

Spis treści

OD AUTORÓW	11
1. WPROWADZENIE	13
1.1. Czym jest komputer i do czego służy?	13
1.2. Jakie komputery i kiedy zapoczątkowały rozwój informatyki na świecie?	13
1.3. Który komputer można uznać za pierwszy?	14
1.4. Co to jest koncepcja von Neumanna (komputer von neumannowski)?	15
1.5. Jak działa komputer?	15
1.6. Z czego buduje się komputery?	16
1.7. Jak klasyfikuje się komputery?	17
1.8. Co to jest komputer wirtualny?	18
1.9. Co znaczy określenie „generacja komputerów”?	18
1.10. Jakie znamy generacje komputerów?	19
1.11. Jakie firmy zajmują się produkcją komputerów?	19
2. BUDOWA KOMPUTERA I ZASADY JEGO DZIAŁANIA	21
2.1. W jaki sposób przedstawiane są informacje wewnątrz komputera?	21
2.2. Jak są reprezentowane wewnątrz komputera informacje, które nie są liczbami?	22
2.3. Czy przy określaniu informacji przechowywanych wewnątrz komputera używa się wyłącznie pozycyjnego systemu dwójkowego?	23
2.4. Co zdecydowało o takim, a nie innym sposobie przedstawiania informacji wewnątrz komputera?	24
2.5. Z jakich bloków funkcjonalnych buduje się komputery?	25
2.6. Jak jest zbudowany komputer?	26
2.7. Jakie podstawowe parametry charakteryzują procesor?	27
2.8. Jakie podstawowe parametry charakteryzują pamięć komputera?	28
2.9. Jakie podstawowe parametry charakteryzują urządzenia wejścia/wyjścia?	29
2.10. W jaki sposób procesor współpracuje z pamięcią operacyjną?	29
2.11. W jaki sposób komputer współpracuje z urządzeniami wejścia/wyjścia?	30
2.12. W jaki sposób procesor komunikuje się z urządzeniami wejścia/wyjścia?	30
2.13. Jak są zasilane obecne komputery?	31
2.14. Co to jest UPS i jaką rolę pełni w zasilaniu komputera?	31
3. PROCESOR	33
3.1. Jak jest zbudowany procesor?	33
3.2. Jak działa procesor?	34
3.3. Co to jest adresowanie i jakie znamy najważniejsze rodzaje adresowania?	35
3.4. Jaką postać mają rozkazy komputera?	36
3.5. Jak klasyfikujemy rozkazy komputera?	36
3.6. Jak są przetwarzane dane w procesorze potokowym, skalarnym?	37

3.7.	Do czego służą i jak są wykonywane rozkazy skoku?	37
3.8.	Na czym polegają rozgałęzienia i w jaki sposób są realizowane?	38
3.9.	Co to są pętle?	38
3.10.	Co to są i do czego służą znaczniki?	38
3.11.	Co to jest podprogram i w jaki sposób się z niego korzysta?	39
3.12.	Co to jest stos?.....	41
3.13.	Do czego służy stos?	41
3.14.	Na czym polega przerwanie i jakie są rodzaje przerwania?.....	42
3.15.	Jak przebiega obsługa przerwania?	42
3.16.	Jak działa arytmometr?.....	43
3.17.	Jakie rodzaje liczb można przedstawić w komputerze?	43
3.18.	Na czym polega zapis w systemie znak-moduł?	44
3.19.	Co to jest uzupełnienie dwójkowe?.....	44
3.20.	Na czym polega zapis stałopozycyjny i do czego służy?	45
3.21.	Jaką postać mają liczby zmiennopozycyjne i dlaczego się je stosuje?.....	47
3.22.	Co to jest koprocessor?	47
3.23.	Jak działa sterowanie?	48
3.24.	Od czego zależy szybkość działania procesora?	49
3.25.	Do jakiej wartości można zwiększać częstotliwość taktowania procesora?	49
3.26.	Co to jest mikroprocesor i kiedy go stworzono?.....	50
3.27.	W jakich okolicznościach doszło do skonstruowania mikroprocesora?	50
3.28.	Co powoduje dużą popularność mikroprocesorów?.....	51
3.29.	Co to są procesory CISC, RISC?.....	51
3.30.	Dlaczego wprowadzono procesory RISC?.....	52
3.31.	Gdzie w praktyce spotykamy procesory o architekturze RISC?	52
3.32.	Jakie procesory stosuje się do budowy współczesnych komputerów PC?.....	53
4.	PAMIĘĆ	55
4.1.	Po co jest pamięć i co się przechowuje w pamięci?.....	55
4.2.	Jak jest zbudowana pamięć RAM?.....	55
4.3.	Jakie nośniki informacji stosuje się w pamięciach komputerów?.....	56
4.4.	Co to znaczy, że pamięć jest ulotna?	57
4.5.	Czy informacje zapisane w pamięci można przechowywać dowolnie długo?.....	57
4.6.	Czy zawartość pamięci można zmieniać?	57
4.7.	Jak klasyfikuje się pamięci komputerów?	58
4.8.	Jak jest zbudowana pamięć półprzewodnikowa?	59
4.9.	Czym się różni pamięć półprzewodnikowa dynamiczna od statycznej?.....	60
4.10.	Co to jest pamięć podręczna cache i dlaczego się ją stosuje?.....	61
4.11.	Co to jest i dlaczego stosujemy pamięć wirtualną?	63
4.12.	Co to jest ramdysk i po się go stosuje?.....	63
4.13.	Jak są zbudowane pamięci stałe (ROM)?.....	64
4.14.	Do czego służą pamięci zewnętrzne (masowe) i jakie są jej rodzaje?	65

4.15. Na czym polega zasada działania pamięci z ruchomym nośnikiem magnetycznym?	65
4.16. Jakie metody kodowania informacji stosuje się w pamięciach z ruchomym nośnikiem magnetycznym?	66
4.17. Jak działają dyski twarde (typu Winchester)?	67
4.18. Jak działa napęd dysków elastycznych?	67
4.19. Co to są dyski wymienne?	68
4.20. Na jakiej zasadzie działają pamięci magnetoptyczne?	69
4.21. Na czym polega zasada działania pamięci z ruchomym nośnikiem optycznym CD, CD-R, CD-RW?	69
4.22. W jaki sposób koduje się informacje w pamięciach z nośnikiem optycznym?	70
4.23. Co to jest DVD?	71
4.24. Jak jest zbudowany streamer?	71
4.25. Co to jest backup? W jaki sposób archiwizuje się informacje i dlaczego?	72
4.26. Co to jest mirroring?	73
4.27. Na czym polega kompresja archiwizowanych informacji?	74
4.28. Jakie środki ostrożności należy zachowywać przy przechowywaniu nośników magnetycznych?	74
5. SYSTEM WEJŚCIA/WYJŚCIA.....	75
5.1. Jakie znamy urządzenia wejścia/wyjścia?	75
5.2. Dlaczego w nazewnictwie urządzeń wejścia/wyjścia jest tak dużo nazw angielskich?	75
5.3. Jakie parametry techniczne i cechy funkcjonalne urządzeń wejścia/wyjścia są istotne z punktu widzenia komputera?	76
5.4. Co to jest system wejścia/wyjścia?	76
5.5. W jaki sposób przedstawia się informacje przesyłane do i z urządzeń wejścia/wyjścia?	77
5.6. Jak w komputerze przedstawia się znaki narodowe?	77
5.7. W jaki sposób sprzęga się urządzenia wejścia/wyjścia z komputerem?	78
5.8. Jakie funkcje pełni BIOS w systemie komputerowym?	79
5.9. W jaki sposób zmieniamy konfigurację BIOS-u?	79
5.10. Co to jest interfejs?	79
5.11. Co to jest Centronix?	80
5.12. Czym różnią się tryby pracy portu równoległego SPP, BPP, EPP, ECP?	80
5.13. Co to jest RS-232?	81
5.14. Jak procesor obsługuje urządzenia wejścia/wyjścia?	81
5.15. Dlaczego stosuje się przerwania?	82
5.16. Jak klasyfikujemy przyczyny przerwania?	82
5.17. Jakie wyróżniamy etapy obsługi przerwania?	83
5.18. Jak działa monitor graficzny wyposażony w kineskop?	83
5.19. Jakimi parametrami charakteryzuje się monitor graficzny?	84
5.20. Jakie wyróżniamy rodzaje drukarek?	85
5.21. Jakie parametry charakteryzują drukarkę?	85

5.22. Jak działa drukarka igłowa?	86
5.23. Na czym polega zasada działania drukarki atramentowej?	86
5.24. Jaka jest zasada działania drukarki laserowej?	86
5.25. Jak działa drukarka termiczna?	87
5.26. Czym różni się ploter od drukarki?	87
5.27. Jak działa skaner?	87
5.28. Co to są digitizery?	88
5.29. Czy do komputera można podłączyć kamerę cyfrową?	88
5.30. Czy komputer może wytwarzać sygnały dźwiękowe i czy może rejestrować dźwięki?	89
5.31. Czym różnią się produkowane obecnie karty dźwiękowe?	89
5.32. Czy można wydawać komputerowi polecenia głosowe?	89
5.33. Czy komputer analizuje obrazy?	90
5.34. Czy możliwe jest rozpoznawanie pisma?	90
5.35. W jaki sposób niewidomi odbierają wyniki pracy komputera?	90
5.36. Jakie znamy tryby transmisji danych?	91
5.37. Jakimi parametrami określa się łącze transmisji danych?	91
5.38. Co to jest modem i jak się porozumiewa z komputerem?	92
5.39. Jakie środki techniczne są niezbędne, aby można korzystać z modemu?	93
5.40. Dlaczego do połączeń na duże odległości używa się modemów?	93
5.41. W jaki sposób modemy różnych producentów współpracują ze sobą?	94
5.42. Jakie stosuje się metody zabezpieczeń przed zniekształceniem przesyłanej informacji?	95
5.43. Jakie stosuje się metody kompresji danych w modemach?	95
5.44. Co podłączamy do złącza PS2?	96
5.45. Jak działa i jak się komunikuje z komputerem mysz?	96
5.46. Czym można zastąpić mysz?	96
5.47. Dlaczego monitory LCD stają się coraz bardziej popularne?	97
5.48. W jakim celu stosujemy ekrany dotykowe?	97
5.49. Jakie zalety mają kody kreskowe?	98
5.50. Jak są przesyłane dane z klawiatury do komputera?	98
5.51. Do czego służy i czym się charakteryzuje złącze PCMCIA?	99
5.52. Co to jest USB?	99
5.53. Jakie wyróżniamy standardy magistral i złączy kart rozszerzeń?	99
5.54. Jakie zadania spełnia sterownik sprzętowy dysku?	100
5.55. Co należy rozumieć pod pojęciem chipset?	100
5.56. Dlaczego magistrala PCI jest lepsza od ISA?	101
5.57. Czym różnią się magistrale IDE i SCSI?	101
5.58. Do czego służy port AGP?	102
5.59. Jaką rolę pełnią w systemie komputerowym sterowniki programowe?	102

6. ARCHITEKTURA SYSTEMÓW KOMPUTEROWYCH.....	103
6.1. Co spowodowało, że zamiast określenia „komputer” zaczęto używać terminu „system komputerowy”?.....	103
6.2. Co należy rozumieć przez określenie „architektura” w odniesieniu do systemu komputerowego?	103
6.3. Jak klasyfikujemy systemy komputerowe?	104
6.4. Co to jest skalowalność?.....	105
6.5. Co oznacza pojęcie: wielowątkowość?	105
6.6. Na czym polega wieloprogramowość (wielozadaniowość)?.....	105
6.7. Jakie powinny być racjonalne reguły przydzielania priorytetów programom użytkowym w systemie wielozadaniowym?	106
6.8. Czy jest możliwe, aby w systemie wielozadaniowym dwa programy zostały wykonywane w takim samym czasie, jaki jest potrzebny do wykonania każdego z nich w systemie jednozadaniowym?.....	107
6.9. Jak działa system wielodostępny?	107
6.10. Co to są systemy komputerowe czasu rzeczywistego?.....	107
6.11. Co spowodowało wprowadzenie systemów wieloprocesorowych?	108
6.12. Co to są systemy wieloprocesorowe i jakie znamy ich struktury?	108
6.13. Jakie są wady systemów wieloprocesorowych?	109
6.14. Co to są komputery sterowane danymi ?	110
6.15. Do obsługi jakich architektur przystosowany jest system Windows 95, Windows NT, a do jakich UNIX?	110
6.16. Jakie przyczyny spowodowały powstanie sieci komputerowych?	111
7. SIECI KOMPUTEROWE – PODSTAWY	113
7.1. Co to jest sieć komputerowa i jakie są cele jej użytkowania?.....	113
7.2. Co to jest model warstwowy ISO-OSI?	113
7.3. Jakie zadania spełniają poszczególne warstwy modelu ISO-OSI?.....	114
7.4. Dlaczego stosujemy model warstwowy?.....	115
7.5. Jakie zasoby można udostępnić w sieci komputerowej?.....	116
7.6. Co to jest segment sieci?	116
7.7. Jakie wyróżnia się rodzaje sieci komputerowych?.....	116
7.8. Jakie są topologie lokalnych sieci komputerowych?.....	117
7.9. Co to jest protokół sieciowy (komunikacyjny)?.....	120
7.10. Co to jest Ethernet ?.....	121
7.11. Jakie wyróżniamy rodzaje sieci Ethernet?.....	121
7.12. Jakie rodzaje okablowania stosowane są przy budowie sieci Ethernet?	121
7.13. Jakie urządzenia stosuje się do budowy i łączenia sieci LAN?.....	122
7.14. Jakie protokoły sieciowe są najczęściej używane w sieciach LAN?.....	122
7.15. Jakie media używane są do połączeń sieciowych?.....	123
7.16. Jakie są kategorie okablowania?.....	123
7.17. Co rozumiemy pod pojęciami inteligentny budynek, okablowanie strukturalne?	124
7.18. Jakie wyróżnia się prędkości przesyłania informacji w sieciach komputerowych?.....	125

7.19. Czy informacje przesyłane w sieci są zabezpieczane przed błędami?	125
7.20. Czy można w prosty sposób połączyć siecią Ethernet dwa komputery bez dodatkowych urządzeń?	125
7.21. Jaka jest różnica między routerem a bramą?	126
7.22. Czym różni się wzmacniacz od przełącznika?	126
7.23. Co to jest serwer?	127
7.24. Co to jest EDI?	127
8. INTERNET.....	129
8.1. Jak powstał Internet?	129
8.2. Co to jest Internet i jak działa?	129
8.3. Kto administruje Internetem?	130
8.4. Co to jest etykieta sieciowa?	131
8.5. W jaki sposób można uzyskać dostęp do Internetu?	131
8.6. Jakie podstawowe protokoły transmisji dostępne są w Internecie i jak działają? ..	132
8.7. Jak są rozróżniane poszczególne komputery w Internecie?	133
8.8. Według jakiego klucza są nadawane nazwy i adresy komputerom pracującym w ramach Internetu?	133
8.9. Jak jest odniesienie modelu OSI do protokołu TCP/IP?	134
8.10. Dlaczego powstał protokół IPv6?	134
8.11. Jak sprawdzić działanie połączenia z Internetem?	135
8.12. Jakie podstawowe serwisy są dostępne w Internecie?	135
8.13. Co to jest DNS i jak działa?	136
8.14. Co to jest serwis WWW i jak on działa?	136
8.15. W jaki sposób można przysyłać pliki w Internecie?	137
8.16. Czy można pracować na zdalnym komputerze podłączonym do Internetu? Jeśli tak, to jak to zrobić?	137
8.17. Jak działa poczta w Internecie?	138
8.18. Do czego służy serwis news?	138
8.19. Jak można prowadzić pogawędki w Internecie?	139
8.20. Czy można słuchać radia lub oglądać telewizję przez Internet?	140
8.21. Czy możliwe jest połączenie wideokonferencyjne z innym użytkownikiem Internetu?	140
8.22. Jak wysłać faks przez Internet?	140
8.23. Czy można korzystać z Internetu przez telefon komórkowy?	141
8.24. Czy połączenia w Internecie są bezpieczne?	141
8.25. Czy można robić zakupy i operacje bankowe przez Internet?	142
8.26. Co to jest SSL?	142
9. OPROGRAMOWANIE SYSTEMÓW KOMPUTEROWYCH.....	143
9.1. Co to jest algorytm?	143
9.2. Jaki jest źródłowy termin algorytm?	143
9.3. Jakie są typowe konstrukcje algorytmiczne?	144

9.4. Jaką postać będzie miał przykładowy algorytm dla zadania poszukiwania maksymalnego elementu z ciągu danych n liczb?	144
9.5. Co rozumiemy pod pojęciem język programowania, program i programowanie?	145
9.6. Jak klasyfikujemy języki programowania?	146
9.7. Dlaczego programowanie w języku wewnętrznym jest uciążliwe?	147
9.8. Czym się różni język assemblerowy od języka wewnętrznego?	147
9.9. Jaka jest postać instrukcji w języku assemblerowym?	148
9.10. Jakie są wady języków assemblerowych?	149
9.11. Dlaczego języki assemblerowe są nadal stosowane mimo ich wad?	150
9.12. Dlaczego języki wysokiego poziomu stanowią duże ułatwienie w pracy programisty?	150
9.13. Jakie są najważniejsze języki wysokiego poziomu?	151
9.14. Co to jest programowanie strukturalne?	151
9.15. Jaką postać w języku Pascal miałyby przykładowy program znajdowania maksymalnego elementu z ciągu danych n liczb a_1, a_2, \dots, a_n ?	152
9.16. Co to jest programowanie obiektowe?	153
9.17. Jaką postać w języku C++ miałyby przykładowy program znajdowania maksymalnego elementu ze stu losowych liczb?	153
9.18. Jakie są kryteria oceny języków programowania?	155
9.19. Jakie są zasadnicze różnice między językami naturalnymi a językami programowania?	155
9.20. Co to są metody numeryczne?	155
9.21. Czy komputer popełnia błędy arytmetyczne?	156
9.22. Jakimi wyróżniamy etapy opracowania programu?	156
9.23. Jakie cechy powinien mieć dobry program?	157
9.24. Jaka jest wydajność programistów?	158
9.25. Jakie znamy sposoby testowania programów?	159
9.26. Co to jest oprogramowanie systemu komputerowego i jak je klasyfikujemy? ..	160
9.27. Co to jest system operacyjny i do czego służy?	161
9.28. Jak doszło do powstania systemów operacyjnych?	161
9.29. Jaka jest różnica między procesem a programem?	163
9.30. Na czym polega działanie systemu operacyjnego?	163
9.31. Jaka jest typowa struktura systemu operacyjnego?	163
9.32. W jaki sposób system operacyjny zarządza czasem procesora?	165
9.33. W jaki sposób system operacyjny zarządza pamięcią operacyjną?	167
9.34. W jaki sposób system operacyjny zarządza informacją?	167
9.35. W jaki sposób system operacyjny zarządza urządzeniami i operacjami wejścia/wyjścia?	169
9.36. W jaki sposób przebiega komunikacja operatora komputera z systemem operacyjnym?	170
9.37. Co to są translatory i jaka jest różnica między kompilatorami a interpreterami? ..	170
9.38. Co to jest odwrotna notacja polska (ONP)?	172
9.39. Dlaczego stosujemy narzędzia wspomagające tworzenie programów?	174

9.40. Co rozumiemy pod pojęciem CASE?.....	174
10. EWOLUCJA KOMPUTERÓW I ICH ZASTOSOWANIA	175
10.1. W jakich dziedzinach stosujemy komputery?	175
10.2. Co kryje się pod pojęciem multimedia?	176
10.3. Jakie są tendencje rozwoju technik multimedialnych?	176
10.4. Co to jest sztuczna inteligencja?.....	176
10.5. Co to są systemy eksperckie?	177
10.6. Czy współczesne komputery mogą rozpoznawać mowę?.....	177
10.7. Jak działają urządzenia analizujące obrazy? Jakie mają zastosowania?	178
10.8. Co to są sieci neuronowe?	179
10.9. Co to są komputery optyczne?	180
10.10. Jaki wpływ mają komputery na społeczeństwo?	180
10.11. Na czym polega wizja społeczeństwa informacyjnego?	181
10.12. Jakie cechy mają komputery piątej generacji?	182
11. SPIS LITERATURY	183
12. SKOROWIDZ.....	187

Od autorów

Szanowny Czytelniku

Jeśli właśnie sięgnąłeś na półkę księgarską po naszą książkę, to chcemy wyjaśnić czym ona nie jest. Otóż **nie jest ona instrukcją obsługi kolejnego programu użytkowego bądź systemu operacyjnego**. Podręczników uczących szybko (?) i krok po kroku, jakie klawisze naciskać oraz gdzie kliknąć myszą, aby osiągnąć żądany efekt w komputerze i zobaczyć to na ekranie monitora, jest bardzo dużo. Oczywiście nie deprecjonujemy ich wartości. Są one potrzebne do nabycia niektórych sprawności manualnych.

