założeniu że przedział digitalizacji  $\beta$  jest nie mniejszy niż 0,01? Zastanów się, czy przy takiej digitalizacji można użyć kodowania ośmio-bitowego lub szesnastobitowego.

## 1.15. Biblioteka na dysku komputera

Księgozbiór pewnej biblioteki liczy 208 566 książek. Przyjmijmy, że każda książka średnio składa się z 500 stron. Na każdej stronie można zapisać 30 linii tekstu, a w każdej linii znajduje się 70 znaków. Ile takich książek można przechować na dysku o dostępnej pojemności użytkowej 160 GB, przy założeniu, że każdy znak jest zapisany:

- a) w kodzie ASCII,
- b) w kodzie UTF?

Czy wszystkie książki z biblioteki przy takim założeniu mogłyby zmieścić się na jednym dysku? Odpowiedź uzasadnij.

## 1.16. Skanowane zdjęcie

Mamy kolorowe zdjęcie o wymiarach 10 na 15 cm. Chcemy je zeskanować w rozdzielczości 300 ppi (dpi), korzystając z modelu barw RGB oraz CMYK. Jaki rozmiar w megabajtach będzie miał plik wynikowy, jeśli przyjmujemy, że do zakodowania pojedynczej składowej koloru użyto 8 bitów.

Rozmiary plików wyznacz dla obu modeli barw.

## 1.17. Obliczamy kolor

Wiadomo, że oko człowieka jest najbardziej czułe na kolor zielony, a najmniej na kolor niebieski. Projektanci kolorowych monitorów przyjmują, że człowiek rozróżnia kolory kodowane w systemie RGB zgodnie z współczynnikami korekcji 30 (Red), 59 (Green), 11 (Blue). Załóżmy, że mamy do zbadania pewien wzorec koloru, np. taki jak w sklepach z farbami. Zaznaczmy kolor tego wzorca za pomocą składowych jako  $C_w = (R_w, G_w, B_w)$ . Nie mamy gwarancji, że monitor będzie w stanie precyzyjnie odwzorować kolor wzorca, np. na zdjęciu fotograficznym, bo ma on ograniczoną paletę kolorów. Wynika to z ograniczenia długości kodu binarnego zapisującego kolor w pamięci komputera. Zaproponuj sposób odnajdywania koloru najbliższego wzorcowi z dostępnej palety  $C_i = [R_i, G_i, B_i]$  zawierającej  $N$  różnych kolorów.

# Rozwiązania

## 1.1. Trwałość hasła

Hasło zawiera 9 znaków z alfabetu 26 liter, bo tyle małych liter można wprowadzić z klawiatury o układzie klawiszy typu QWERTY. Wszystkich kombinacji hasła złożonego z tych liter jest zatem 26 do potęgi 9 = 5 429 503 678 976.

Procesor z zegarem 3 GHz wykonuje  $3 \cdot 10^9$  taktów na sekundę. Wiemy, że w ciągu 100 taktów zegarowych jest sprawdzany jeden wariant hasła, czyli w ciągu sekundy można sprawdzić  $3 \cdot 10^7$  haseł. Aby złamać hasło Michała, będziemy potrzebowali maksymalnie około 180 983,46 sekund, czyli około 50 godzin, 16 minut i 23 sekund pracy komputera.

## 1.2. Umowa na Internet dla telewizora

Objętość informacyjna jednej klatki filmu bez kompresji wynosi (w bitach):

- a) w rozdzielczości Standard =  $720 \cdot 576 \cdot 24 = 9\,953\,280$  b,
- b) w rozdzielczości HD Ready =  $1366 \cdot 768 \cdot 24 = 25\,178\,112$  b,

c) w rozdzielczości Full HD =  $1920 \cdot 1080 \cdot 24 = 49\,766\,400$  b.

Jedna sekunda filmu bez kompresji potrzebuje transmisji (bitów):

- a) Standard =  $9\,953\,280 \text{ b} \cdot 30 = 298\,598\,400 \text{ b}$ ,
- b) HD Ready =  $25\,178\,112 \text{ b} \cdot 30 = 755\,343\,360 \text{ b}$ ,
- c) Full HD =  $49\,766\,400 \text{ b} \cdot 30 = 1\,492\,992\,000 \text{ b}$ .

Jedna sekunda filmu z kompresją MPEG-4 potrzebuje transmisji (bitów):

- a) Standard =  $298\,598\,400 \text{ b} : 60 = 4\,976\,640 \text{ b}$ ,
- b) HD Ready =  $755\,343\,360 \text{ b} : 60 = 12\,589\,056 \text{ b}$ ,
- c) Full HD =  $1\,492\,992\,000 \text{ b} : 60 = 24\,883\,200 \text{ b}$ .

Minimalna przepustowość łącza internetowego wynosi (dla poszczególnych rozdzielczości):

- a) Standard =  $4\,976\,640 \text{ b/s} : 1024 = 4860 \text{ kb/s} : 1024 = 4,746 \text{ Mb/s}$ ,
- b) HD Ready =  $12\,589\,056 \text{ b/s} : 1024 = 12\,294 \text{ kb/s} : 1024 = 12,006 \text{ Mb/s}$ ,
- c) Full HD =  $24\,883\,200 \text{ b/s} : 1024 = 24\,300 \text{ kb/s} : 1024 = 23,730 \text{ Mb/s}$ .

Tak więc przepustowość, jaką dysponujemy, nie wystarcza nawet do płynnego oglądania telewizji internetowej w rozdzielczości Standard.

### 1.3. Wbudowana pamięć skanera

Niech  $Y$  oznacza nieznaną (stałą) liczbę pikseli w każdym obrazku. Wtedy możemy napisać równanie:

$$\begin{aligned} 7 \cdot (Y \cdot \log_2(2^{32}) + 256 \cdot 1024 \cdot 8) + 2,25 \cdot 1024 \cdot 1024 \cdot 8 &= \\ = 10 \cdot (Y \cdot \log_2(2^{24}) + 256 \cdot 1024 \cdot 8) \end{aligned}$$

Wykonując przekształcenia, dochodzimy do odpowiedzi:

$$\begin{aligned} 7 \cdot 256 \cdot 1024 \cdot 8 + 2,25 \cdot 1024 \cdot 1024 \cdot 8 - 10 \cdot 256 \cdot 1024 \cdot 8 &= \\ = 10 \cdot Y \cdot 24 - 7 \cdot Y \cdot 32 \end{aligned}$$

$$2^3 \cdot 2^{10} \cdot 2^8 \cdot (7 + 2,25 \cdot 2^2 - 10) = Y \cdot (24 \cdot 10 - 7 \cdot 32)$$

$$Y = 3 \cdot 2^{18} \text{ (pikseli)}$$

Zatem całą objętość pamięci obliczamy następująco:

$$10 \cdot (3 \cdot 2^{18} \cdot \log_2(2^{24}) + 256 \cdot 1024 \cdot 8) : (8 \cdot 1024 \cdot 1024) =$$

$$10 \cdot (3 \cdot 2^{18} \cdot 3 \cdot 2^3 + 2^{21}) : 2^{23} =$$

$$10 \cdot 2^{21} \cdot (9 + 1) : 2^{23} =$$

$$2 \cdot 5 \cdot 2^{21} \cdot 2 \cdot 5 : 2^{23} =$$

$$25 \cdot 2^{23} : 2^{23} = 25 \text{ MB}$$

## 1.4. Pamięć wideo w monitorze

Przeanalizujmy dane wejściowe. Przy liczbie kolorów  $N = 256$  wymagana długość słowa kodującego kolor pojedynczego punktu wynosi  $i = \log_2 256 = 8$  bitów. Zarezerwowana objętość pamięci wideo wynosi  $P_w = 10\,800\,000$  bitów. Powierzchnia ekranu wynosi  $7,5 \cdot 12,5 \text{ cm}^2 = 3 \cdot 5 \text{ cal}^2 = 15 \text{ cal}^2$ .