Celem, jaki sobie stawiamy, jest dostarczenie Czytelnikowi fundamentalnej wiedzy o komputerach, czyli solidnych i zrozumiałych podstaw. W naszym rozumieniu tylko zrozumienie fundamentów pozwala „nie bać się komputera” i adaptować się do szybko zachodzących zmian w tej dziedzinie. Chcemy dostarczyć wiedzy, która szybko się nie starzeje i oprócz odpowiedzi na pytania „Jak coś zrobić?” udziela odpowiedzi przede wszystkim na pytania „Dlaczego tak a nie inaczej trzeba coś zrobić?”

Spośród wielu form prezentacji wiedzy zdecydowaliśmy się na stworzenie zbioru pytań i odpowiedzi pogrupowanych tematycznie w rozdziały. Podział tematyczny omawianych zagadnień odróżnia naszą książkę od słownika (leksykonu) terminów informatycznych, w którym znajdujemy jedynie wyjaśnienia ułożonych alfabetycznie haseł bez ich hierarchicznego uporządkowania.

W naszych intencjach kierujemy tę pozycję do bardzo szerokiego grona uczniów, studentów, absolwentów uczelni mających różne – najczęściej małe – doświadczenie w pracy z komputerami, którzy zaledwie otarli się o środowisko komputerowe. Większość z nich wykorzystuje konkretne programy użytkowe, nie znając podstaw, i oczekuje książki – elementarza. Nasze **opracowanie ma być takim elementarzem i powinno umożliwić lepsze zrozumienie tego, co być może dotychczas wykonywane było całkiem nieźle, lecz w sposób słabo uświadomiony**. Jednocześnie pozycja ta powinna ułatwić świadome studiowanie innych, bardziej zaawansowanych, książek dostępnych na rynku wydawniczym.

Będziemy bardzo wdzięczni Czytelnikom za przekazanie wszelkich uwag na temat tej książki (e-mail: molmar@atr.bydgoszcz.pl)

Bydgoszcz 31.12.2000

Sopot 1.01.2001

1. Wprowadzenie

1.1. Czym jest komputer i do czego służy?

Komputer jest urządzeniem elektronicznym przeznaczonym do przetwarzania informacji według określonego algorytmu (patrz pyt. 9.1). Ważną cechą komputera jest możliwość zmiany w czasie realizowanych przez niego funkcji poprzez zmianę algorytmu.

Zakres zastosowań komputerów jest bardzo szeroki. Zmieniał się on wraz z rozwojem techniki. Najstarsze z komputerów służyły głównie do obliczeń numerycznych lub dla celów wojskowych. Wraz z rozwojem techniki, wzrostem niezawodności, zmniejszaniem rozmiarów i spadkiem cen wzrastała liczba zastosowań komputerów. Korzyści wynikające z używania komputerów w dużych przedsiębiorstwach i organizacjach handlowych spowodowały szybki wzrost zainteresowania tymi urządzeniami. Rosnący popyt wymuszał kolejne udoskonalenia, powodował wzrost produkcji i dalszą obniżkę cen. Miniaturyzacja i wzrost niezawodności pozwoliły zastosować komputery do sterowania różnymi obiektami i procesami technologicznymi. Gdy ludzie zaczęli łączyć ze sobą komputery, powstało nowe zastosowanie. Komputery ułatwiają porozumiewanie się ludzi (społeczeństwo informacyjne) oraz przekazywanie danych na odległość. Powstanie dużych sieci umożliwiło zastosowanie komputerów w wirtualnych sklepach, bankach, biurach itd.

Obecnie trudno jest wymienić dziedzinę, w której nie można by używać komputera.

1.2. Jakie komputery i kiedy zapoczątkowały rozwój informatyki na świecie?

W roku 1937 w Harvard University Howard Aiken rozpoczął pracę nad tzw. automatyczną maszyną cyfrową. Dziś termin maszyna cyfrowa wydaje się anachroniczny. Maszyna przecież przetwarza energię, a komputer przetwarza informacje. Konstrukcja była w użyciu od 1944 roku znana pod nazwą Mark I. Ta elektromechaniczna maszyna cyfrowa zbudowana była z kółek licznikowych sterowanych przekaźnikami i nie posiadała pamięci. Informacje wprowadzano do niej za pomocą taśmy dziurkowanej. Program był modyfikowany poprzez zmianę sieci połączeń kablowych. Mark I został zbudowany z około 750 tys. części połączonych przewodami o długości ok. 900 km. Osiągane wówczas czasy wykonywania operacji: dodawanie – 0,3 s, mnożenie – 4 s, dzielenie – 10 s. Mark I pracował w systemie dziesiętnym.

Pierwszą elektroniczną maszyną cyfrową był ENIAC. Zaprojektowali go Presper Eckert i John Mauchly. Został on zbudowany w 1946 roku. Zbudowano go z około 18 tys. lamp próżniowych, a ważył 30 ton i potrzebował 175 kW mocy. Ukończenie budowy ENIAC-a uznaje się za początek ery komputerów (patrz pyt. 1.3).

Mówiąc o pierwszym komputerze, należy wspomnieć o wątku Johna Atanasoffa (Amerykanin bułgarskiego pochodzenia), który był profesorem matematyki i fizyki Uniwersytetu Stanowego Iowa w USA. W latach 1937-42 pracował nad komputerem, którego prototyp zaczął działać już w 1939 roku. Nosił nazwę ABC (Atanasoff – Berry Computer). Wykorzystywał on arytmetykę binarną, a jako pamięć – lampy próżniowe i kondensatory. W 1940 roku doszło do spotkania Johna Mauchly'ego z Johnem Atanasoffem, na którym Atanasoff zaznajomił Mauchly'ego ze szczegółami konstrukcyjnymi. Mauchly i Eckert opatentowali swoje pomysły, jednak na przełomie lat 1960-70 odbył się proces sądowy, którego wyrok unieważnił ich patenty. Sąd uznał, że prawa do patentu powinny przysługiwać Atanasoffowi.

Kolejnym etapem w rozwoju komputerów było opublikowanie szczegółów konstrukcyjnych komputera EDVAC. Uczynił to John von Neumann w 1945 roku. **Jego konstrukcja wykorzystywała system dwójkowy i pamięć programu. Obecnie większość komputerów wykorzystuje koncepcję von Neumanna** (patrz pyt. 1.3).

Mniej więcej w tym samym czasie na Uniwersytecie w Cambridge ukończono pracę nad maszyną EDSAC. Nowością było wprowadzenie licznika rozkazów i specjalnego rejestru zwanego akumulatorem.

Pierwszym komputerem sprzedawanym komercyjnie był UNIVAC. Pierwszy egzemplarz tego komputera trafił do amerykańskiego biura rządowego w 1951 roku. Od tego momentu zaczyna się era komercyjnych zastosowań komputerów.

1.3. Który komputer można uznać za pierwszy?

Nie można jednoznacznie wskazać jednej konstrukcji, która byłaby uznawana za pierwszy komputer. To, który komputer był pierwszy, zależy od tego, jakie urządzenie będziemy uważali za komputer.

Jeśli za kryterium będzie uznawany dwójkowy system liczenia, wówczas pierwszym komputerem był ABC.

Jeśli takim kryterium będzie konstrukcja elektroniczna, wtedy pierwszym komputerem był ENIAC.

Kryterium może być też sposób programowania urządzenia. Komputery MARK I czy ENIAC były programowane poprzez zmianę sieci połączeń. Pierwszym komputerem, którego program znajdował się w pamięci był EDVAC skonstruowany przez Johna von Neumanna (patrz pyt. 1.4).

1.4. Co to jest koncepcja von Neumanna (komputer von neumannowski)?

Koncepcja von Neumanna przewiduje umieszczenie danych i programu w pamięci. Dane i program są przechowywane w pamięci w postaci binarnej (dwójkowej – ang. *binary*). Obecnie większość komputerów wykorzystuje architekturę von Neumanna (ang. *control flow computer* – komputer sterowany przepływem rozkazów) – w przeciwieństwie do komputerów sterowanych przepływem danych (ang. *data flow computer*).

1.5. Jak działa komputer?

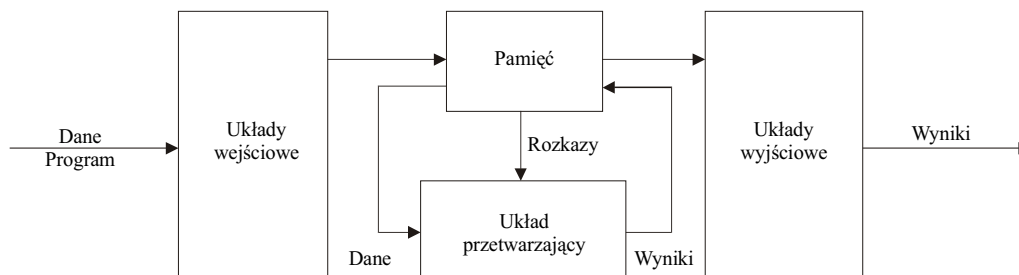
Jak już zostało to zdefiniowane w odpowiedzi 1.1, komputer służy do przetwarzania informacji według określonego algorytmu. Informacje wejściowe i program są wprowadzane do jednostki dokonującej obliczenia, a wyniki są z niej wyprowadzane.

Algorytm jest przepisem określającym sposób postępowania z danymi. To algorytm ustala, jakie operacje i w jakiej kolejności należy wykonać aby otrzymać zamierzony rezultat. Język rozumiany przez komputer jest inny niż ten, którym posługują się ludzie. Składa się on z ograniczonej liczby rozkazów, które zdefiniował jego twórca. Konkretnemu rozkazowi odpowiada ściśle określona reakcja procesora (patrz pyt. 3.4). Przykładowe rozkazy przez niego rozumiane to: „dodaj”, „porównaj”, „podziel”, „prześlij”, „skocz do”, „zatrzymaj”. Zbiór wszystkich rozumianych przez komputer rozkazów to **lista rozkazów języka wewnętrznego**.

Algorytm zapisany w tym języku to program. W komputerach obecnie stosowanych program jest umieszczany w pamięci operacyjnej komputera. W pierwszych komputerach algorytm był zmieniany poprzez zmianę fizycznej sieci połączeń.

Sposób pobierania danych i wysyłania wyników zależy od konkretnego zastosowania komputera. Informacje podlegające przetwarzaniu mogą być wprowadzane do komputera z klawiatury, różnych urządzeń wejścia/wyjścia lub z pamięci. Wyniki obliczeń mogą być prezentowane na monitorze, drukarce lub bezpośrednio mogą sterować jakimś urządzeniem zewnętrznym.

W praktyce, komputer nie wszystkie wyniki obliczeń wysyła na określone urządzenie; niektóre wyniki (np. pośrednie) są kierowane do pamięci i używane ponownie jako dane wejściowe. Takie podejście przyspiesza wykonywanie obliczeń – brak konieczności odczytywania wyników pośrednich i wprowadzania ich ponownie jako danych wejściowych.



Rysunek 1.1. Uproszczony model komputera rzeczywistego

1.6. Z czego buduje się komputery?

Komputery buduje się z półprzewodnikowych układów scalonych. We współczesnych komputerach używa się wielu układów scalonych, pełniących różne funkcje. Sam procesor jest oddzielnym układem scalonym, inne służą przechowywaniu informacji (pamięci), a odrębne zapewniają komunikację z otoczeniem.

Układ scalony zbudowany jest ze struktury półprzewodnikowej, na której znajduje się wiele milionów tranzystorów, struktura zamknięta jest w plastikowej obudowie wyposażonej w wyprowadzenia. W pojedynczym układzie scalonym na niewielkiej powierzchni znajduje się wiele (miliony) elementów elektronicznych. Układy scalone różnią się stopniem złożoności, zwanym stopniem scalenia. Stopień scalenia określa liczbę elementów (diody, tranzystory, rezystory) zawartych w pojedynczym układzie. Rozróżniamy następujące stopnie scalenia:

- mały (SSI) – do kilkudziesięciu elementów
- średni (MSI) – kilkaset elementów
- duży (LSI) – kilka do kilkudziesięciu tysięcy elementów
- wielki (VLSI) – powyżej 100 tys. elementów.

Są to podziały umowne, których nie należy traktować z surową dokładnością. Istnieją inne klasyfikacje, np. takie, w których za układy scalone SSI uznaje się układy zawierające bramki i przerzutniki (patrz pyt. 2.4).

Zaletą układów scalonych są małe wymiary, mały pobór mocy i wysoka niezawodność. Liczba elementów w jednym układzie scalonym stale się zwiększa. Przykładowo w 1971 roku mikroprocesor Intel 4004 zawierał około 2300 tranzystorów, a obecnie procesor Pentium III zawiera kilka milionów elementów.

1.7. Jak klasyfikuje się komputery?

Przyjęta tutaj klasyfikacja uwzględnia zastosowania i budowę komputerów:

- superkomputery
- stacje robocze
- serwery
- komputery osobiste
- komputery przenośne
- komputery przemysłowe.

Superkomputery służą do wykonywania złożonych obliczeń matematycznych. Są bardziej złożone od serwerów. Mają bardzo wiele procesorów pracujących równoległe, kilka GB pamięci RAM i kilkadziesiąt GB pamięci dyskowych.

Stacje robocze to bardzo zubożone wersje komputerów, przeznaczone do współpracy z wydajniejszymi serwerami. Stacja taka korzysta z zasobów systemowych i możliwości obliczeniowych serwera, sama zaś pozostaje tylko pośrednikiem między użytkownikiem a systemem dużego komputera. Dzięki temu możliwości dużego komputera mogą być wykorzystywane przez wielu użytkowników pracujących na stacjach roboczych.

Serwery mogą mieć więcej procesorów, większe pojemności pamięci operacyjnej, większą liczbę dysków twardych lub macierze dyskowe, podtrzymanie napięcia zasilającego.

Komputery osobiste są używane zarówno do pracy, jak i do zabawy. Do tej grupy zaliczyć można komputery wielu producentów. Są one zróżnicowane pod względem budowy i wydajności. Jeszcze parę lat temu były to komputery rodzin takich jak: IBM PC, Apple, Atari serii ST i Amiga firmy Commodore. Obecnie pozostały na rynku komputery klasy IBM PC.

Do klasy **komputerów przenośnych** zaliczamy notebooki, palmtopy, notatniki kieszonkowe, słowniki elektroniczne. Komputery te charakteryzuje ich mobilność, mniejszy pobór mocy, ale też i mniejsza wydajność.

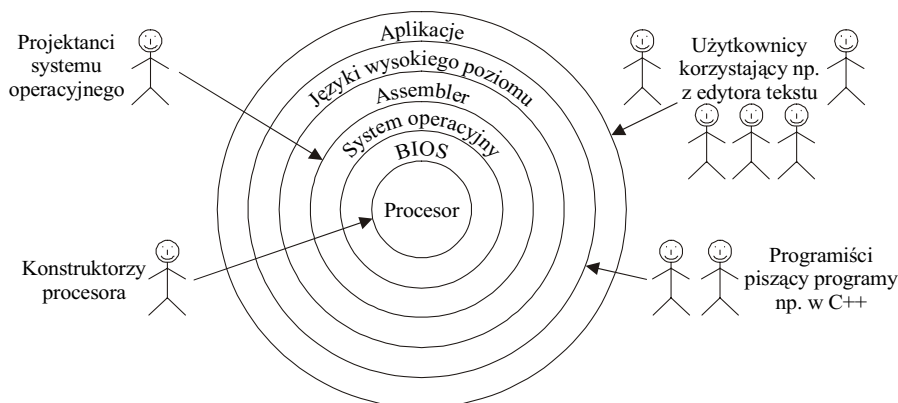
Komputery przemysłowe są to różnego rodzaju sterowniki. Mają one zastosowanie zarówno w domu, jak i w przemyśle. W domu spotyka się je np. w sprzęcie audio-video, w telefonach, w sprzęcie AGD, w przemyśle zaś jako sterowniki procesów technologicznych.

1.8. Co to jest komputer wirtualny?

Termin „virtual” w informatyce zrobił karierę. Według słownika [41] wirtualny (średniowieczno-łaciński virtualis=skuteczny, z łaciny virtus=moc, cnota) to 'mogący zaistnieć, (teoretycznie) możliwy'.