Oznaczmy przez  $x^2$  liczbę punktów znajdujących się na powierzchni w jednym calu kwadratowym, wtedy ekran powinien zawierać  $k = x^2 \cdot 15$  punktów. Jako że każdy punkt wymaga 8 bitów informacji, ogólna ilość informacji niezbędna do sterowania ekranem wynosi  $I = 120x^2$  bitów. Właśnie do przechowywania tej informacji przeznaczona jest pamięć wideo, więc przyrównując  $I$  do  $P_w$ , otrzymujemy równanie  $120x^2 = 10800000$ . Stąd  $x^2 = 90000$ , a  $x = 300$  punktów na cal, czyli 300 ppi (ang. *pixels per inch*). Jest to bardzo wysoka rozdzielczość, porównywalna z tą charakterystyczną dla wyświetlaczy Apple iPhone i lepszych modeli Samsung Galaxy.

## 1.5. Czytnik e-booków

1 GB to 1024 MB. W związku z tym na jedną wczytaną książkę przypada następujący rozmiar ( $RK_i$ ) w MB (patrz tabela 1.2).

**Tabela 1.2.** Wyniki obliczenia średniej objętości książek w kolejnych próbach

Lp.	Liczba książek	Liczba MB na 1 wczytaną książkę
1	1020	$RK_1 = 1024:1020 \approx 1,004 \text{ MB}$
2	789	$RK_2 = 1024:789 \approx 1,298 \text{ MB}$
3	654	$RK_3 = 1024:654 \approx 1,566 \text{ MB}$
4	943	$RK_4 = 1024:943 \approx 1,086 \text{ MB}$
5	654	$RK_5 = 1024:654 \approx 1,566 \text{ MB}$

Średni rozmiar książki w badaniach Tomka wynosi:

$$Avg_{RK} = (1,004 \text{ MB} + 1,298 \text{ MB} + 1,566 \text{ MB} + 1,086 \text{ MB} + 1,566 \text{ MB}) : 5 \approx 1,304 \text{ MB}$$

Odchylenie standardowe dla rozmiaru książki z próby pięcioelementowej ( $n = 5$ ) wyznaczamy z zależności:

$$Std = \sqrt{\frac{\sum_{i=1}^n (RK_i - Avg_{RK})^2}{n-1}}$$

Czyli w tym przypadku:

$$\text{Std} = \sqrt{\frac{\sum_{i=1}^5 (\text{RK}_i - \text{Avg}_{\text{RK}})^2}{5-1}} = \sqrt{\frac{0,274848}{4}} = \sqrt{0,068712} = 0,26213 \text{ MB}$$

Ostatecznie średni rozmiar książki w wersji elektronicznej wynosi około  $1,3 \pm 0,3$  MB.

## 1.6. Język angielski dla informatyków

Najpierw ustalmy, jak duża jest ta grupa. Niech język angielski będzie wybrany przez  $A$  studentów. Skoro 12 studentów nie wybrało języka angielskiego, to grupa składa się z  $n = (A + 12)$  studentów. Informacja o wyborze przez jednego ze studentów języka angielskiego może być obliczona następująco:  $i_A = \log_2[(12 + A) : A] = \log_2 3$ . Obliczamy  $A = 6$ , wtedy  $n = 12 + 6 = 18$ , więc grupa jest osiemnastoosobowa.

Niech język francuski będzie wybrany przez  $F$  studentów, wtedy informacja o wyborze przez drugiego studenta języka francuskiego może być obliczona ze wzoru:  $i_F = \log_2(18 : F) = 1$ . Wynika stąd, że  $F = 9$ , więc liczba studentów  $N$ , którzy wybrali język niemiecki, wynosi:  $N = (n - A - F) = (18 - 6 - 9) = 3$ .

## 1.7. Kamera wideo

Jest to bardzo proste zadanie — aby je rozwiązać, wystarczy przedstawić wszystkie dane w tych samych jednostkach. Wybierzmy gigabajt jako jednostkę do porównania wielkości pliku z pojemnością nośników pamięci. Wtedy  $5\,134\,567\,367$  bajtów to  $5\,014\,225,944$  kB =  $4896,705$  MB =  $4,78$  GB. Po kompresji wielkość pliku wynosi  $4,78 \text{ GB} \cdot 0,85 = 4,063$  GB. W związku z tym kolega nie zabierze do domu pełnego filmu ze studniówki, ponieważ dysponuje on pamięcią  $0,5 + 1 + 2 = 3,5$  GB, tak więc brakuje jeszcze  $0,563$  GB.

## 1.8. Genom ludzki

Ilość informacji, która odpowiada pojedynczej literze alfabetu, wynosi:  $I = \log_2 N = \log_2 4 = 2$ . W celu reprezentowania sekwencji DNA na komputerze musielibyśmy być w stanie reprezentować wszystkie 4 litery w postaci binarnej. Można to zrobić za pomocą minimum 2 bitów, co daje 4 różne kombinacje bitowe (00, 01, 10 i 11). Każde 2-bitowe połączenie stanowiłoby jedną sekwencję DNA. Jeden bajt (lub 8 bitów) może wtedy reprezentować 4 sekwencje DNA. Aby zapisać cały ludzki genom na komputerze bez kompresji, potrzebujemy około  $3 \cdot 10^9 : 4 = 750\,000\,000$  bajtów lub  $715,26$  megabajtów pamięci. Do zapisu wystarczy zatem np. pamięć USB o pojemności 1 GB.

## 1.9. Wyniki kartkówki

Grupa ma jednego starostę, starosta mógł zaliczyć kartkówkę (zdarzenie A) lub nie zaliczyć kartkówki (zdarzenie B). Zgodnie ze **wzorem Shannona** informacja dotycząca i-tego niezależnego zdarzenia „część grupy ze starosta na czele nie zaliczyła kartkówki” może być obliczona w sposób następujący:  $I_i = \log_2 P_i = 4$ , gdzie  $P_i$  to prawdopodobieństwo zdarzenia, które zaszło. Stąd  $P_i = 2^{(-I_i)} = 2^{(-4)} = 1 : 16$ . Nie mieliśmy informacji dodatkowej o wynikach kartkówki, tak więc możemy uważać, że prawdopodobieństwo to zależy tylko od liczby studentów, którzy nie zaliczyli kartkówki. Część grupy, która nie zaliczyła kartkówki, wynosi jedną szesnastą z 32, czyli jest to tylko dwóch studentów, wśród których znalazł się starosta grupy. Odpowiedź brzmi: kartkówkę zaliczyło trzydziestu studentów.

## 1.10. Tryb High Color

Głębia koloru przy podanej liczbie wyświetlanych kolorów wynosi:  $G_K = \log_2 65536 = 16$  bitów. Oznacza to, że w pamięci wideo sygnał dla każdego punktu (piksele) wyświetlanego obrazka kodowany jest dwoma bajtami. Liczba punktów obrazu, dla których niezależnie kodowany będzie kolor przy podanej rozdzielczości, wynosi:  $L_P = 1024 \cdot 768 = 786\,432$ . Stąd minimalna pamięć niezbędna do obsługi wyświetlacza w naszej sytuacji może być obliczona następująco:  $16 \cdot 786\,432 = 12\,582\,912 \text{ b} \approx 1,5 \text{ MB}$ .

Większa czułość oka na kolor zielony podpowiada, że do kanału zielonego należy przydzielić więcej pamięci wideo, aby umożliwić kodowanie większej liczby odcieni zieleni. Mając do podziału 16 bitów, można by wybrać schemat podziału 5-6-5 lub 4-8-4 (według schematu RGB: czerwony-zielony-niebieski). Przyjęty przemysłowy standard rozwiązania tego problemu (poparty badaniami) polega na tym, że kolory czerwony i niebieski będą miały kodowanie 5-bitowe, da to 32 możliwe poziomy jasności. A kolor zielony będzie kodowany za pomocą pozostałych 6 bitów, da to 64 możliwe poziomy jasności (więcej, ponieważ oko ludzkie rozróżnia więcej odcieni koloru zielonego). Zatem kolory czerwony i niebieski zajmą po 0,47 MB pamięci wideo, a na kolor zielony przypadnie 0,56 MB.