To, jak człowiek widzi komputer, zależy od tego, jak bardzo zagłębia się w jego strukturę. W taki sposób komputer tworzy pewien model warstwowy. Do jednej warstwy ma dostęp projektant procesora, do innej – programista piszący program zwany systemem operacyjnym (patrz pyt. 9.27), a jeszcze do innej osoba korzystająca z aplikacji (programu użytkowego).

Użytkownik korzystający z warstwy wyższej nie interesuje się działaniem warstw niższych. Na przykład programista, używający języka wysokiego poziomu, nie interesuje się bezpośrednim działaniem sprzętu ani systemu operacyjnego. Model takiego komputera przedstawia rysunek 1.2:



* BIOS (patrz pyt. 5.8)

Rysunek 1.2. Model komputera wirtualnego

1.9. Co znaczy określenie „generacja komputerów”?

Wszyscy rozumiemy termin „generacja pokoleniowa”, szczególnie w kontekście starsi a młodzież. Termin generacja komputerów należy rozumieć analogicznie.

Generacje są wyznaczone przez dominujący w danym czasie typ technologii układów elektronicznych używanych do budowy komputerów. Technologia jest kryterium podstawowym, jednak poszczególne generacje charakteryzuje się za pomocą wybra-

nych parametrów funkcjonalnych, których poziom wynika bezpośrednio z możliwości stwarzanych przez nową technologię.

1.10. Jakie znamy generacje komputerów?

W historii rozwoju techniki obliczeniowej wyróżnia się cztery generacje technologiczne.

Czasami wyróżnia się generację zerową, czyli komputery zbudowane z przekaźników.

W **pierwszej generacji** do budowy komputerów stosowano próżniowe lampy elektronowe. Budowane wtedy komputery były urządzeniami olbrzymimi, zawodnymi i powolnymi. Ich obsługa była niewygodna i skomplikowana.

Komputery **drugiej generacji** pojawiły się z chwilą powszechnego użycia do ich budowy układów tranzystorowych. Zdecydowanie zmniejszyła się ich objętość i pobór mocy. Znacznie wzrosła niezawodność i szybkość działania. W komputerach drugiej generacji wprowadzono pamięci ferrytowe (dziś już nie stosowane), co również przyczyniło się do wzrostu prędkości obliczeń.

Kolejny skok technologiczny, wyznaczający początek **trzeciej generacji**, to użycie układu scalonego (patrz pyt. 1.6). Efektem tego był dalszy wzrost niezawodności, zmniejszenie wymiarów i poboru mocy przy dużym wzroście możliwości obliczeniowych. Wówczas to pojawiły się liczne udogodnienia ułatwiające pracę, wśród których za największe trzeba uznać systemy operacyjne (patrz pyt. 9.27).

Za początek **czwartej generacji** przyjmuje się zbudowanie mikroprocesora, którego możliwości porównywane były z możliwościami prostych zestawów minikomputerowych. Wiązało się to z wprowadzeniem układów wykonywanych w technologii LSI oraz VLSI. Cechą komputerów czwartej generacji jest ich zgodność (ang. *compatibility*) oraz możliwość stosowania tych samych programów na komputerach różnych producentów.

Przypuszczalnie komputery kolejnej generacji będą cechowały się sztuczną inteligencją, zdolnością podejmowania samodzielnych decyzji. Przewiduje się, że komputery te będą miały inną budowę. Zamiast struktur półprzewodnikowych będą to komputery kwantowe, optyczne i biologiczne.

1.11. Jakie firmy zajmują się produkcją komputerów?

Obecnie większość sprzedawanych komputerów to komputery zgodne z architekturą (patrz pyt. 6.2) IBM PC. Tak duża popularność komputerów tej klasy jest spowodowana brakiem patentu firmy IBM (świadome działanie tej firmy) na swój produkt.

Dzięki temu wiele małych przedsiębiorstw mogło skorzystać z pomysłu IBM, produkować tańsze i w 100% zgodne z pierwowzorem komputery. Efektem tego było bardzo szybkie rozpowszechnienie się komputerów tej klasy. Duży udział w produkcji tego typu sprzętu mają firmy z Korei, Tajwanu, Singapuru i Hongkongu (czyli tak zwane dalekowschodnie tygrysy).

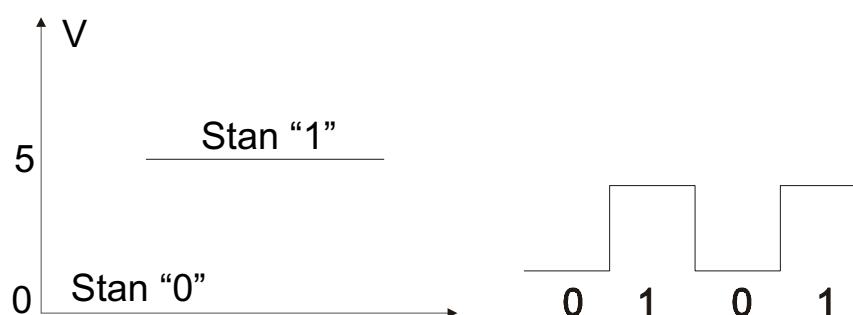
Obecnie produkowane są też komputery firmowe zgodne z IBM PC. Znanymi producentami są IBM, Hewlett Packard, DEC, Dell, Compaq, NEC, Acer, Toshiba.

Znani są też producenci komputerów o specjalnych przeznaczeniach. Często jako serwery sieciowe używane są komputery firmy SUN, natomiast jako superkomputery – NEC, Cray, Silicon Graphics, IBM, Hewlett Packard, SUN.

2. Budowa komputera i zasady jego działania

2.1. W jaki sposób przedstawiane są informacje wewnątrz komputera?

Zgodnie z podaną wcześniej definicją, komputer jest urządzeniem elektronicznym, a zatem nośnikami informacji muszą w nim być sygnały elektryczne. Są to sygnały dwustanowe, tzn. takie, że ich parametry zawsze mieszczą się w jednym z dwóch dopuszczalnych przedziałów. W celu rozróżnienia obu stanów przypisuje im się dwa różne symbole – cyfry 0 i 1 (rys. 2.1).



Rysunek 2.1. Sygnał dwustanowy

Wszystkie informacje przechowywane i przetwarzane przez komputer są zapisywane w systemie dwójkowym. Systemem, jaki stosujemy na co dzień, jest system dziesiętny. Warto sobie uświadomić, że $L = 523$ to $L = 5 \cdot 10^2 + 2 \cdot 10^1 + 3 \cdot 10^0$. Tę zasadę można zapisać wzorem:

$$L_{10} = \sum_{i=0}^{n-1} c_i \cdot 10^i = c_{n-1} \cdot 10^{n-1} + c_{n-2} \cdot 10^{n-2} + \dots + c_1 \cdot 10^1 + c_0 \cdot 10^0$$

w którym n oznacza liczbę cyfr, a c_i – i -tą cyfrę liczby dziesiętnej ($c_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$).

System dziesiętny jest szczególnym przypadkiem pozycyjnego systemu zapisu o dowolnej podstawie, w którym wartość liczby L_r oblicza się wg wzoru:

$$L_r = \sum_{i=0}^{n-1} b_i * r^i = b_{n-1} * r^{n-1} + b_{n-2} * r^{n-2} + \dots + b_1 * r^1 + b_0 * r^0$$

gdzie współczynnik b_i oznacza i -tą cyfrę liczby a r podstawę systemu ($b_i \in \{0, 1, \dots, r-1\}$). W systemie dziesiętnym $r = 10$. Liczba różnych cyfr równa się podstawie systemu.

Ze względu na charakter sygnałów stosowanych w komputerach najwygodniejszym systemem przedstawiania liczb wewnątrz komputera jest pozycyjny system dwójkowy.

$$L_2 = \sum_{i=0}^{n-1} a_i * 2^i = a_{n-1} * 2^{n-1} + a_{n-2} * 2^{n-2} + \dots + a_1 * 2^1 + a_0 * 2^0$$

gdzie współczynnik a_i oznacza i -tą cyfrę liczby dwójkowej ($a_i \in \{0, 1\}$).

Cyfry dwójkowe nazywamy bitami. Jest to powszechnie przyjęty skrót odpowiedniego terminu angielskiego (ang. *binary digit*).

Do określania rozmiarów pamięci komputerowych używane jest pojęcie bajtu (ang. *byte*). Bajt składa się z ośmiu bitów. Za pomocą ośmiu bitów można zapisać $2^8 = 256$ ciągów zerjedynkowych, co umożliwi reprezentację liter, cyfr, semigrafiki oraz znaków narodowych. W kodzie ASCII (patrz pyt. 5.5) każdy znak jest reprezentowany za pomocą ośmiu bitów.

2.2. Jak są reprezentowane wewnątrz komputera informacje, które nie są liczbami?

Wszystkie tak zwane alfanumeryczne informacje (np. litery, cyfry, znaki) wewnątrz komputera trzeba przedstawiać za pomocą wybranych ciągów zerjedynkowych. Takie przyporządkowanie ciągów obiektom, które mają reprezentować, nazywa się kodowaniem. Natomiast ciągi zerjedynkowe reprezentujące określone informacje nazywa się kodami. Kody definiuje się za pomocą tablic.

Wśród wielu możliwych kodów najistotniejsze są kody rozkazów komputera i kody znaków alfanumerycznych stosowane do przedstawiania tekstów zrozumiałych dla człowieka. Tablice definiujące kody rozkazów są zawarte w dokumentacji technicznej dostarczanej przez producenta procesora. Kody znaków alfanumerycznych są znormalizowane i ujęte w odpowiednie standardy takich organizacji, jak ISO czy ANSI.

2.3. Czy przy określaniu informacji przechowywanych wewnątrz komputera używa się wyłącznie pozycyjnego systemu dwójkowego?

Wewnątrz komputera używa się tylko pozycyjnego systemu dwójkowego. System ten jest wygodny z punktu widzenia konstruktorów komputera, natomiast jest bardzo kłopotliwy dla jego użytkowników. Kody rozkazów mają po kilkadziesiąt bitów. Programista, który chciałby za pomocą tak długich ciągów zer i jedynek zapisywać kolejne rozkazy swojego programu, myliłby się nieustannie. W praktyce stosuje się dwa wygodne dla człowieka sposoby zapisu liczb **prezentujące na zewnątrz informacje wewnętrzne komputera**. Są to systemy pozycyjne: ósemkowy i szesnastkowy.

W systemie ósemkowym korzysta się z cyfr należących do zbioru $\{0, 1, \dots, 7\}$, a wyrażenie określające wartość liczby uzyskuje się podstawiając $r = 8$ we wzorze z pytania 2.1. Ciąg cyfr $465_{(8)}$ tworzący liczbę ósemkową ma więc wartość: $4 \cdot 8^2 + 6 \cdot 8^1 + 5 \cdot 8^0 = 309_{(10)}$. Indeksy (8) i (10) oznaczają – odpowiednio – formę zapisu ósemkową i dziesiętną.

W systemie szesnastkowym używa się z 16 symboli. Najczęściej korzysta się z następującego zbioru: $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F\}$. Symbole literowe oznaczają cyfry, których wartości dziesiętne wynoszą kolejno: 10, 11, 12, 13, 14 i 15. Wyrażenie wyznaczające wartość liczby uzyskuje się podstawiając $r = 16$ we wzorze z pytania 2.1.

Popularność systemów ósemkowego i szesnastkowego do reprezentacji informacji wewnętrznych komputera wynika z łatwości konwersji wartości binarnych na te systemy. Aby dokonać zamiany liczby binarnej na szesnastkową, należy pogrupować bity po cztery (poczynając od prawej strony) i takie czwórki zastąpić symbolami systemu szesnastkowego (rys. 2.2).

$$\begin{array}{cccc} \underbrace{1001} & \underbrace{0010} & \underbrace{1010} & \underbrace{1100} & \text{- liczba binarna} \\ 9 & 2 & A & C & \text{- liczba szesnastkowa} \end{array}$$

Rysunek 2.2. Zamiana liczby binarnej na szesnastkową

Podobnie wygląda zamiana na liczbę w systemie ósemkowym. W tym przypadku bity grupujemy po trzy (poczynając od prawej strony) i dokonujemy zamiany na symbol systemu ósemkowego (rys. 2.3).

$$\begin{array}{cccccc} \underbrace{001} & \underbrace{001} & \underbrace{001} & \underbrace{010} & \underbrace{101} & \underbrace{100} & \text{- liczba binarna} \\ 1 & 1 & 1 & 2 & 5 & 4 & \text{- liczba ósemkowa} \end{array}$$

Rysunek 2.3. Zamiana liczby binarnej na ósemkową

Prawidłowość obliczeń można sprawdzić poprzez system dziesiętny.

$$92AC_{(16)} = 9 \cdot 16^3 + 2 \cdot 16^2 + A \cdot 16^1 + C \cdot 16^0 = 37548_{(10)}$$

$$111254_{(8)} = 1 \cdot 8^5 + 1 \cdot 8^4 + 1 \cdot 8^3 + 2 \cdot 8^2 + 5 \cdot 8^1 + 4 \cdot 8^0 = 37548_{(10)}$$

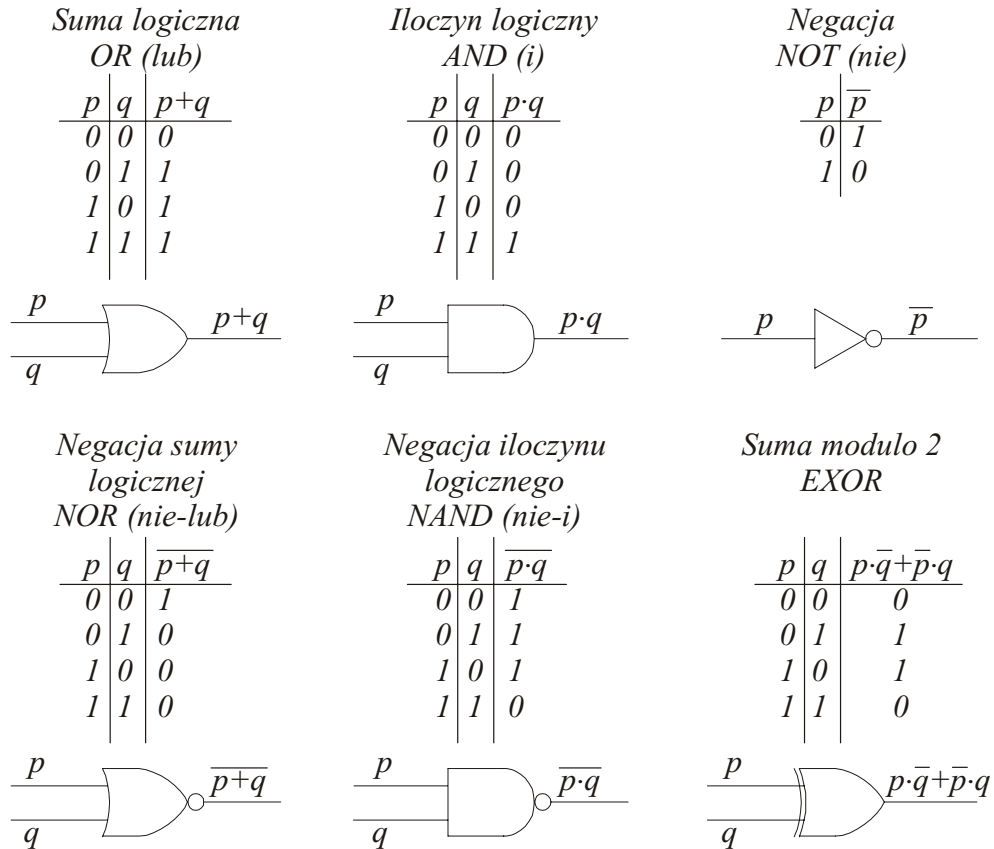
Można sprawdzić, że:

$$1001001010101100_{(2)} = 1 \cdot 2^{15} + 0 \cdot 2^{14} + \dots + 1 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0 = 37548_{(10)}$$

2.4. Co zadecydowało o takim, a nie innym sposobie przedstawiania informacji wewnątrz komputera?

Podstawową przyczyną używania systemu dwójkowego jest łatwość wytwarzania i duża niezawodność elektronicznych układów dwustanowych. Nietrudno to zrozumieć, jeśli uświadomimy sobie, że najprostszym elementem dwustanowym jest zwykły wyłącznik przerywający lub zamykający obwód elektryczny. Co ciekawsze, taki właśnie „mikrowyłącznik” półprzewodnikowy stanowi podstawowy element układów, z których buduje się komputery. Co do niezawodności, to należy ją rozumieć jako pewność rozpoznawania stanów sygnału. Przy dwóch stanach wyrażanych przez dwa różne, dostatecznie odległe od siebie poziomy napięcia, pewność ta jest największa. Wadą sygnałów dwustanowych jest mała ilość przekazywanej informacji. Aby zwiększyć pojemność informacyjną urządzeń budowanych z elementów dwustanowych, wystarczy zwiększyć liczbę sygnałów wyjściowych układu, tzn. zastosować odpowiednią kombinację wielu układów elementarnych. Jeżeli urządzenie będzie miało dwa sygnały wyjściowe, to liczba możliwych stanów wyjściowych wyniesie cztery. Jeżeli będzie miało trzy sygnały wyjściowe, to możliwych będzie osiem stanów wyjściowych. Ogólnie liczba możliwych stanów wyjściowych jest równa 2^n , gdzie n oznacza liczbę sygnałów wyjściowych. Takie urządzenia budowane z elementów dwustanowych są prostsze i łatwiej je wykonać, niż urządzenia korzystające z sygnałów wielostanowych.

Drugą ważną przyczyną, decydującą o monopolu elementów dwustanowych, jest łatwość ich opisu, analizy i syntezy. Stany 0 i 1, utożsamiane poprzednio z cyframi dwójkowymi, można potraktować jako wartości logiczne „fałsz” i „prawda”. Wówczas okaże się, że działanie elementów dwustanowych opisują wprost operacje dwuelementowej algebry Boole’a. Algebrę tę definiuje dwuelementowy zbiór $\{0, 1\}$ oraz trzy operacje: $+$, \cdot , oraz $\bar{}$. Zmienne należące do zbioru $\{0, 1\}$ oraz wymienione operacje nazywamy zmiennymi i operacjami logicznymi. Układy realizujące operacje logiczne nazywamy bramkami (ang. *gates*) logicznymi. Na rysunku 2.4 przedstawiono tablice definiujące poszczególne operacje oraz symbole odpowiednich bramek. Oprócz bramek podstawowych przedstawiono także bramki NOR, NAND i EXOR. Znaczenie bramek NAND i NOR wynika z faktu, iż za pomocą każdej z nich można zrealizować dowolną operację podstawową.