## 1.11. Bardzo duży plik wideo

Na dysku o systemie plików FAT16 można przechowywać pliki o rozmiarze do 2,1 GB.

Na dysku o systemie plików FAT32 można przechowywać pliki o rozmiarze do 4 GB.

Na dysku o systemie plików NTFS można przechowywać pliki o rozmiarze do 16 TB.

System plików APPS nie istnieje, więc ciężko w nim cokolwiek przechować.

Plik o wielkości 123 456 789 000 B to 114,978 GB, czyli prawie 115 GB. W związku z tym na poszczególnych partycjach danych plik musi zostać podzielony na następującą liczbę mniejszych plików, aby można było go zarchiwizować:

1. Dysk A — 55 plików.
2. Dysk B — 29 plików.
3. Dysk C — nie trzeba dzielić pliku.
4. Dysk D — taki system plików nie istnieje.

## 1.12. Jakość obrazków w gazetce studenckiej

Mamy do czynienia z kodowaniem informacji, wykorzystującym alfabet z 256 znaków (kolorów). W tej sytuacji informacja przekazywana za pomocą pojedynczego koloru wynosi:

$$i = \log_2 256 = 8 \text{ bitów.}$$

Powierzchnia obrazka wynosi 3 na 5 = 15 cali kwadratowych. Oznaczmy jako  $X$  rozdzielczość drukarki w pikselach na cal, wtedy ogólna liczba pikseli obrazka będzie wynosiła  $15X^2$ , a liczba bitów informacji dla całej bitmapy —  $8 \cdot 15X^2$ . Możemy przyrównać to wyrażenie do rozmiaru pliku, czyli:

$$120X^2 = 43200000.$$

Po rozwiązaniu równania kwadratowego dowiadujemy się, że szukana rozdzielczość wynosi 600 pikseli na cal, jest to bardzo wysoka rozdzielczość, stosowana przy wydrukach fotograficznych.

## 1.13. Zapisujemy fale dźwiękowe

Częstotliwość próbkowania wynosi 44,1 kHz. W związku z tym w ciągu jednej sekundy zostanie pobranych 44 100 próbek.

Każda próbka będzie kodowana ciągiem 24-bitowym. Dlatego natężenie strumienia danych pomiarowych będzie wynosiło 1 058 400 b/s, co jest równoważne 132 300 B/s. W związku z tym w ciągu 30 minut zostanie zapisanych na dysku  $30 \cdot 60 \cdot 132 300$  bajtów informacji. Dlatego całkowita pamięć potrzebna do archiwizacji tej próbki dźwiękowej wyniesie około 227 MB (dokładniej 227,108 MB).



## 1.14. Co to jest digitalizacja?

Przedział zmienności wartości analogowej jest następujący:  $x \in \langle -5; 8 \rangle$ . W związku z tym można zapisać, że  $x_{\min} = -5$ ,  $x_{\max} = 8$ . Najmniejsza liczba bitów  $LB$  kodująca wartości z tego przedziału w ciągu binarnym musi spełniać zależność:  $2^{LB} \geq \frac{x_{\max} - x_{\min}}{\beta} + 1$ ,

$$\text{czyli: } 2^{LB} \geq \frac{8+5}{0,01} + 1.$$

Wynika stąd, że:  $2^{LB} \geq 1301$ .

W związku z tym najmniejsza liczba bitów  $LB$  spełniająca powyższą nierówność wynosi 11, tak więc kodowanie ośmiobitowe będzie tutaj niewystarczające, a kodowanie szesnastobitowe będzie nadmiarowe. Optymalne w tej sytuacji okaże się kodowanie jedenastobitowe, a wśród popularnych standardów kodowania najlepsze będzie dwunastobitowe (półtorabajtowe).