Rysunek 2.4. Podstawowe operacje i bramki logiczne

2.5. Z jakich bloków funkcjonalnych buduje się komputery?

Podstawowymi blokami, jakie można wyróżnić we współczesnym komputerze, są:

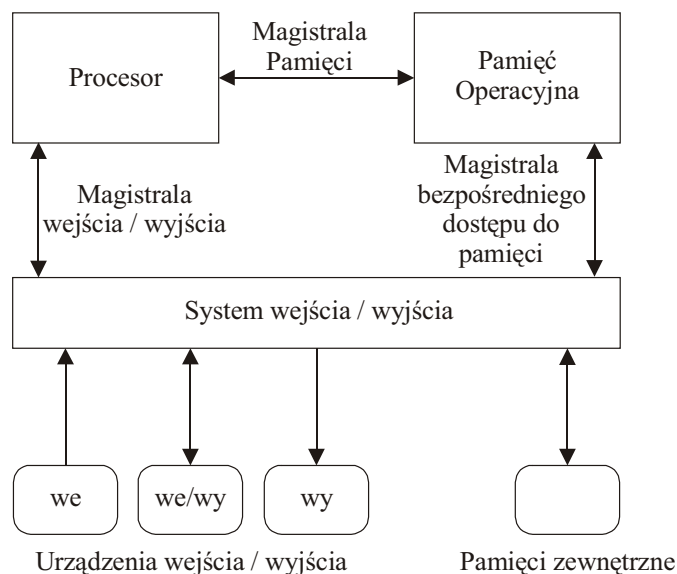
- procesor lub procesory
- pamięć operacyjna
- pamięci zewnętrzne
- urządzenia wejścia/wyjścia.

2.6. Jak jest zbudowany komputer?

Punktem wyjścia do odpowiedzi jest model komputera przedstawiony w odpowiedzi na pytanie 2.5.

Najpierw zostanie omówiony układ przetwarzający. W rzeczywistych komputerach nosi on nazwę procesora. Można w nim wyodrębnić: jednostkę arytmetyczno-logiczną (ALU), część sterującą i rejestry. Część przetwarzająca zawiera arytmometr oraz różnego rodzaju rejestry np. do pamiętania stanu procesora, pamiętania wyników pośrednich. Część sterująca odpowiada za pobieranie z pamięci i wykonywanie kolejnych rozkazów programu. Procesor współpracuje bezpośrednio z pamięcią operacyjną. W pamięci przechowywane są rozkazy sterujące (ciąg rozkazów to program) oraz dane podlegające przetwarzaniu oraz wyniki. Komunikowanie się procesora z otoczeniem umożliwiają urządzenia wejścia/wyjścia.

Urządzenia wejścia/wyjścia komunikują się z procesorem poprzez system wejścia/wyjścia, zapewniający odpowiednie dopasowanie i konwersję sygnałów przesyłanych między procesorem i urządzeniami.



Rysunek 2.5. Schemat blokowy komputera

Wszystkie wymienione bloki są połączone za pomocą magistral. Zamiast łączyć każdy blok z każdym innym blokiem ($\frac{n \cdot (n-1)}{2}$ połączeń, gdzie n – liczba bloków)

wymyślono magistralę, czyli w dużym uproszczeniu wiązkę przewodów (lub zespół ścieżek na obwodzie drukowanym), po to aby móc transmitować informacje. Do magistrali urządzenia są dołączane równolegle. Oczywiście jednocześnie nadawcą infor-

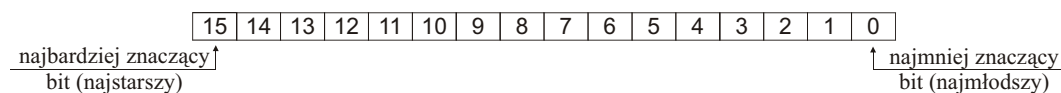
macji do magistrali może być jeden blok, a odbiorcami kilka bloków. Rozwiązanie takie wymaga stosowania układów sterujących dostępem do magistrali.

2.7. Jakie podstawowe parametry charakteryzują procesor?

Podstawowymi parametrami charakteryzującymi procesor są:

- długość słowa
- częstotliwość taktowania magistrali
- częstotliwość taktowania procesora
- liczba rejestrów i ich przeznaczenie
- wielkość pamięci pomocniczej typu cache
- lista rozkazów.

Przez słowo komputera rozumiemy ciąg bitów o ustalonej długości, który w różnych operacjach traktujemy niepodzielnie. Długość słowa decyduje bezpośrednio lub pośrednio o wielu innych parametrach komputera. Typowe długości słów są zwykle wielokrotnościami bajtu, czyli ciągu 8-bitowego. Najczęściej spotykane słowa mają 16, 32 lub 64 bity. Na rysunku 2.6 przedstawiono sposób numerowania bitów w słowie.



Rysunek 2.6. Słowo komputera

Warto zauważyć, że tak określone numery bitów są równocześnie wykładnikami potęg stanowiących wagi poszczególnych bitów we wzorze z pytania 2.1.

Szybkość taktowania procesora jest ograniczana przez technologię wykonania samego układu scalonego. Ustalana jest ona za pomocą układu zwanego zegarem i wymuszającego rytm pracy procesora. Im szybkość taktowania jest większa, tym krócej procesor wykonuje poszczególne operacje.

Częstotliwość taktowania magistrali nie jest jednoznaczna z częstotliwością pracy samego procesora. Obecnie sprzedawane procesory mogą pracować z częstotliwościami bliskimi kilku GHz, jednak magistrala pracuje z częstotliwością rzędu setek MHz. Ta rozbieżność wynika ze stosowania tak zwanych mnożników w procesorach, przez które mnożona jest prędkość taktowania szyny.

Rejestry procesora mają najczęściej długość jednego słowa komputera. Duża część rejestrów wewnętrznych procesora stanowi pamięć podręczną (ang. *cache*) słu-

zając przechowywaniu wyników pośrednich operacji. Część rejestrów jest związana z częścią sterującą procesora, część z jednostką arytmetyczno-logiczną.

2.8. Jakie podstawowe parametry charakteryzują pamięć komputera?

Podstawowymi parametrami charakteryzującymi pamięć są: pojemność i czas dostępu. Pojemność pamięci określa liczba komórek przechowujących informacje. Podstawowym elementem, jaki można zapamiętać, jest jeden bit. W większości pamięci nie ma możliwości zapisywania i odczytywania pojedynczych bitów. Najmniejszą porcją informacji przesyłaną między procesorem i pamięcią jest zwykle zawartość komórki wielobitowej. Zwykle jedna komórka zawiera jedno słowo komputera, często jednak jest od niego krótsza. Wówczas jedno słowo trzeba przechowywać w dwu lub więcej komórkach pamięci.

Pojemności pamięci są zwykle dość znacznymi liczbami, dlatego wyraża się je w tysiącach, milionach bądź miliardach komórek.

$$1 \text{ k} = 2^{10} = 1024 \approx 1000 \quad (\text{k} - \text{kilo})$$

$$1 \text{ M} = 2^{20} = 1048576 \quad (\text{M} - \text{mega})$$

$$1 \text{ G} = 2^{30} = 1073741824 \quad (\text{G} - \text{giga})$$

Tak więc pamięć o pojemności np. 64 M komórek ma:

$$64 * 1048576 = 67108864 \text{ komórek.}$$

Zwykle operuje się bajtami, a więc używa się skrótów kB, MB, GB. Każda komórka pamięci ma swój jednoznaczny identyfikator, zwany adresem. Adres jest liczbą całkowitą z przedziału $[0, M-1]$, gdzie M oznacza pojemność pamięci.

Adresy komórek pamięci wyraża się liczbami dwójkowymi. Liczba bitów przeznaczona na zapisanie adresu decyduje o maksymalnej wartości tego adresu, a zatem i maksymalnej pojemności pamięci, z jaką będzie mógł współpracować dany procesor.

Drugim ważnym parametrem charakteryzującym pamięć jest czas dostępu. Określa on szybkość pamięci, czyli minimalny czas od chwili zmiany adresu na wejściu pamięci, do ustalenia danych na wyjściu pamięci. Czas ten jest zależny od technologii wykonania układu scalonego pamięci.

2.9. Jakie podstawowe parametry charakteryzują urządzenia wejścia/wyjścia?

Trudno jest wyczerpująco odpowiedzieć na to pytanie. Parametry urządzenia zewnętrznego są ściśle związane z rodzajem urządzenia, rodzajem informacji wymienianych pomiędzy komputerem a urządzeniem i sposobem jego podłączenia. Do najważniejszych, wspólnych dla wielu urządzeń zaliczyć można:

- sposób podłączenia urządzenia do komputera (poprzez port standardowy, czy rozwiązanie niestandardowe)
- szybkość wprowadzania i wyprowadzania informacji
- sposób wprowadzania i wyprowadzania informacji (szeregowy, równoległy)
- sposób kodowania i postać przesyłanej informacji.

2.10. W jaki sposób procesor współpracuje z pamięcią operacyjną?

Procesor jest połączony z pamięcią trzema magistralami: sterującą, adresową i danych. Liczba linii magistrali adresowej jest zależna od pojemności pamięci i jest równa liczbie bitów adresu (patrz pyt. 2.8). Liczba linii magistrali danych jest równa długości komórki pamięci. Budowa magistrali sterującej zależy od konkretnego rozwiązania konstrukcyjnego. Typowa szyna sterująca zawiera linię wyboru operacji (odczyt, zapis) oraz linie sygnalizujące stan pamięci.

W celu odczytania zawartości jakiejś komórki procesor przesyła do pamięci, liniami adresowymi, adres tej komórki. Następnie wysyłany jest sygnał czytania, który uruchamia układy sterujące pamięci. Na podstawie adresu układ sterujący lokalizuje wybraną komórkę. Jej zawartość przesyłana jest na wyjście pamięci, skąd magistralą danych jest przekazywana do procesora.

Operacja zapisu przebiega podobnie. Procesor na magistrali adresowej umieszcza adres komórki, informacje przeznaczone do zapisania na magistrali danych, a następnie wysyła sygnał zapisu magistralą sterującą. Na podstawie adresu układ sterujący odszukuje wybraną komórkę, a następnie uaktualnia jej zawartość informacjami pobranymi z magistrali danych.

2.11. W jaki sposób komputer współpracuje z urządzeniami wejścia/wyjścia?

Urządzenia zewnętrzne zwykle nie są podłączane bezpośrednio do procesora. Wynika to z faktu, że urządzenia te zawierają zwykle oprócz układów elektronicznych również elementy elektromechaniczne. Stąd ich parametry dość znacznie odbiegają od parametrów procesora (poziomy napięcie i prądów, szybkość działania). Aby zapewnić współpracę procesora z układami wejścia/wyjścia, opracowano specjalne układy dopasowujące. Zbiór układów elektronicznych zapewniających dopasowanie, wraz z regułami wymiany danych i informacji sterujących, nosi nazwę systemu wejścia/wyjścia. W skład tego systemu wchodzi m.in. magistrala wejścia/wyjścia. Zawiera ona linie adresowe pozwalające procesorowi wskazać urządzenie, z którym chce współpracować, linie danych oraz linie sterujące i linie stanu. Komunikacja procesora z systemem odbywa się podobnie jak z pamięcią.

Magistrala wejścia/wyjścia może być wydzieloną szyną komputera lub może być współdzielona z magistralą pamięci. W pierwszym rozwiązaniu, do komunikacji procesora z urządzeniami wejścia/wyjścia używa się innych rozkazów, niż do wymiany informacji z pamięcią. W drugim przypadku, gdy magistrala pamięci i systemu wejścia/wyjścia jest wspólna, urządzenia wejścia/wyjścia są przez procesor traktowane jak obszar pamięci o odrębnych adresach. Wówczas do wymiany informacji z pamięcią i systemem wejścia/wyjścia używa się tych samych rozkazów.

2.12. W jaki sposób procesor komunikuje się z urządzeniami wejścia/wyjścia?

Urządzenia mogą się komunikować z komputerem na trzy sposoby:

- poprzez procesor – obsługa programowa
- za pomocą przerwań
- poprzez bezpośredni dostęp do pamięci – DMA (ang. *Direct Memory Access*).

Pierwsze rozwiązanie wiąże się ze znacznym obciążeniem procesora. W tym przypadku procesor monitoruje stan urządzenia oczekując nadejścia danych.

W drugim rozwiązaniu, urządzenia informują procesor o konieczności reakcji. Metoda ta jest bardziej efektywna, gdyż pozwala zwolnić procesor z ciągłego badania stanu urządzeń. W tej metodzie, wszystkie informacje między urządzeniami i procesorem komputera przesyłane są za pośrednictwem rejestrów procesora.

Trzeci sposób przekazywania danych między urządzeniami a pamięcią jest najszybszy. W tym przypadku procesor jedynie steruje przepływem danych, które są przesyłane bezpośrednio między określonym urządzeniem a pamięcią. Procesor tylko

nadzoruje transmisję. Dzięki temu oszczędza się czas tracony na podwójne przenoszenie danych (z urządzenia do procesora i z procesora do pamięci).

2.13. Jak są zasilane obecne komputery?

Typowe komputery stacjonarne są zasilane prądem przemiennym o napięciu 220V, który jest doprowadzany do zasilacza umieszczonego wewnątrz obudowy. W chwili obecnej używa się wyłącznie zasilaczy impulsowych. Ich zaletą są niewielkie wymiary, mała waga i duża sprawność przy dość dużych mocach. Zasada pracy takiego zasilacza polega na:

- wyprostowaniu prądu zasilającego
- zmodulowaniu napięcia przebiegiem o częstotliwości rzędu kHz
- poddaniu transformacji przebiegu o wysokiej częstotliwości
- wyprostowaniu prądu otrzymanego z transformatora.

Operacje te mają na celu zmniejszenie rozmiarów transformatora (przetwarzanie prądu o wyższej częstotliwości umożliwia zastosowanie transformatora o mniejszych wymiarach).

Na wyjściu takiego zasilacza otrzymuje się napięcia stałe o wartościach: +5 V, +12 V, -5 V i -12 V.

Wewnątrz zasilacza znajduje się również wentylator, którego zadaniem jest wymuszenie przepływu powietrza wewnątrz komputera.

Nowsze zasilacze mogą być włączane i wyłączane przez płytę główną komputera. Umożliwia to automatyczne włączanie komputera o określonej godzinie, po wciśnięciu przycisku myszy lub klawiatury oraz wyłączanie komputera po programowym, a nie ręcznym zamknięciu systemu.

2.14. Co to jest UPS i jaką rolę pełni w zasilaniu komputera?

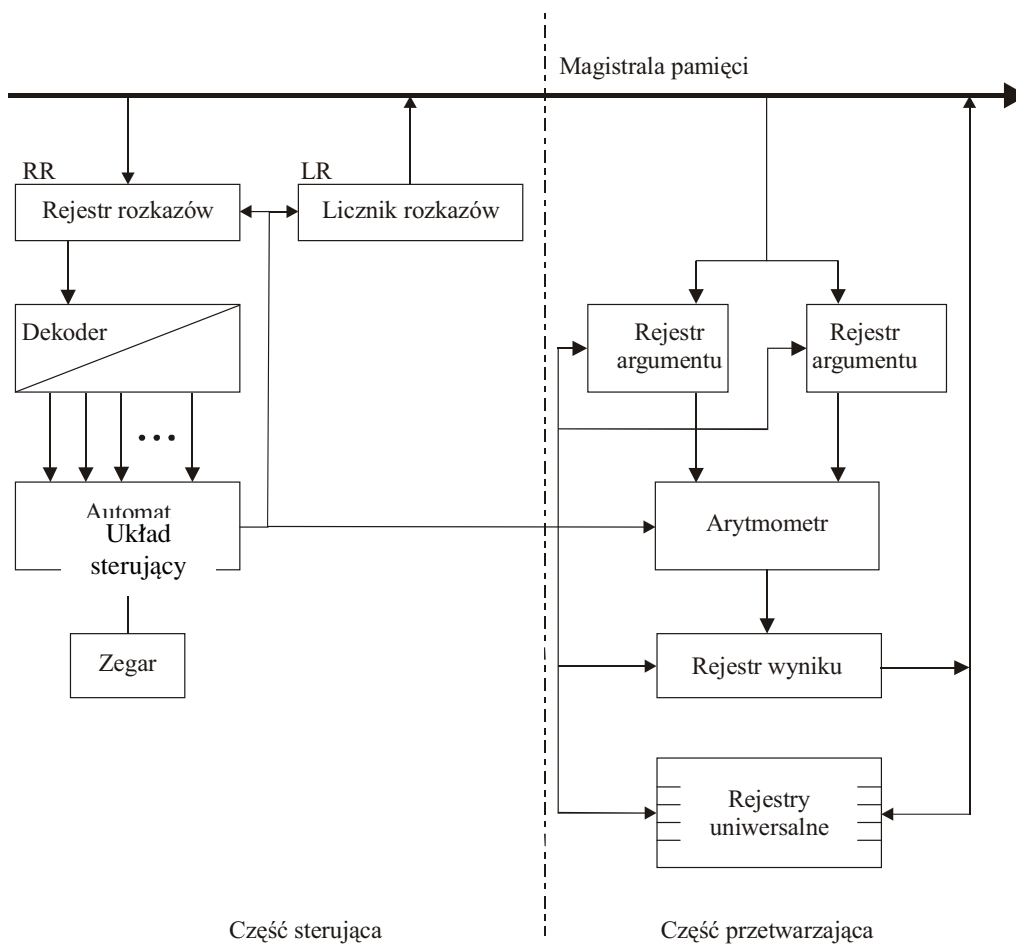
UPS (ang. *Uninterruptable Power System* – bezprzerwowo system zasilania) służy podtrzymaniu napięcia zasilającego podczas niedługich (kilka do kilkudziesięciu minut – zależy to od obciążenia systemu) przerw w dostawach energii elektrycznej. Przerwy te są jedną z najczęstszych przyczyn utraty danych. UPS-y pozwalają jedynie poprawnie (a nie raptownie) zakończyć pracę systemu komputerowego (po to by po powrocie energii elektrycznej można byłoby ja poprawnie kontynuować). UPS-y nie są więc zasilaczami awaryjnymi tak jak agregaty prądotwórcze (np. dla bloku operacyjnego w szpitalu). Obecnie produkowane UPS-y zabezpieczają także przed przepięciami i spadkami napięcia. UPS-y mogą samoczynnie zamykać system, gdy zapas energii w akumulatorach spadnie poniżej określonego poziomu.