## 1.15. Biblioteka na dysku komputera

Liczba znaków znajdujących się na jednej stronie wynosi:  $30 \cdot 70 = 2100$ . Każda książka średnio składa się z 500 stron, więc liczba znaków w książce wynosi:  $2100 \cdot 500 = 1\,050\,000$ .

W kodzie ASCII jeden znak zapisywany jest przy użyciu 8 bitów (1 bajt).

W kodzie UTF, w zależności od wariantu, jeden znak jest zapisywany przy użyciu 8 bitów (1 bajt — kod UTF-8) lub 16 bitów (2 bajty — kod UTF-16).

Zatem aby zapisać na dysku jedną książkę, należy wykorzystać:

$$1\,050\,000 \cdot 1 = 1\,050\,000 \text{ B (kod ASCII oraz UTF-8),}$$

$$1\,050\,000 \cdot 2 = 2\,100\,000 \text{ B (kod UTF-16).}$$

Pamiętając, że 160 GB to 171 798 691 840 B, dzielimy pojemność dysku przez wielkość pamięci wymaganą do zapisania jednej książki. Po przeprowadzeniu obliczeń widzimy, że na dysku o dostępnej pojemności użytkowej 160 GB można zapisać:

$$163\,617 \text{ książek (kod ASCII oraz UTF-8),}$$

$$81\,808 \text{ książek (kod UTF-16).}$$

Jak widać, dysk jest za mały dla księgozbioru zawartego w bibliotece, ale dwa takie dyski mogą być wystarczające.

## 1.16. Skanowane zdjęcie

Kolorowe zdjęcie o wymiarach 10 na 15 cm w przeliczeniu na cale ma rozmiar 3,937 cali na 5,906 cali. Jeśli chcemy je zeskanować w rozdzielczości 300 ppi (dpi) (300 punktów na 1 cal), wówczas rozmiar zdjęcia, liczony w pikselach, będzie wynosił 1181 na 1771. W trybie **RGB** występują 3 składowe koloru (**R**ed; **G**reen; **B**lue), a w trybie **CMYK** występują 4 składowe koloru (**C**yan; **M**agenta; **Y**ellow; **blacK**).

Jeśli przyjmujemy, że pojedyncza składowa koloru jest zakodowana przez 8 bitów, wówczas 1 piksel będzie zapisany przy użyciu 24 bitów (RGB) lub 32 bitów (CMYK).

Zatem rozmiar pliku będzie równy:  $1181 \cdot 1771 \cdot 24$  bity (RGB) lub  $1181 \cdot 1771 \cdot 32$  bity (CMYK). Czyli około 5,98 MB (RGB) lub 7,98 MB (CMYK).

## 1.17. Obliczamy kolor

Poszukiwany kolor będzie odpowiadał minimalnej odległości do wzorca w przestrzeni kolorów, czyli minimum funkcji:

$$f_i = \text{SQRT}(30 \cdot (R_i - R_w)^2 + 59 \cdot (G_i - G_w)^2 + 11 \cdot (B_i - B_w)^2).$$

Najprostszym rozwiązaniem będzie obliczenie odległości do wzorca dla wszystkich  $N$  kolorów i następnie wybranie takiego z nich, któremu będzie odpowiadała najmniejsza wartość  $f_i$ .

## Podsumowanie

Jednym z bardzo istotnych obszarów aktywności informatyków zawodowych jest prowadzenie i obsługa olbrzymich zbiorów danych. Dane te opisują przebieg procesów biznesowych istotnych dla gospodarki opartej na wiedzy. Zbiory danych są prowadzone w postaci baz danych, które podporządkowują się ściśle określonym regułom. Systemy informatyczne zbudowane na podstawie tych baz danych udostępniają klientom zamawiane informacje. Mogłeś słyszeć nawet takie określenie jak **hurtownia danych** (ang. *data warehouse*), jest to rodzaj bazy danych zoptymalizowanej pod kątem pewnego biznesu. Stosuje się również określenie **baza wiedzy** (od ang. *knowledge base*), jest to rozległy zbiór powiązanych logicznie danych dotyczących danego obszaru tematycznego i umożliwiających zautomatyzowane wnioskowanie logiczne. Trudno wyobrazić sobie — i tym bardziej całościowo ogarnąć — zbiory danych prowadzone przez takich gigantów informacyjnych jak Google, Yahoo czy Bing. Pewnie i Ty korzystasz z usług takich systemów informatycznych dostępnych w Internecie.