UPS-y mogą być typu on-line, off-line i mieszane. W UPS-ach typu on-line komputer cały czas jest zasilany prądem, wytworzonym z energii zmagazynowanej w akumulatorze, a prąd z sieci energetycznej służy jedynie doładowywaniu akumulatora. Ten typ UPS-a jest najlepszy, ale i najdroższy. Przy stosowaniu UPS-ów typu off-line energia z baterii służy jedynie podtrzymaniu napięcia, w przypadku zaników napięcia w sieci energetycznej. Podczas normalnej pracy komputer jest zasilany bezpośrednio z sieci, a akumulator jest doładowywany prądem konserwującym. W rozwiązaniach mieszanych podczas normalnej pracy napięcie sieci jest jedynie korygowane.

3. Procesor

3.1. Jak jest zbudowany procesor?

Na rysunku 3.1 pokazano uproszczony schemat funkcjonalny procesora. Linia przerywana oddziela dwie główne jego części: sterującą i przetwarzającą.



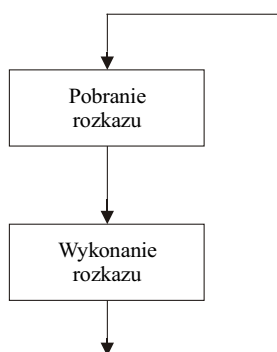
Rysunek 3.1. Uproszczony schemat funkcjonalny procesora

Rdzeń części przetwarzającej stanowi arytmometr, wykonujący podstawowe operacje arytmetyczne i logiczne. Z wejściem i wyjściem arytmometru współpracują rejestry, przechowujące argumenty i wynik operacji. Specjalnie wyróżnionym rejestrem, przechowującym argument operacji i wynik, jest akumulator. Oprócz rejestrów przypisanych do arytmometru, w części przetwarzającej, mogą występować jeszcze rejestry uniwersalne, wykorzystywane przy różnych operacjach wewnątrz procesora. Arytmometr komunikuje się poprzez magistrale wewnętrzne z różnymi układami wewnątrz procesora oraz poprzez magistralę pamięci z pamięcią komputera. Oprócz linii danych do procesora dochodzą także sygnały sterujące.

Część sterująca zawiera układ taktowany zegarem, generujący synchronicznie wszystkie niezbędne sygnały sterujące pracą układów wewnętrznych procesora oraz układów zewnętrznych. Głównymi rejestrami znajdującymi się w tej części są: licznik rozkazów (LR) wskazujący zawsze miejsce w pamięci, z którego należy pobrać kolejny rozkaz, oraz rejestr rozkazów (RR) przechowujący aktualny rozkaz. Licznik rozkazów współpracuje z częścią adresową magistrali pamięci oraz z arytmometrem. Rejestr rozkazów jest połączony z liniami danych magistrali pamięci oraz dekoderym wytwarzającym odpowiednie sygnały identyfikujące wymaganą operację. W części sterującej mogą również wystąpić specjalne rejestry wykorzystywane przy wyznaczeniu argumentów operacji.

3.2. Jak działa procesor?

Działanie procesora sprowadza się do cyklicznego powtarzania dwóch podstawowych operacji: pobrania rozkazu i wykonania rozkazu.



Rysunek 3.2. Podstawowy cykl pracy procesora

Przerwanie tego cyklu może nastąpić tylko w przypadku napotkania specjalnego rozkazu, nakazującego zatrzymanie procesora lub w wyniku interwencji z zewnątrz.

Na każdą z wymienionych czynności składa się ciąg czynności elementarnych, zwanych krokami. Cały cykl procesora składa się z następujących kroków:

- odczytanie z pamięci rozkazu wskazanego przez licznik rozkazów LR i umieszczenie go w rejestrze rozkazów RR
- zmiana zawartości LR, tak aby wskazywał następny rozkaz
- dekodowanie rozkazu, określenie jego typu i liczby argumentów
- wyznaczenie adresów i pobranie niezbędnych argumentów
- wykonanie wymaganych czynności wskazanych w treści rozkazu i zapamiętanie wyniku
- przejście do pobrania kolejnego rozkazu.

Podczas wykonywania każdego kroku są generowane niezbędne sekwencje sygnałów sterujących. Sygnały te powodują przesyłanie danych, uruchamiają układy arytmetyczne i logiczne, testują obecność określonych sygnałów.

3.3. Co to jest adresowanie i jakie znamy najważniejsze rodzaje adresowania?

Przez adresowanie rozumiemy zasady określania i wyznaczania argumentów rozkazu. Jeżeli w danym komputerze są możliwe różne rodzaje adresowania, to sposób kodowania rozkazu musi uwzględniać możliwość wyboru jednego z nich. Rodzaj adresowania może wynikać z kodu operacji lub być wskazany przez specjalne pole trybu adresowania, związane z polem argumentu.

Do najważniejszych rodzajów adresowania zaliczamy:

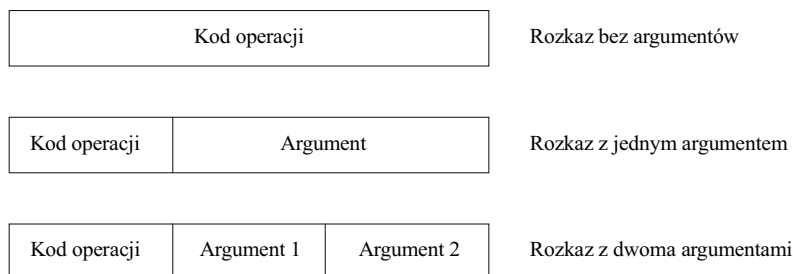
- adresowanie natychmiastowe (argument operacji zawarty bezpośrednio w treści rozkazu)
- adresowanie bezpośrednie (w rozkazie oprócz kodu operacji występuje adres argumentu)
- adresowanie pośrednie (adres zawarty w treści rozkazu nie wskazuje argumentu ani jego adresu, lecz miejsce, w którym znajduje się adres argumentu – czyli adres adresu argumentu)
- indeksowanie (adres argumentu powstaje w wyniku sumowania części adresowej rozkazu i zawartości specjalnego rejestru indeksowego procesora)
- adresowanie względem rejestru bazowego (adres argumentu powstaje w wyniku sumowania odpowiedniego pola rozkazu z zawartością specjalnego rejestru bazowego procesora)
- adresowanie względem licznika rozkazów (adres argumentu otrzymuje się przez dodanie części adresowej rozkazu do zawartości licznika rozkazów).

3.4. Jaka postać mają rozkazy komputera?

Wewnątrz komputera rozkazy, tak jak inne informacje, są ciągami bitów o ustalonej długości, którym odpowiadają pewne liczby. Liczby te są nazywane kodami rozkazów, a ich znaczenie jest określone przez konstruktora procesora. Zbiór rozkazów „rozumianych” przez procesor tworzy tak zwaną listę rozkazów procesora. Nie wszystkie rozkazy są niezbędne, jednak istnienie ich ułatwia pracę programistów, umożliwia tworzenie złożonych struktur danych (patrz rozdz. 9)

W różnych komputerach długość rozkazów może być różna. W niektórych komputerach wszystkie rozkazy mają jednakową długość jednego słowa. Częściej jednak procesor zawiera rozkazy różnej długości, od jednego do kilku słów. Długość rozkazu zależy głównie od liczby bitów w słowie. Im słowo jest krótsze, tym częściej występują rozkazy wielosłowne.

Każdy rozkaz składa się z kilku pól. Jedno z nich występuje zawsze i nosi nawet pola kodu operacji. Kod ten definiuje funkcję rozkazu, czyli czynności jakie należy wykonać. Pozostałe pola zawierają argumenty operacji. Liczba tych pól zależy od rodzaju operacji, jakiej odpowiada rozkaz. W rozkazach bez argumentów (wyzeruj coś, dodaj jedynkę do...) pola dodatkowe nie występują. Sposób rozmieszczenia wymienionych elementów w słowie lub słowach komputera nazywamy formatem rozkazu.



Rysunek 3.3. Typowe formaty rozkazów procesora

Długość kodu operacji jest uzależniona od liczby rozkazów obsługiwanych przez procesor. Pole n-bitowe może pomieścić tylko $2^n - 1$ różnych ciągów zerojedynkowych. Na długość całego rozkazu znacznie większy wpływ ma liczba i typ jego argumentów.

3.5. Jak klasyfikujemy rozkazy komputera?

Jeśli jako kryterium klasyfikacji przyjmie się podział funkcjonalny, wówczas można wyróżnić następujące grupy rozkazów:

- rozkazy przesyłające informacje (powodują przemieszczenia informacji bez zmieniania jej treści)
- rozkazy arytmetyczne i logiczne (wykonują operacje na argumentach)
- rozkazy sterujące programem (umożliwiają zmianę kolejności wykonywania rozkazów, cykliczne powtarzanie rozkazów – pętle, podejmowanie decyzji – rozgałęzienia)
- rozkazy wejścia/wyjścia (służą wprowadzaniu i wyprowadzaniu informacji do i z procesora).

3.6. Jak są przetwarzane dane w procesorze potokowym, skalarnym?

Procesory tego typu przetwarzają jednocześnie kilka rozkazów. Koncepcja ta wynika z konieczności zwiększenia wydajności procesorów. W standardowym procesorze fazy rozkazu (pobranie, wykonanie) były wykonywane kolejno. W ten sposób kolejne układy procesora czekały beczynnie na krok, w którym będą potrzebne.

Koncepcja przetwarzania potokowego polega na jednoczesnym wykonywaniu większej liczby rozkazów. W tym przypadku, każdy z układów procesora po zakończeniu pracy nad bieżącym rozkazem, przechodzi do wykonywania następnego.

W przetwarzaniu równoległym procesor jest wyposażony w wiele jednostek zdolnych do wykonywania rozkazów. Przy takim podejściu polecenia są wykonywane wtedy, gdy dostępne są argumenty niezbędne do ich realizacji (np. wyniki obliczeń dokonywane przez inne jednostki procesora). Procesor musi tak rozdzielać rozkazy, aby nie dochodziło do konfliktu argumentów. Przykładem architektury równoległej są procesory superskalarne. Współczesne procesory ogólnego przeznaczenia (np. Pentium II, Pentium III) posiadają taką właśnie architekturę.

3.7. Do czego służą i jak są wykonywane rozkazy skoku?

W programach komputerowych bardzo często zachodzi konieczność wielokrotnego wykonania pewnego ciągu rozkazów, pominięcia części programu oraz zmiany kolejności wykonywania operacji. W takich sytuacjach niezbędne są rozkazy modyfikujące licznik rozkazów (LR). Programowa zmiana wartości tego licznika wiąże się z zakłóceniem kolejności wykonywania poleceń. Rozkazy zmieniające zawartość licznika rozkazów, to rozkazy skoku. Mają one jeden argument. Jest nim adres komórki, do której chcemy przekazać sterowanie. W wyniku wykonania takiego rozkazu do rejestru LR, trafia właśnie ten adres. Z miejsca, które wskazuje zmodyfikowany licznik rozkazów, jest pobierany kolejny rozkaz.

3.8. Na czym polegają rozgałęzienia i w jaki sposób są realizowane?

Rozgałęzieniami, czyli rozkazami warunkowymi nazywamy te miejsca w programie, w których podejmuje się jakąś decyzję. Do realizacji rozgałęzień są potrzebne rozkazy zwane skokami warunkowymi. Podczas ich wykonywania procesor sprawdza, czy określony warunek (np. czy zachodzi równość $i = 7$) jest spełniony czy też nie. W zależności od wyniku, wykonuje się skok lub kontynuuje się wykonywanie programu od następnego rozkazu. Argumentami w rozkazie skoku są: sprawdzany warunek oraz adres kolejnego rozkazu.

3.9. Co to są pętle?

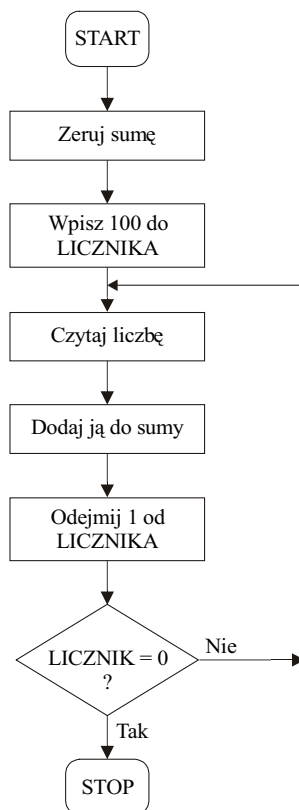
Pętle są swoistą odmianą skoku warunkowego. Służą one wykonywaniu sekwencji czynności określoną liczbę razy. Liczba powtórzeń wykonania zbioru rozkazów może być na stałe ustalona w czasie pisania programu lub może być uzależniona od wartości odpowiedniej zmiennej, modyfikowanej w trakcie wykonywania programu. Pętlę realizuje się poprzez użycie skoku warunkowego, wykonywanego aż do spełnienia określonego warunku.

Jeśli pętla ma być wykonana zawsze taką samą liczbę razy, warunkiem jej zakończenia jest osiągnięcie przez zmienną, będącą licznikiem zwiększanym przy każdym przejściu sekwencji, założonej wartości. Jeśli liczba wykonań pętli ma być uzależniona np. od stanu wejść procesora, wówczas warunkiem kończącym pętlę nie jest wartość licznika, lecz czynnik zewnętrzny.

Zdarza się, że należy zawiesić działanie programu. W tym celu można zastosować pętlę, która nie kończy się nigdy. Wówczas w skoku warunkowym należy umieścić warunek niemożliwy do spełnienia np. $1 = 2$.

3.10. Co to są i do czego służą znaczniki?

Znaczniki stanu, zwane także znacznikami warunków, są jednobitowymi układami pamiętającymi (zwanymi przerzutnikami), które zapamiętują wynik porównania dwóch argumentów lub cechę ostatnio wykonanej operacji arytmetycznej. W drugim przypadku stan jednego z przerzutników świadczy o tym, że wynik był: dodatni, ujemny lub zerowy. W procesorze są też znaczniki mówiące o tym, że podczas wykonywania operacji wystąpił błąd lub wynik operacji jest nieprawidłowy np. przekroczono dopuszczalny zakres liczb.



Rysunek 3.4. Algorytm programu sumowania stu liczb z wykorzystaniem pętli

Dla porównywania dwu argumentów często służy specjalny rozkaz, który nie zmieniając wartości argumentów wyznacza ich różnicę i ustawia odpowiedni znacznik tak, jakby wykonano rzeczywiste odejmowanie.

Znaczniki stanu mogą być badane przez rozkazy skoków warunkowych. Kod warunku występujący w tych rozkazach określa, o który znacznik lub o jaką ich kombinację w danym przypadku chodzi. Znaczniki potocznie nazywane są flagami (ang. *flags*).

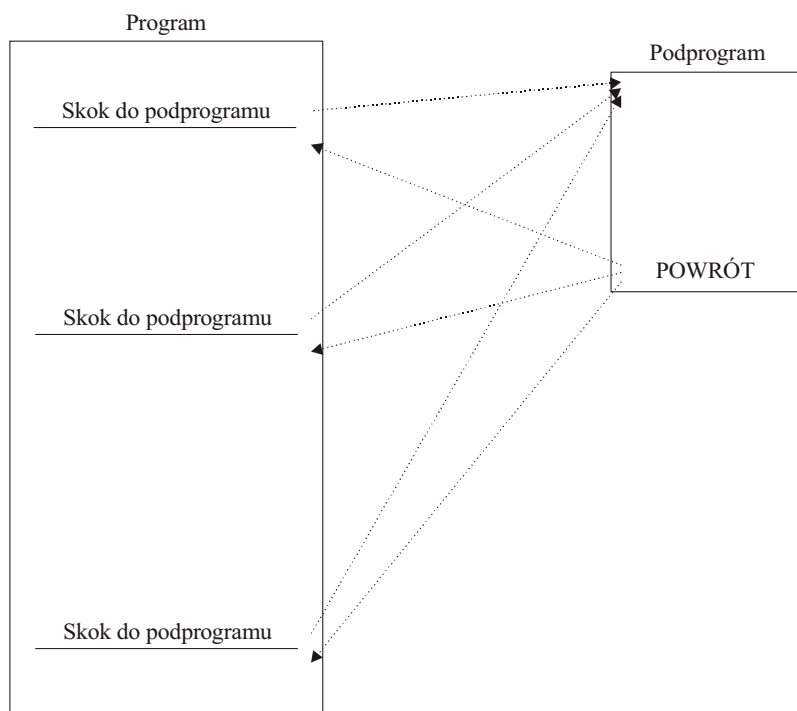
3.11. Co to jest podprogram i w jaki sposób się z niego korzysta?

Podprogramów używa się, jeśli identyczny ciąg rozkazów w programie jest często wykorzystywany w różnych miejscach tegoż programu. Przykładem może być wyprowadzanie specjalnie sformatowanych wyników obliczeń do drukarki. Każda chęć wysłania czegoś na to wyjście, musi być poprzedzona dodatkowymi obliczeniami. Za każdym razem jest jednak potrzebny ten sam ciąg rozkazów, który należy wstawić we wszystkie miejsca programu, w których następuje wyprowadzanie wyników. Wsta-

wianie wielokrotnie tego zbioru rozkazów jest rozrzutne z punktu widzenia rozmiaru programu i pracochłonne.

Można to zrobić prościej, korzystając z rozkazów skoku i rozkazu powrotu (rys. 3.5). Przed każdym wysłaniem wyniku do drukarki, należy wykonać skok do stałego miejsca, gdzie są umieszczone rozkazy konwersji i wykonać je. Ciąg rozkazów konwersji musi się kończyć specjalnym rozkazem powrotu, tak by po wysłaniu wyniku do drukarki program mógł kontynuować obliczenia. Aby taki powrót był możliwy, przed wykonaniem skoku niezbędne jest zapamiętanie aktualnej wartości licznika rozkazów. Do tego celu używa się stosu omówionego w pytaniu 3.12.

Fragment programu wykorzystywany w ten charakterystyczny sposób nosi nazwę podprogramu. Skok powodujący jego uaktywnienie oraz zapamiętanie licznika rozkazów nazywa się wywołaniem podprogramu. Powrót do miejsca, z którego nastąpiło wywołanie, nazywa się rozkazem powrotu z podprogramu.



Rysunek 3.5. Ilustracja wykorzystania podprogramu

3.12. Co to jest stos?

Stos jest to specjalnie zorganizowany fragment pamięci (w postaci jednowymiarowej tablicy) przeznaczony do przechowywania niewielkiej liczby danych. Dostęp do tej tablicy jest możliwy tylko w pewien charakterystyczny sposób. Dane na stosie można umieszczać lub zdejmować.

Umieszczanie danych przypomina układanie książek, jedną na drugiej. Kolejnego elementu, który chcemy umieścić na stosie, nie można wstawić między istniejące już na stosie elementy, lecz można go tylko położyć na wierzchołku. W wyniku umieszczania kolejnych elementów stos „rośnie”. Element ostatnio położony na stosie nazywa się wierzchołkiem stosu.

Pobieranie danych ze stosu przebiega w odwrotnej kolejności. Jedynym dostępnym elementem, który można pobrać jest wierzchołek stosu. Nie można „wyjąć” żadnych elementów ze środka stosu. W wyniku pobierania elementów stos „obniża się”. Po pobraniu wszystkich elementów stos jest pusty.

Działania na stosie są możliwe dzięki istnieniu specjalnego rejestru, zwanego wskaźnikiem stosu. Zawiera on zawsze adres aktualnego wierzchołka stosu. Rozkaz zapisu danej na stosie zwiększa wartość wskaźnika stosu, a następnie umieszcza ją w miejscu przez niego wskazywanym. Rozkaz pobierania informacji ze stosu odczytuje daną wskazywaną przez wskaźnik i zmniejsza jego wartość. W ten sposób zawsze znane jest położenie jedynie wierzchołka stosu.

3.13. Do czego służy stos?

Stos jest szczególnie przydatny do przechowywania adresów powrotu z podprogramów. Rozkazy wywołania podprogramu automatycznie przesyłają na stos aktualną wartość licznika rozkazów. Natomiast rozkazy powrotu z podprogramu (również automatycznie) pobierają ze stosu wartość licznika rozkazów i wykonują skok w tak odtworzone miejsce. Rozwiązanie to umożliwia wielopoziomowe wywoływanie podprogramów przez podprogramy.

W podobny sposób jest wykorzystywany stos przy obsłudze przerw (patrz pyt. 3.14). Przed przystąpieniem do wykonywania rozkazu podprogramu obsługi przerwania procesor zapisuje na stosie aktualne położenie w wykonywanym programie oraz stan procesora. Po zakończeniu obsługi przerwania stan rejestrów procesora jest odtwarzany i program może być kontynuowany.

3.14. Na czym polega przerwanie i jakie są rodzaje przerwania?

Przerwanie polega na wstrzymaniu aktualnie wykonywanego programu i automatycznym uruchomieniu jednego ze specjalnych podprogramów wykonujących czynności zwane obsługą przerwania.

Przyczynami przerwania mogą być błędy, spowodowane przez program takie, jak próba dzielenia przez zero, próba wykonania nielegalnego rozkazu lub odwołanie do zawartości nieistniejącej komórki pamięci. Takie sytuacje są wykrywane przez układy procesora, które zawieszają wówczas wykonywanie programu i uruchamiają podprogram, np. wysyłający komunikat o błędzie. Po zakończeniu podprogramu obsługi przerwania następuje wznowienie programu, który został przerwany. Wyjątkiem są sytuacje, w których oprogramowanie nadzorujące pracę komputera (system operacyjny – rozdz. 9) uzna, że dalsze jego wykonywanie nie ma sensu.

Oprócz przerwania mających swe źródło w samym programie, istnieją jeszcze przerwanie zewnętrzne. Mogą one być spowodowane np. naciśnięciem klawisza, zakończeniem lub błędem transmisji przez układ wejścia/wyjścia. Dzięki zastosowaniu przerwania do obsługi urządzeń wejścia/wyjścia, jest możliwe zwolnienie procesora z bezpośredniego obsługiwanie transmisji. Układy komunikacyjne wymieniają informacje między sobą, a procesor jest jedynie informowany o obecności odebranych już informacji. To rozwiązanie pozwala wykorzystać czas procesora na wykonanie innych zadań, podczas gdy wolne urządzenia transmisyjne same prowadzą transmisję.

3.15. Jak przebiega obsługa przerwania?

Przykładem obrazującym działanie systemu przerwania może być praca sekretarki, która pisze coś ważnego. W trakcie pisania dzwoni telefon, powodując przerwanie pracy i konieczność „obsługi” tego zdarzenia. Po zakończeniu rozmowy telefonicznej sekretarka powraca do pisania. Za chwilę przychodzi jej szef i prosi o zrobienie kawy. Czynność ta jest ważniejsza od pisania, więc musi zostać „obsłużona” wcześniej. Po zrobieniu kawy sekretarka wraca do swoich obowiązków.

W komputerze po zgłoszeniu sygnału przerwania na stos jest wysyłany adres rozkazu, od którego należy wznowić program po zakończeniu obsługi przerwania oraz inne informacje określające aktualny stan programu. Podprogram obsługi przerwania musi być zakończony rozkazem powrotu z przerwania. Działanie tego rozkazu jest podobne do rozkazu powrotu z podprogramu. Rozkaz powrotu z przerwania odtwarza dodatkowo stan programu, jaki był w chwili przyjęcia przerwania.

Podczas obsługi przerwania może zostać zgłoszone inne przerwanie. W zależności od tego, czy przerwanie obecnie obsługiwane jest ważniejsze (ma wyższy priorytet) czy ważniejsze jest przerwanie właśnie zgłaszane, procesor może ignorować sygnał zgłoszenia lub przystąpić do jego obsługi. Jeśli przerwanie zgłaszane w czasie obsługi

innego przerwania ma niższy priorytet, procesor nie przyjmuje go, aż do chwili zakończenia obsługi poprzedniego przerwania, o wyższym priorytecie. Jeśli natomiast nowo zgłaszane przerwanie ma wyższy priorytet, procesor zawiesza wykonywanie poprzedniego podprogramu obsługi i wykonuje podprogram nowego przerwania. Po zakończeniu jego obsługi jest kontynuowany podprogram obsługi poprzedniego przerwania.

3.16. Jak działa arytmometr?

Arytmometr wraz ze współpracującymi z nim rejestrami umożliwia wykonywanie jedynie podstawowych operacji arytmetycznych i logicznych. Zbiór tych operacji jest przeważnie uboższy, niż zbiór rozkazów arytmetycznych komputera. Zawiera on najczęściej:

- dodawanie
- odejmowanie
- zwiększanie lub zmniejszanie o jeden
- przesuwanie w prawo lub w lewo
- funkcje logiczne: NOT, AND, OR, EXOR (patrz pyt. 2.4).

Niemal w każdym komputerze zbiór ten jest inny. Zależy to głównie od sposobu rozwiązania układów sterowania arytmometrem i założonych parametrów użytkowych. Najmniejszy możliwy zbiór operacji podstawowych, za pomocą których można wykonać dowolne działanie, jest jeszcze mniej liczny niż podany powyżej. Jednak im więcej funkcji bezpośrednio wbudowanych w arytmometr, tym szybciej odbywa się przetwarzanie.

Z arytmometrem współpracują rejestry robocze, najczęściej o długości słowa procesora. Na poszczególne wejścia są podawane równolegle wszystkie bity obydwu argumentów. Wynik trafia do rejestru przyłączonego do wyjścia arytmometru. Oprócz rejestrów dostępnych z zewnątrz z arytmometrem mogą współpracować rejestry „niewidoczne” z zewnątrz. Dodatkowe układy pozwalają arytmometrowi wykonywać bardziej skomplikowane operacje, np. mnożenie liczb. Odbywa się to poprzez powtarzanie w odpowiedniej kolejności operacji elementarnych (w przypadku mnożenia będzie to ciąg dodawań i przesunięć).

3.17. Jakie rodzaje liczb można przedstawić w komputerze?

Wewnątrz komputera, ze względu na ograniczoną liczbę bitów, na których mogą być zapisane liczby, mamy do czynienia z podzbiorem liczb wymiernych. Ze względu na sposób zapisu wyróżniamy w tym podzbiore liczy całkowite.

Liczby całkowite przeważnie zapisuje się w naturalnym kodzie dwójkowym. Czasami, w celu ułatwienia przetwarzania liczb, stosuje się zapis w kodzie BCD (ang. *Binary Coded Decimal*) – każdą cyfrę liczby dziesiętnej koduje się osobno w naturalnym kodzie dwójkowym. Problem powstaje wówczas, gdy trzeba przedstawić liczbę całkowitą ze znakiem. W tym przypadku najczęściej stosuje się zapis w systemie znak-moduł lub tzw. uzupełnienie dwójkowe.

Do zapisu liczb wymiernych najczęściej wykorzystuje się zapis zmiennopozycyjny, który pozwala objąć znacznie większy zakres liczb niż systemy stałopozycyjne.

3.18. Na czym polega zapis w systemie znak-moduł?

Jest to zapis podobny do stosowanego w systemie dziesiętnym. Liczba składa się z dwóch części: znaku i wartości bezwzględnej. Wartość bezwzględna jest kodowana w zwykły sposób, jak każda liczba dwójkowa, natomiast znak określa się za pomocą oddzielnego bitu. Zwykle jest to lewy skrajny bit słowa. Jeśli bit ten ma wartość „0”, to liczba jest interpretowana jako dodatnia, natomiast wartość bitu „1” odpowiada liczbie ujemnej.

Liczby, które można przedstawić w ten sposób w słowie n -bitowym, należą do przedziału $[-(2^{n-1} - 1); +(2^{n-1} - 1)]$. Pewną niedogodnością zapisu w tym systemie jest niejednoznaczny sposób zapisu zera. W zapisie znak-moduł zero może mieć wartość dodatnią lub ujemną. Najważniejszą wadą tego systemu są trudności występujące podczas wykonywania działań arytmetycznych.

3.19. Co to jest uzupełnienie dwójkowe?

Uzupełnieniem dwójkowym liczby x ($x \neq 0$), zapisanej za pomocą n bitów, nazywa się liczbę $x_u = 2^n - x$. Jeśli weźmie się liczbę czterobitową $x = 0101$ ($x = 5$), to zgodnie z definicją, jej uzupełnienie dwójkowe ma wartość $x_u = 2^4 - 0101 = 1011$ ($x_u = 16 - 5 = 11$). Niech $x_u = y = 1011$. Wówczas $y_u = 2^4 - 1011 = 0101$ ($y_u = 16 - 11 = 5$). Otrzymuje się dość ciekawy wynik: $y_u = x$, czyli $(x_u)_u = x$. Operator „u” działa identycznie jak znana operacja w arytmetyce zwana zmianą znaku liczby, gdyż $(x_u)_u = x$ podobnie jak $-(-x) = x$. Wynika z tego, że uzupełnienie dwójkowe danej liczby, reprezentuje liczbę przeciwną. Warto również zauważyć, że ciąg bitów stanowiący uzupełnienie dwójkowe liczby dodatniej, czyli kod reprezentujący liczbę ujemną, ma jedynekę na najbardziej znaczącym bicie. Łatwo więc identyfikować liczby ujemne (rys. 3.6).

Liczba dziesiętna	Liczba U2			
	wagi (*)			
	(-8)	4	2	1
-8	1	0	0	0
-7	1	0	0	1
-6	1	0	1	0
5	1	0	1	1
-4	1	1	0	0
-3	1	1	0	1
-2	1	1	1	0
-1	1	1	1	1
+0	0	0	0	0
+1	0	0	0	1
+2	0	0	1	0
+3	0	0	1	1
+4	0	1	0	0
+5	1	1	0	1
+6	0	1	1	0
+7	0	1	1	1

Warto zaznaczyć, że wagi dla kolejnych bitów to -8, 4, 2, 1, w ogólności dla n bitów: $-2^{n-1}, 2^{n-2}, \dots, 2^0$

Rysunek 3.6. Liczby z przedziału [-8;7] zapisane w systemie uzupełnienia do dwóch

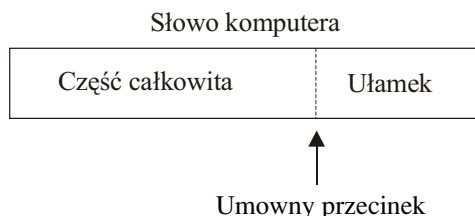
Zero w tym systemie jest jednoznacznie reprezentowane, a operacje dodawania i odejmowania są wykonywane identycznie dla liczb dodatnich i ujemnych. Odejmowanie dwóch liczb wykonuje się według zasady:

$$A-B = A+(-B) = A+(B)_u$$

Wszystkie możliwe liczby, które można w ten sposób zapisać w słowie n-bitowym należą do przedziału: $[-2^{n-1}; 2^{n-1}-1]$. Jest to przedział niesymetryczny np. za pomocą ośmiu bitów można zapisać liczby z zakresu od -128 do +127.

3.20. Na czym polega zapis stałopozycyjny i do czego służy?

Omówione powyżej sposoby umożliwiają zapis liczb całkowitych. Zapis stałopozycyjny umożliwia natomiast przedstawianie liczb zawierających część ułamkową, czyli liczb wymiernych. W tym celu, w słowie komputera ustala się umowną pozycję przecinka, oddzielającą część całkowitą liczby od części ułamkowej.



Rysunek 3.7. Format liczby stałopozycyjnej

Ponieważ sposób zapisu liczb całkowitych omówiono już wcześniej, należy zastanowić się nad sposobem zapisu części ułamkowej liczb. W tym przypadku również stosuje się zapis w systemie pozycyjnym o podstawie 2. Tu jednak, do wyznaczania wartości poszczególnych liczb, stosuje się ujemne wykładniki potęg.

Jeśli na zapis ułamka właściwego przeznaczony jest k bitów, to jego wartość określa wzór:

$$\sum_{i=1}^k a_i \cdot 2^{-i} = a_1 \cdot 2^{-1} + a_2 \cdot 2^{-2} + \dots + a_k \cdot 2^{-k}$$

w którym współczynniki a_i oznaczają i -te cyfry dwójkowe, licząc od przecinka w prawo.

Zaletą zapisu stałopozycyjnego jest jego prostota i elastyczność. Wynika to z tego, że pozycja przecinka w liczbie zależy od umowy. Zakres liczb, które przedstawia się w ten sposób, można zmieniać w zależności od potrzeb. System ten ma również wiele wad. Przede wszystkim, istotnie ogranicza możliwy zakres liczb, ze względu na konieczność wydzielenia w słowie części ułamkowej. Nie pozwala on również na zbyt dużą dokładność obliczeń, gdyż wynik musi być zawsze tak zaokrąglony, by mieścił się w słowie. Przykładowo, ułamek właściwy mający skończone rozwinięcie w systemie dziesiętnym (np. 0,1) nie ma takiego w systemie dwójkowym. Czytelnik może sprawdzić, że:

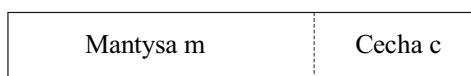
$$0,1_{(10)} = 0,0 \underbrace{0011\ 0011\ 0011}_{\text{okres}}_{(2)}$$

Arytmetyka liczb, przedstawionych w ten sposób, jest z reguły realizowana programowo, tzn. nie buduje się arytmetrów wykonujących operacje na liczbach stałopozycyjnych. Wszystkimi szczegółami związanymi z przeskalowaniem argumentów i wyników operacji, musi się zajmować programista. Zmniejsza to oczywiście tempo ich wykonywania.

3.21. Jaka postać mają liczby zmiennopozycyjne i dlaczego się je stosuje?

Liczby zmiennopozycyjne składają się z dwóch części:

- części ułamkowej zwanej mantysą (m)
- części całkowitej zwanej cechą (c).



Rysunek 3.8. Format liczby zmiennopozycyjnej

Podział liczby na dwie części jest podobny jak w zapisie stałopozycyjnym. Podobieństwo to jest tylko pozorne. W przypadku zapisu stałopozycyjnego, wartość liczby jest sumą obu części, a ich długość jest umowna. W przypadku liczb zmiennopozycyjnych, długość mantysy i cechy jest stała, a wartość liczby określa wzór:

$$x = m \cdot 2^c$$

Taki sposób zapisu liczb wymiernych pozwala objąć bardzo duży zakres liczb, kosztem nierównomiernego ich rozmieszczenia na osi liczbowej. Dla przykładu można podać zakresy liczb typu integer (stałopozycyjne) i real (zmiennopozycyjne) w Pascalu:

- Integer -32768...32767
- Real 2,9e⁻³⁹ ... 1,7e³⁸

W wielu zastosowaniach, np. w fizyce, gdzie występuje olbrzymia rozpiętość danych liczbowych (np. masa elektronu i prędkość światła), jest to jednak konieczne.

Mantysę liczby zmiennopozycyjnej przedstawia się najczęściej w systemie znak-moduł. Długość mantysy ma wpływ na dokładność liczb. Cechę natomiast przeważnie przedstawia się w systemie uzupełnieniowym do 2.

3.22. Co to jest koprocesor?

Koprocesory są wyspecjalizowanymi układami współpracującymi z procesorem. Najczęściej spotyka się koprocesory przyspieszające działania na liczbach zmiennoprzecinkowych. Są to koprocesory arytmetyczne. Często używa się również koprocesorów graficznych, zwanych też akceleratorami graficznymi.

Dawniej koprocesory arytmetyczne były budowane jako dodatkowe układy scalone i umieszczane w płycie głównej komputera. Począwszy od i486 koprocesor aryt-

metryczny i procesor są umieszczone na jednej strukturze. Wyjątkiem jest procesor i486SX, w którym działanie koprocatora zostało zablokowane.

Akceleratory graficzne są wyspecjalizowanymi układami przetwarzającymi grafikę. Znalazły one zastosowanie zwłaszcza przy pracach z grafiką przestrzenną. Koprocessor graficzny może sprzętowo dokonywać obrotów i innych transformacji obiektów, bez obciążania złożonymi obliczeniami procesora głównego.

3.23. Jak działa sterowanie?

Sterowanie realizuje podstawowy algorytm działań procesora. Składa się on z następujących kroków (patrz pyt. 3.2):

- odczytanie z pamięci rozkazu wskazanego przez licznik rozkazów LR i umieszczenie go w rejestrze rozkazów RR
- zmiana zawartości LR tak, aby wskazywał następny rozkaz
- dekodowanie rozkazu, określenie jego typu i liczby argumentów
- wyznaczenie adresów i pobranie niezbędnych argumentów
- wykonanie wymaganych czynności, wskazanych w treści rozkazu i zapamiętanie wyniku
- przejście do pobrania kolejnego rozkazu.

Każdy z tych kroków składa się z wielu czynności elementarnych. Na przykład pobranie danej z pamięci wymaga:

- otwarcia bramek umożliwiających przesłanie adresu do rejestru adresowego pamięci
- wysłanie sygnału odczytu
- otwarcie bramek umożliwiających przesłanie danych z rejestru buforowego pamięci do miejsca przeznaczenia.

Wykonanie operacji arytmetycznej wymaga:

- otwarcia odpowiednich dróg łączących rejestry zawierające argumenty z wejściami arytmometru
- ustawienia odpowiednich sygnałów sterujących działaniem arytmometru
- zbadania, czy operacja zakończyła się
- otwarcia drogi umożliwiającej przekazanie wyniku
- ustawienia odpowiednich znaczników stanu, odzwierciedlających cechy wyniku.