Już przed końcem poprzedniego stulecia wyklarowała się odrębna dziedzina działalności informatyków zwana **zarządzaniem informacją** (ang. *Information management*). Obejmuje ona gromadzenie i zarządzanie informacjami z wielu źródeł i dostarczanie tych informacji do odbior-

ców. Systemy zarządzania informacją wiążą tych, którzy mają udział w powstaniu informacji lub prawo do wykorzystania tych informacji. Zarządzanie informacjami nie jest zadaniem prostym, wymaga ono rozumienia teorii i technologii gromadzenia, przechowywania, dystrybuowania i ochrony informacji. Działania te są ściśle związane i pokrywają się z praktyką zarządzania dużymi zbiorami danych.

Jeszcze jednym ciekawym obszarem pracy informatyków jest tak zwane **przetwarzanie informacji** (od ang. *Information processing*). Ta działalność polega na zastosowaniu metod transformacji dostępnych informacji zgodnie z określonymi potrzebami informacyjnymi kierownictwa organizacji. Typowe procedury transformacji informacji przewidują m.in. agregację informacji, sortowanie, wyszukiwanie i selekcję, statystyczną obróbkę, rangowanie, interpretację. Przetwarzanie informacji jest ściśle związane z działalnością menedżerów, tak więc informatycy mają za zadanie wspomagać ich działania. Przetwarzanie informacji umożliwia efektywne zarządzanie, jeśli udostępni odpowiednie (właściwe, użyteczne, cenne) informacje dotyczące przedsiębiorstwa i jego kontrahentów.

Wymiarowanie informacji ma znaczenie praktyczne w kontekście ciągle rozwijającego się biznesu informacyjnego. Przecież już płacimy za megabajty transferu danych w sieciach komórkowych, w niektórych serwisach jesteśmy rozliczani za ilość pobieranych danych, musimy realnie oceniać swoje możliwości i ograniczenia w tym zakresie. Tak samo powinniśmy umieć realnie oceniać przestrzeń informacyjną posiadanych nośników informacji, objętość kopiowanych lub wysyłanych multimediiów (plików audio, wideo lub zdjęć elektronicznych). Mamy nadzieję, że rozwiązane w tym rozdziale zadania pomogą Ci w doskonaleniu umiejętności oraz w głębszym zrozumieniu zachodzących zjawisk natury informacyjnej.

Przedstawione w rozdziale zadania oczywiście nie wyczerpują zagadnień teorii informacji, a raczej pokazują niektóre proste jej zastosowania. Studiując informatykę, będziesz mógł zapoznać się z wieloma ciekawymi zagadnieniami związanymi z pojęciem informacji i z jej wymiarowaniem. Zagadnienia zaawansowane, takie jak architektura informacji, frameworki informacyjne, bazy wiedzy czy inżynieria informacji, poznasz w trakcie studiów. Obecnie możesz skorzystać z dodatkowych źródeł wiedzy podanych we wstępie do niniejszego rozdziału oraz zapoznać się z działaniem polecanych przez nas serwisów internetowych:

1. Kalkulatory bitowe: <http://www.matisse.net/bitcalc/>, <http://www.bit-calculator.com/>.
2. Obliczeniowa maszyna wiedzy WolframAlpha: <http://www.wolframalpha.com/>.
3. Narzędzie do informacyjnej analizy tekstów Textalyser: <http://textalyser.net/>.
4. Narzędzia typu open source do zarządzania wiedzą: <http://sourceforge.net/directory/business-enterprise/knowledgemanagement/>.
5. „Information is beautiful” — serwis internetowy prezentujący wizualizację pomysłów, problemów, wiedzy, danych: <http://www.informationisbeautiful.net/>.