Układy sterowania muszą także reagować na sygnały przerwania i odpowiednio modyfikować cykl pracy procesora.

3.24. Od czego zależy szybkość działania procesora?

Szybkość działania procesora zależy od jego konstrukcji. Główne cechy mające wpływ na szybkość procesora to:

- szerokość magistrali danych
- wewnętrzna częstotliwość taktowania procesora (liczba cykli wykonywanych przez procesor w ciągu sekundy)
- zewnętrzna częstotliwość taktowania (częstotliwość magistrali z jaką procesor uzyskuje dostęp do pamięci i innych układów)
- wielkość pamięci podręcznej pierwszego poziomu (patrz pyt. 4.10)
- wielkość pamięci podręcznej drugiego poziomu (patrz pyt. 4.10).

Wymienione parametry wynikają z modelu procesora i jego wersji. Obecnie produkowane procesory posiadają 64-bitową magistralę danych, wewnętrzną częstotliwość taktowania rzędu 1 GHz, zewnętrzną częstotliwość 133 MHz, wielkość pamięci podręcznej pierwszego poziomu 64 kB, rozmiar pamięci podręcznej drugiego poziomu do 1024 kB.

Szybkość procesora jest wyrażana np. w MIPS-ach (ang. *Milion Instructions Per Second*) lub GFLOPS-ach (ang. *Giga Floating Operations Per Second*). Pomiaru tych parametrów dokonuje się specjalnymi programami, reprezentującymi „mieszanki obliczeniowe” dla danej klasy zastosowań i wówczas ocenia szybkość komputera.

3.25. Do jakiej wartości można zwiększać częstotliwość taktowania procesora?

Zwiększaniu częstotliwości taktowania procesora towarzyszy zwiększanie mocy wydzielanej z procesora. Aby zwiększyć częstotliwość taktowania, producenci obniżają napięcie zasilania układów oraz wytwarzają struktury z tranzystorami o coraz mniejszych wymiarach. Procesor 486 był zbudowany z tranzystorów wielkości 1 μm , procesor Pentium 0,6-0,35 μm , a procesor Pentium II 0,25 μm . Fizyczna granica wielkości pojedynczego tranzystora jest rzędu 0,05 μm i prawdopodobnie zostanie osiągnięta już około 2005 roku. Przy tej wielkości mogą wystąpić problemy na poziomie struktury atomowej.

Kolejne wersje procesorów są złożone z coraz większej liczby tranzystorów. Procesor 486 miał ponad 1mln tranzystorów, natomiast Pentium II posiada 6,5 mln tranzystorów.

Drugim ograniczeniem są względy ekonomiczne. Wraz ze wzrostem upakowania elementów na strukturze, rośnie koszt wytworzenia procesora.

3.26. Co to jest mikroprocesor i kiedy go stworzono?

Mikroprocesor jest cyfrowym układem scalonym o dużym stopniu scalenia i funkcjonalnej uniwersalności. Rodzaj zadania, jakie wykonuje taki układ, nie jest bowiem zdeterminowany konfiguracją połączeń, lecz jest określony przez program zawarty w pamięci. Cechą charakterystyczną urządzeń cyfrowych, zbudowanych z mikroprocesorów, jest łatwa zmiana rodzaju wykonywanego zadania. Polega ona jedynie na wymianie programu sterującego. Pod tym względem mikroprocesor jest uniwersalnym układem cyfrowym.

Za początek ery mikroprocesorów przyjmuje się rok 1971. Wówczas został wyprodukowany pierwszy, czterobitowy mikroprocesor – Intel 4004. Układ ten, o strukturze typowej dla kalkulatora, nadawał się już do wykonywania prostych urządzeń. Był zbudowany z ok. 2300 tranzystorów i taktowany częstotliwością 100 kHz. W następnym roku Intel wprowadził na rynek mikroprocesor ośmiobitowy (8008). Mikroprocesory ośmiobitowe są nadal stosowane w prostych układach sterujących.

3.27. W jakich okolicznościach doszło do skonstruowania mikroprocesora?

Po zaprojektowaniu i wykonaniu urządzenia funkcje przez nie wykonywane nie mogą być zmieniane. Zaprojektowany układ może być wykorzystywany tylko zgodnie ze swoim przeznaczeniem i np. układ elektroniczny sterujący pralką automatyczną nie nadaje się do regulacji zapłonu w samochodzie. Istnieje jednak układ cyfrowy nie mający takich ograniczeń – mikroprocesor. Rodzaj zastosowania urządzenia zbudowanego z mikroprocesora nie jest zdeterminowany przez sztywną konstrukcję połączeń, lecz jest określony przez program zawarty w pamięci współpracującej z mikroprocesorem. Zmiana rodzaju wykonywanego zadania sprowadza się zatem do zmiany programu. W tym znaczeniu mikroprocesor jest uniwersalnym (programowalnym) układem cyfrowym. Za pomocą tego uniwersalnego układu można skonstruować specjalizowane urządzenia wykonujące przez cały czas tę samą funkcję (np. mikroprocesorowy sterownik kuchenki mikrofalowej) bądź urządzenia uniwersalne jak np. komputer do różnych zastosowań.

Można powiedzieć, że rozwój mikroprocesorów rozpoczął się prawie przypadkowo. Firma Display Terminals Corporation, która produkowała „inteligentne” terminale i małe systemy mikrokomputerowe, zamierzała wprowadzić na rynek prosty oraz tani komputer i zamówiła w roku 1969 w firmach Intel i Texas Instruments układ scalony LSI przeznaczony do sterowania monitorem ekranowym. Firma Intel zbudowała taki układ, lecz był on dziesięć razy wolniejszy od zamawianego. Zleceniodawca nie przyjął więc zamówionego wyrobu. Firma Intel miała do wyboru dwie drogi – odstąpienie wyrobu na półki i poniesienie kosztów opracowania wyrobu bądź wprowadzenia

dzenie wyrobu na rynek do sondażowej sprzedaży. Wybrała drugi wariant. W roku 1972 rozpoczęto sprzedaż układu nazwanego Intel 8008.

3.28. Co powoduje dużą popularność mikroprocesorów?

Przyczynami dużej popularności mikroprocesorów są następujące właściwości:

- niska cena przy dużych możliwościach funkcjonalnych
- duża niezawodność
- małe wymiary
- mały pobór mocy.

Tak duża popularność mikroprocesorów wynika głównie z faktu, że w jednym układzie scalonym można umieścić prawie każdy układ cyfrowy. Stało się to możliwe dzięki umieszczeniu układu procesora, pamięci programu, pamięci RAM, układów wejścia/wyjścia na jednej strukturze. Mikroprocesor stał się układem bazowym niemal każdego urządzenia elektronicznego. Powszechność stosowania układów mikroprocesorowych oraz ich masowa produkcja przyczyniły się do spadku ceny, a tym samym wzrosła ich atrakcyjność w stosunku do tradycyjnych układów cyfrowych.

Zwiększający się stopień scalenia oraz wzrost złożoności układów scalonych, doprowadziły do powstania coraz bardziej wyspecjalizowanych układów, o coraz krótszych seriach. Jak wiadomo, produkcja układów scalonych jest opłacalna przy długich seriach. Wprowadzenie na rynek mikroprocesora, jako układu uniwersalnego, usatysfakcjonowało zarówno producentów jak i użytkowników.

3.29. Co to są procesory CISC, RISC?

Procesory ze złożoną listą instrukcji są określane jako CISC (ang. *Complex Instruction Set Computer*). Charakteryzują się:

- dużą liczbą rozkazów o różnych długościach
- dużą liczbą trybów adresowania
- łatwością tworzenia oprogramowania.

Natomiast procesory ze zredukowaną, uproszczoną listą rozkazów – RISC (ang. *Reduced Instruction Set Computer*) charakteryzują się:

- niewielką liczbą rozkazów
- małą liczbą trybów adresowania
- prostą i szybką układową jednostką sterującą
- trudnością w tworzeniu oprogramowania.

3.30. Dlaczego wprowadzono procesory RISC?

Projektanci procesorów, w celu zwiększenia ich wydajności, dodawali nowe bloki oraz zwiększali listę rozkazów. Coraz bogatsza lista rozkazów ułatwiała pracę programistom, a także pozwalała skrócić program. Powodowała jednak wzrost złożoności jednostki sterującej procesora i spowolnienie jej pracy.

Badania statystyczne prowadzone nad stopniem wykorzystania listy rozkazów procesorów CISC wykazały, że tylko niewielka ich część jest wykorzystywana przez programistów. Najczęściej wykorzystywali oni skomplikowane rozkazy, a więc najdłużej wykonujące się. Podobnie przedstawiały się wyniki badań nad wykorzystaniem trybów adresowania.

Opierając się na tych wynikach, zakwestionowano potrzebę zwiększania listy rozkazów procesorów. Efektem było pojawienie się procesorów o zredukowanej liście rozkazów (RISC). W procesorach RISC stosuje się rozkazy o jednakowej długości, łatwe do dekodowania, wykonywane w pojedynczym cyklu zegara. W procesorach tych używa się przetwarzania potokowego.

Ograniczona lista rozkazów wiąże się ze znaczną złożonością kompilatorów, (patrz pyt. 9.37) o dużych możliwościach optymalizacji programu wynikowego. Program wynikowy dla procesora RISC jest zwykle dłuższy niż dla procesora CISC, jednak procesor RISC i tak wykonuje go szybciej.

3.31. Gdzie w praktyce spotykamy procesory o architekturze RISC?

Procesory RISC są stosowane w wydajnych serwerach sieciowych. Przykładami mogą być procesory: Alpha, Sparc, PowerPC.

Procesory Alpha firmy Compaq (dawny Digital) są procesorami RISC. Przykładem jest 21264, który jest procesorem 64-bitowym. Wszystkie rozkazy mają długość 32-bitów. Procesor jest taktowany z częstotliwością 466-733 MHz.

Procesor Ultra Sparc II produkowany przez Sun Microsystems jest 64-bitowym procesorem RISC. Pracuje z częstotliwością taktowania 250-400 MHz.

Struktura PowerPC została przyjęta przez koncerny IBM, Apple i Motorola. PowerPC 750 jest procesorem 64-bitowym. W jednym cyklu zegara może być wykonywanych równolegle sześć rozkazów. Procesor ten pracuje z częstotliwością taktowania 200-366 MHz.

3.32. Jakie procesory stosuje się do budowy współczesnych komputerów PC?

W komputerach powszechnego użytku najczęściej spotkać można procesory trzech firm: Intel, AMD i Cyrix. Wśród nich dominują:

- Intel Celeron
- Intel Pentium II
- Intel Pentium III,
- Intel Pentium 4
- AMD Athlon
- AMD Duron
- VIA Cyrix III.

Trzech wiodących producentów prowadzi ze sobą zawziętą rywalizację, czego efektami są prezentowane co kilka tygodni procesory o coraz to wyższej wydajności. Częstotliwość taktowania procesora jest często uważana za wyznacznik wydajności procesora, jednak nie zawsze tak jest.

Ze względu na bardzo duże tempo rozwoju procesorów trudne jest operowanie konkretnymi częstotliwościami. Na początku roku 2001 częstotliwości procesorów sięgnęła 1,5 GHz. W pewnym przybliżeniu sprawdza się prawo Moore'a, które mówi że wydajność komputerów podwaja się co osiemnaście miesięcy.

W erze rozwoju usług multimedialnych i internetowych nowe procesory są wyposażane w dodatkowe instrukcje przyspieszające przetwarzanie obrazów i dźwięku.

Obecnie wytwarzane procesory posiadają dwupoziomą pamięć podręczną.

4. Pamięć

4.1. Po co jest pamięć i co się przechowuje w pamięci?

Pamięć służy przechowywaniu informacji i jest niezbędnym elementem komputera. W komputerze używa się pamięci ulotnych RAM (ang. *Random Access Memory*), pamięci ROM (ang. *Read Only Memory*) oraz pamięci masowych.

Pamięci RAM i ROM są wykonywane w postaci półprzewodnikowych układów scalonych. Aby informacje w pamięci RAM mogły być przechowywane, musi być ona zasilana. Po odłączeniu napięcia dane zostają utracone. Inaczej wygląda sytuacja w przypadku pamięci ROM. W pamięci tej informacje mogą być przechowywane również po odłączeniu zasilania. Pamięci RAM i ROM służą do przechowywania niewielkich ilości danych.

Zawartość pamięci RAM może być dowolnie modyfikowana (zapisywana i odczytywana). Pamięć RAM jest używana jako pamięć operacyjna, w której przechowywane są kopie aktualnie wykonywanych programów, dane potrzebne programom oraz wyniki ich działania. Należy podkreślić, że pamięć RAM nazywa się często pamięcią o dostępie natychmiastowym (bezpośrednim, swobodnym) z tego powodu, że czas dostępu do informacji w niej zawartej nie zależy od lokalizacji tej informacji. Inaczej jest np. w pamięciach dyskowych i taśmowych. Nie należy więc niefortunnie i zbyt dosłownie tłumaczyć słowa „random”. Zdarza się niektórym tłumaczenie – „pamięć o dostępie losowym”! To nieporozumienie. O pamięci RAM często mówi się, że jest zapisywalno-odczytywalna.

Pamięć ROM jest przeznaczona tylko do odczytu (jest zapisywana poza komputerem w specjalnych programatorach). Pamięć ROM jest najczęściej wykorzystywana do przechowywania programów niezbędnych do uruchomienia komputera (np. BIOS – patrz pyt. 5.8). W najnowszych komputerach zawartość tej pamięci można zmieniać przy użyciu specjalnego programu.

Przechowywaniu programów i dużych (masowych) ilości informacji służą pamięci masowe. Do pamięci masowych należą dyski twarde, dyskietki, dyski optyczne i magnetoptyczne oraz pamięci taśmowe.

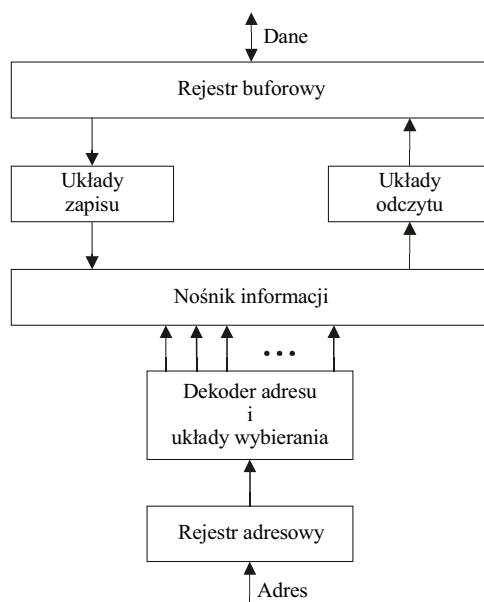
4.2. Jak jest zbudowana pamięć RAM?

Uproszczony schemat typowej pamięci komputerowej przedstawia rysunek 4.1.

Centralnym elementem tego schematu jest prostokąt symbolizujący podstawowy zespół funkcjonalny pamięci, a mianowicie nośnik informacji. Jest to właściwy maga-

zyn, w którym przechowuje się informacje. W pamięciach DRAM (ang. *Dynamic RAM*) są to kondensatory, w pamięciach SRAM (ang. *Static RAM*) – przerzutniki, w dyskach magnetycznych – zespół talerzy o specjalnych właściwościach.

Zapisanie informacji do pamięci odbywa się za pomocą układów zapisu, przekształcających sygnały sterujące i dane na sygnał zmieniający stan nośnika. Układy odczytu identyfikują stan nośnika i generują odpowiednie sygnały sterujące i sygnały danych. Lokalizację odpowiedniego miejsca na nośniku zapewniają układy wybierania, współpracujące z dekodерem adresu.



Rysunek 4.1. Struktura wewnętrzna pamięci komputerowej

4.3. Jakie nośniki informacji stosuje się w pamięciach komputerów?

Rodzaj nośnika jest ściśle zależny od rodzaju pamięci. W pamięciach ROM i RAM są to elementy elektroniczne (wypalona maska połączeń, tranzystory MOS, przerzutniki, kondensatory). Do budowy pamięci masowych używa się materiałów o specjalnych właściwościach magnetycznych lub optycznych.

4.4. Co to znaczy, że pamięć jest ulotna?

Określenie „pamięć ulotna” oznacza, że po odłączeniu zasilania informacje zostają bezpowrotnie utracone. Jest to spowodowane budową pamięci oraz sposobem przechowywania danych. Pojęcie pamięci ulotnej odnosi się do pamięci RAM, gdzie elementami przechowującymi są przerzutniki lub miniaturowe kondensatory. Elementy te, aby mogły przechowywać informacje, wymagają zewnętrznego zasilania.

4.5. Czy informacje zapisane w pamięci można przechowywać dowolnie długo?

Informacji nie można przechowywać dowolnie długo w pamięciach ulotnych, chyba że cały czas układ pamięci będzie zasilany, co jest trudne do zrealizowania. W pamięciach RAM informacja jest przechowywana w przerzutniku lub miniaturowym kondensatorze, które wymagają zasilania.

W pamięci ROM czas przechowywania danych jest taki, jak czas sprawności struktury układu. W tej pamięci informacja jest zapisana w postaci trwałych zmian, wykonanych na etapie produkcji. W przypadku pamięci reprogramowalnych EPROM (ang. *Erasable Programmable ROM*) lub EEPROM (ang. *Electrically EPROM*) czas ten jest ograniczony do około 10 lat.

Podobnie wygląda kwestia trwałości zapisu pamięci masowych. Producenci „miękkich” nośników magnetycznych (dyskiety i taśmy) nie podają danych dotyczących trwałości zapisu. Asekuracyjnie zaleca się aby zapis nie był starszy niż dwa lata lub by liczba cykli zapisu nie była większa niż 300. Nośniki optyczne CD-R i CD-RW, według danych producentów, mają „czas życia” rzędu kilkudziesięciu lat (10–30). Ograniczony czas przechowywania informacji mają także dyski magnetoptyczne (szacuje się go również na kilkadziesiąt lat).

4.6. Czy zawartość pamięci można zmieniać?

Zawartość większości rodzajów pamięci można zmieniać. Należą do nich pamięci RAM, EPROM, EEPROM, Flash, a także pamięci masowe magnetyczne, magnetoptyczne oraz niektóre optyczne.

Zmiana zawartości pamięci SRAM polega na zmianie stanu przerzutnika. Modyfikacja zawartości komórki pamięci DRAM polega na naładowaniu lub rozładowaniu miniaturowego kondensatora.