# PROGRAM PARTNERSKI

GRUPY WYDAWNICZEJ HELION



- 1. ZAREJESTRUJ SIĘ**
- 2. PREZENTUJ KSIĄZKI**
- 3. ZBIERAJ PROWIZJĘ**

Zmień swoją stronę WWW  
w działający bankomat!

**Dowiedz się więcej i dołącz już dzisiaj!**

<http://program-partnerski.helion.pl>

GRUPA WYDAWNICZA

 **Helion SA**

Informatyka to niezwykle dynamicznie rozwijająca się dziedzina wiedzy. Komputery otaczają nas ze wszystkich stron: sterują pracą sprzętów gospodarstwa domowego, pośredniczą w komunikacji, a nawet umożliwiają eksplorację przestrzeni kosmicznej.

Dzisiejszy świat trudno już właściwie wyobrazić sobie bez maszyn cyfrowych i kontrolującego je oprogramowania. Wciąż rośnie zapotrzebowanie rynku na wysokiej klasy specjalistów, którzy dzięki swojej wiedzy i dużemu doświadczeniu potrafią projektować, budować oraz programować systemy komputerowe. Popularność zawodu informatyka potęgują też wysokie zarobki, które zwiększają zainteresowanie młodzieży tą nietatwą dziedziną wiedzy.

Jeśli interesujesz się informatyką, chcesz lepiej poznać jej podstawowe zagadnienia, nauczyć się rozwiązywać jej klasyczne problemy i analizować zadania informatyczne w sposób właściwy prawdziwym programistom, sięgnij po książkę *Chcę zostać informatykiem*. W przystępny sposób prezentuje ona matematyczne mechanizmy wykorzystywane w informatyce, uczy podstaw algorytmiki i wprowadza w świat programowania komputerów. Przedstawia zagadnienia związane z sieciami komputerowymi i bezpieczeństwem informacji oraz sposoby tworzenia gier komputerowych. Przede wszystkim jednak zawiera mnóstwo praktycznych przykładów i ciekawych zadań, które pomogą zainteresowanym kandydatom sprawdzić poziom swojej wiedzy i dostać się na studia informatyczne.

Jeśli marzysz o karierze informatyka lub po prostu interesujesz się programowaniem i chcesz poszerzyć swoją wiedzę na ten temat, trafieś na właściwą książkę!

- PODSTAWY TEORETYCZNE INFORMATYKI NA SKRÓTY
- CIEKAWY ZADANIA INFORMATYCZNE WRAZ Z ROZWIĄZANAMI
- MATEMATYKA, LOGIKA I ALGORYTMIKA W INFORMATYCE
- ANALIZA PODSTAWOWYCH PROBLEMÓW PROGRAMISTYCZNYCH
- WYMIAROWANIE INFORMACJI I BEZPIECZEŃSTWO DANYCH
- PROGRAMOWANIE KOMPUTERÓW I SIECI KOMPUTEROWE
- GRY KOMPUTEROWE

**NIE WAHAJ SIĘ! ZRÓB PIERWSZY KROK,  
ABY POZNAĆ SEKRETY INFORMATYKI!**

**helion.pl**  
księgarnia  
internetowa

Nr katalogowy: 18228

Księgarnia internetowa:  
<http://helion.pl>

Zamówienia telefoniczne:  
**0 801 339900**  
**0 601 339900**



**Helion**

Sprawdź najnowsze promocje:  
• <http://helion.pl/promocje>  
Książki najchętniej czytane:  
• <http://helion.pl/bestsellery>  
Zamów informacje o nowościach:  
• <http://helion.pl/nowosci>

Helion SA  
ul. Kościuszki 1c, 44-100 Gliwice  
tel.: 32 230 98 63  
e-mail: [helion@helion.pl](mailto:helion@helion.pl)  
<http://helion.pl>

sięgnij po WIĘCEJ



KOD KORZYŚCI

ISBN: 978-83-246-8525-7



Cena: 29,90 zł

**Informatyka w najlepszym wydaniu**