Zawartość pamięci EPROM może być zmieniana jedynie poprzez skasowanie całej zawartości pamięci promieniami ultrafioletowymi, a następnie zapisanie nowych informacji. W pamięciach EEPROM i Flash zapisywanie i kasowanie odbywa się sy-

gnąłem elektrycznym. W pamięci EEPROM istnieje możliwość selektywnego kasowania i modyfikowania dowolnej komórki pamięci, natomiast w pamięciach Flash nie ma takiej możliwości.

Zawartość pamięci magnetycznych jest zmieniana poprzez odpowiednie namagnesowanie fragmentu nośnika.

Zmiana zawartości jest również możliwa w przypadku nośników magnetooptycznych. Dokonuje się tego poprzez podgrzanie wybranego punktu nośnika powyżej temperatury Curie (przy użyciu lasera) i odwrócenie biegunów magnetycznych (za pomocą zewnętrznego pola magnetycznego). Zapis na nośnikach optycznych jest możliwy przy użyciu technologii CD-RW (ang. *Compact Disc ReWritable*). Zapis na płycie CD-RW jest możliwy dzięki zastosowaniu lasera o sterowanej mocy i nośnika o odmiennych właściwościach fizycznych. Laser o zmiennej mocy, przy odpowiednim składzie chemicznym nośnika, umożliwia przywracanie stanu początkowego płyty, co pozwala na ponowny zapis.

Istnieją jednak pamięci przeznaczone do jednokrotnego zapisu i wielokrotnego odczytu. Do pamięci takich należą pamięci PROM, dyski CD-R (ang. *Compact Disc Recordable*). Podczas zapisu w nośnikach tych dokonywane są nieodwracalne zmiany, które uniemożliwiają ponowny zapis. W pamięciach PROM programowanie polega na przepaleniu niektórych połączeń matrycy, a w dyskach CD-R na wypaleniu zagłębień w powierzchni dysku.

4.7. Jak klasyfikuje się pamięci komputerów?

Wstępnego podziału dokonano już w odpowiedziach na poprzednie pytania. Ze względu na zastosowany typ nośnika pamięci można podzielić na: pamięci półprzewodnikowe, magnetyczne i optyczne. Z punktu widzenia trwałości przechowywanych informacji i możliwości ich zmiany wyróżnia się: pamięci ulotne i nieulotne oraz pamięci stałe i pamięci, których zawartość można zmieniać.

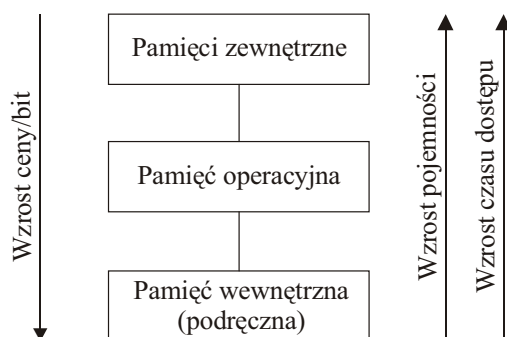
Istnieje jeszcze podział pamięci ze względu na sposób dostępu do informacji. Na podstawie tego kryterium wyróżnia się: pamięci o dostępie natychmiastowym (swobodnym, bezpośrednim), pamięci o dostępie sekwencyjnym oraz pamięci skojarzeniowe.

W pamięciach o dostępie natychmiastowym czas, jaki upływa od podania adresu do uzyskania informacji, jest dla danej pamięci zawsze taki sam i nie zależy od adresu, czyli lokalizacji informacji.

W pamięciach z dostępem sekwencyjnym czas potrzebny na dotarcie do informacji zależy nie tylko od aktualnej lokalizacji informacji, ale od lokalizacji używanej w poprzedniej operacji. Typową pamięcią sekwencyjną jest magnetyczna pamięć taśmowa.

W pamięciach skojarzeniowych (asocjacyjnych) komórki adresowane są poprzez swoją zawartość, a nie za pomocą określonego adresu. Lokalizacja informacji następuje na podstawie jej treści. Pamięci te nazywa się też pamięciami adresowanymi zawartością.

Pod względem funkcjonalnym pamięci komputerowe tworzą strukturę hierarchiczną. Cechą charakterystyczną tej struktury jest współzależność szybkości działania i pojemności pamięci (rys. 4.2).



Rysunek 4.2. Hierarchia pamięci komputerowych

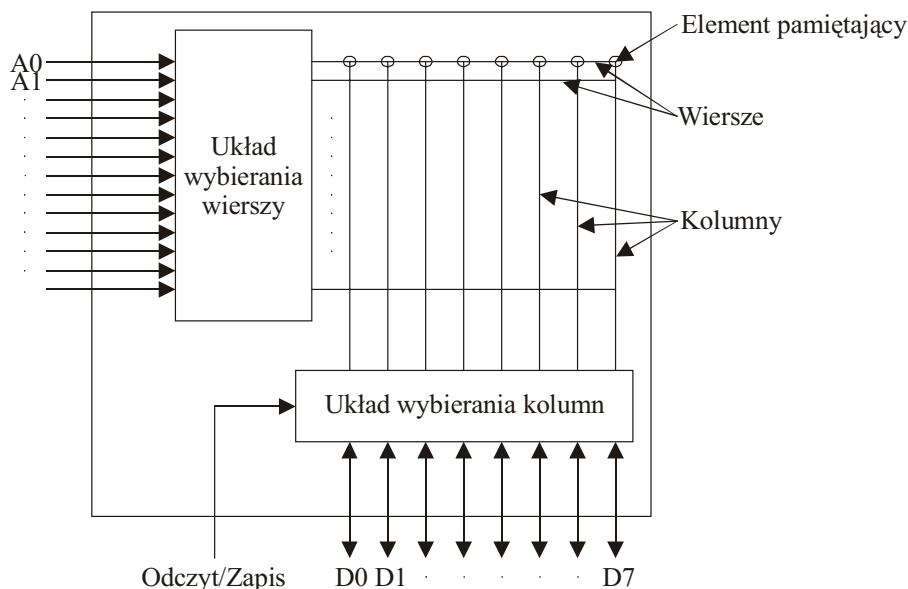
Najszybsze pamięci są najmniej pojemne. Wraz ze wzrostem pojemności szybkość maleje. Zależność ta wynika ze względów ekonomicznych. Pamięci najszybsze są wykonywane jako układy SRAM i używane jako pamięć podręczna procesora. Są one najdroższe w wykonaniu i mają niewielkie pojemności.

Kolejna w hierarchii jest pamięć operacyjna, tańsza w wykonaniu, wolniejsza w działaniu, mająca dużo większą pojemność. Układy pamięci operacyjnej są wykonywane jako pamięci DRAM.

Najwolniejsze i najbardziej pojemne są pamięci masowe, znajdujące się na trzecim poziomie omawianej hierarchii. Do tej kategorii pamięci zalicza się dyski twarde, elastyczne, pamięci optyczne, magnetoptyczne i pamięci taśmowe.

4.8. Jak jest zbudowana pamięć półprzewodnikowa?

Struktura logiczna pamięci półprzewodnikowych nie odbiega od ogólnej struktury pamięci komputerowych, przedstawionej w odpowiedzi do pytania 4.2. Wszystkie bloki funkcjonalne są wykonane w postaci układu scalonego o dużym i wielkim stopniu scalenia.



Rysunek 4.3. Organizacja pamięci półprzewodnikowej

Do zapamiętania pojedynczego bitu informacji potrzebny jest jeden element pamiętający. Wykonany jest w postaci przernutnika (dla pamięci SRAM) lub w postaci miniaturowego kondensatora (dla pamięci DRAM). Elementy te są uporządkowane w postaci matrycy (kolumny i wiersze). Dostęp do konkretnego elementu pamiętającego uzyskuje się przez wybór odpowiedniego wiersza i odczyt lub zapis zawartości odpowiedniej kolumny. Najczęściej buduje się pamięci, w których elementy pamiętające są pogrupowane po osiem. Odpowiada to wielkości bajtu. Wyboru grupy komórek (bajtu) dokonuje się poprzez wybranie odpowiedniego wiersza, a zapis i odczyt jest przeprowadzany dla ośmiu kolumn jednocześnie. Schemat przedstawiony na rysunku przedstawia taką właśnie organizację.

Aby było możliwe przechowywanie w pamięci słów o większej liczbie bitów (16, 32, 64), buduje się moduły zawierające większą liczbę układów scalonych. Na przykład aby uzyskać pamięć o długości pojedynczego słowa 32 bity, należy użyć czterech układów pamięci ośmiobitowych, których linie adresowe łączą się równolegle.

4.9. Czym się różni pamięć półprzewodnikowa dynamiczna od statycznej?

W komputerach stosuje się pamięci statyczne – SRAM (ang. *Static RAM*) oraz dynamiczne – DRAM (ang. *Dynamic RAM*).

Pamięć SRAM jest zbudowana z przerzutników dwustanowych (bistabilnych). Stan, w jakim znajduje się dany przerzutnik, odpowiada jednemu bitowi danych. Informacje w takiej pamięci są tak długo przechowywane, jak długo jest dołączone zasilanie. Budowa pamięci statycznej jest skomplikowana i wykonanie jej jest dość kosztowne. Do wytworzenia jednego elementu pamiętającego (jednego przerzutnika) potrzeba sześciu tranzystorów. Pamięci statyczne są stosunkowo drogie, ale za to szybkie. Są one wykorzystywane głównie jako pamięci podręczne procesorów. Ze względu na duże koszty wytworzenia, nie produkuje się pamięci SRAM o dużych pojemnościach.

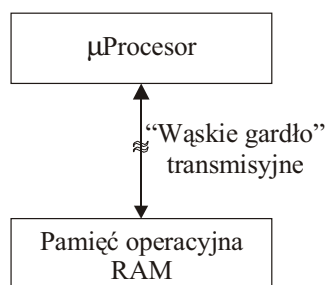
Element pamięci DRAM jest zbudowany z małego kondensatora. To rozwiązanie jest prostsze i tańsze w budowie. Zastosowanie kondensatora jako elementu pamiętającego zobowiązuje jednak projektanta-technologa pamięci do zapewnienia odświeżania jego zawartości. Wynika to z tego, że kondensatory posiadają pewną upływność i po pewnym czasie ich zawartość jest „zapominana”. Częstotliwość odświeżania musi być tak dobrana, aby przed kolejnym cyklem odświeżania zawartość komórki nie została utracona. Konieczność odświeżania spowalnia pracę pamięci.

Pamięci dynamiczne są używane do budowy pamięci operacyjnych komputerów. Są to pamięci o dużych pojemnościach. Można wyróżnić kilka odmian pamięci DRAM: FPM (ang. *Fast Page Mode*), EDO (ang. *Extended Data Out*), SDRAM (ang. *Synchronous DRAM*), RDRAM (ang. *Rambus DRAM*), DRDRAM (ang. *Direct Rambus DRAM*), SLDRAM (ang. *SyncLink DRAM*). Czas dostępu tych pamięci zmienia się od 70 ns (FPM) do 10 ns (SDRAM). Nowe odmiany pamięci DRAM mogą pracować z prędkościami taktowania 400 MHz.

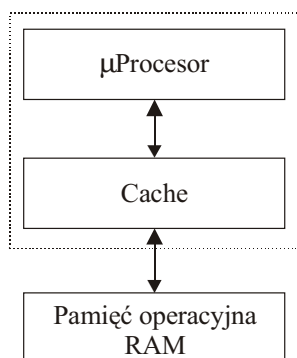
4.10. Co to jest pamięć podręczna cache i dlaczego się ją stosuje?

Pamięć podręczna cache jest zespołem rejestrów pełniących rolę bufora między procesorem a pamięcią operacyjną. Cache jest wykonana jako pamięć statyczna. Jest ona szybka, ale i kosztowna. Ze względów ekonomicznych, pamięć cache ma niewielkie pojemności (rzędu 512 kB). Uzasadnieniem stosowania pamięci podręcznej, jako bufora pamięci operacyjnej, jest znacznie mniejsza prędkość tych pamięci w stosunku do możliwości mikroprocesora (rys. 4.4).

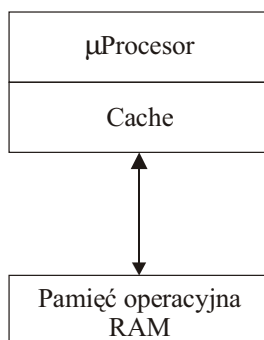
a)



b)



c)



Rysunek 4.4. Komunikacja procesora z pamięcią

- a) bez cache,
- b) z cache oddzielnym od procesora,
- c) z cache wbudowanym w procesor.

Działanie pamięci cache opiera się na właściwości programu, zwanej zasadą lokalności. Zasada ta polega na tym, że podczas wykonywania programów, odwołania do pamięci nie są losowe i dają się przewidzieć. Programy wykorzystują w niewielkich odstępach czasu, te same dane lub te same rozkazy (np. podczas wykonywania pętli). Zadaniem pamięci cache jest zapamiętywanie ostatnio wykorzystywanych danych i rozkazów. Procesor odwołując się do pamięci operacyjnej, najpierw sprawdza czy interesujące go informacje nie mają kopii w pamięci podręcznej. Jeśli szukana informacja zostanie znaleziona w pamięci cache, jest z niej pobierana i wykorzystywana. Jeśli pamięć podręczna nie zawiera poszukiwanej informacji, jest ona pobierana z pamięci operacyjnej, a jej kopia umieszczana w pamięci cache.

Dawniej pamięć podręczna była umieszczana na płycie głównej komputerów w postaci osobnych układów pamięci SRAM. Obecnie pamięć cache jest umieszczana z procesorem w jednej scalonej strukturze. Począwszy od Pentium II w strukturze procesora umieszcza się dwupoziomowy cache. Pamięć podręczna pierwszego poziomu (L1) jest taktowana częstotliwością procesora. Cache drugiego poziomu (L2) jest taktowany połową częstotliwości zegara procesora (w najnowszych również z częstotliwością zegara). Pojemność pamięci L1 w Pentium III wynosi 32 kB, natomiast L2 – 512 kB.

4.11. Co to jest i dlaczego stosujemy pamięć wirtualną?

Pamięć wirtualna jest plikiem stworzonym na dysku komputera. Pamięć ta jest rozszerzeniem pamięci rzeczywistej komputera. W chwili gdy pojemność pamięci operacyjnej jest niewystarczająca, część danych jest przenoszona do pamięci wirtualnej. W pamięci rzeczywistej pozostają tylko te programy lub części programów, które są niezbędne. Zwolniona w ten sposób pamięć rzeczywista pozwala na dalsze wykonywanie programu. Korzystanie z pamięci wirtualnej, ze względu na małe prędkości dysków, powoduje znaczne spowolnienie wykonywanego programu. Jeśli komputer zbyt często korzysta z pamięci wirtualnej, celowe jest rozszerzenie pamięci rzeczywistej, co pozwoli przyspieszyć wykonywanie programów, które dotychczas musiały korzystać z pamięci wirtualnej.

4.12. Co to jest ramdysk i po się go stosuje?

Ramdysk jest pewnego rodzaju „dyskiem”, umieszczonym w pamięci operacyjnej komputera. Ramdysk wytwarza się w pamięci programowo. Charakteryzuje się dużą szybkością i niewielką pojemnością, ograniczoną pojemnością pamięci RAM.

Dysku tego rodzaju używa się np. podczas instalowania systemu operacyjnego, gdy nie można korzystać jeszcze z dysku twardego. Instalacja jest inicjowana z dyskietki, tworzony jest dysk w pamięci RAM, a następnie programy niezbędne do instalacji są rozpakowywane na wytworzony ramdysk.

Wadą ramdysku jest to, że przechowywane informacje są tracone nie tylko po wyłączeniu zasilania, ale także po przeładowaniu (ponownym uruchomieniu) komputera.

4.13. Jak są zbudowane pamięci stałe (ROM)?

Pamięci stałe są przeznaczone tylko do odczytu przechowywanych informacji. Programowanie pamięci ROM odbywa się na etapie produkcji. Pamięci PROM mogą być jednokrotnie programowane przez użytkownika poprzez przepalenie połączeń matrycy. Pamięci reprogramowalne EPROM i EEPROM mogą być wielokrotnie kasowane i programowane przez użytkownika.

Budowa pamięci stałych ROM jest podobna do prezentowanej w pytaniu 4.8. Różnica polega na tym, że pamięć ta nie ma linii zmiany trybu pracy pamięci Zapis/Odczyt. Budowa pojedynczego elementu pamiętającego jest zależna od rodzaju pamięci.

W przypadku pamięci ROM, już na etapie wytwarzania, tylko w wybranych miejscach matrycy są umieszczane tranzystory. W taki sposób powstają bitowe wzorce zer i jedynek. Programowanie na etapie produkcji jest uzasadnione ekonomicznie jedynie przy produkcji wielkoseryjnej.

Odmianą pamięci ROM jednokrotnie programowalnych są pamięci PROM. W przypadku tych pamięci, na etapie produkcji, są wykonywane wszystkie połączenia matrycy. Programowanie polega na przepaleniu wybranych połączeń matrycy. Proces ten prowadzi do fizycznego zniszczenia elementów, a więc ponowne programowanie jest niemożliwe. Zawartość tak zaprogramowanej pamięci jest trwała.

W pamięciach wymazywalnych EPROM i EEPROM pojedynczym elementem pamiętającym, jest specjalny tranzystor MOS z pływającą bramką. Tranzystory znajdujące się na połączeniach kolumn i wierszy macierzy są normalnie otwarte, a zamyka się je poprzez umieszczenie ładunku na ich bramkach. Czas przechowywania tak wprowadzonego ładunku przekracza 10 lat. Czas programowania pamięci EPROM i EEPROM jest wielokrotnie dłuższy niż odczyt. Zaletą tego rodzaju pamięci jest możliwość usunięcia ładunków z bramek tranzystorów, czyli wymazanie ich zawartości i umożliwienie ponownego programowania. Kasowanie pamięci EPROM odbywa się poprzez naświetlenie struktury półprzewodnikowej (poprzez specjalne okienko w obudowie) promieniami ultrafioletowymi. Pamięć EEPROM jest pamięcią kasowaną elektrycznie, co pozwala na selektywne kasowanie wybranych komórek. Elektryczne kasowanie pozwala także na zmianę zawartości pamięci w układzie, w którym pracuje.

Warto tu wspomnieć, że pamięć EEPROM jest niemalże podstawową „roboczą” pamięcią kart elektronicznych (zwaną potocznie chipowymi – wbrew zaleceniom Polskiej Normy).

4.14. Do czego służą pamięci zewnętrzne (masowe) i jakie są jej rodzaje?

Pamięci zewnętrzne służą do przechowywania dużych ilości (masy) informacji. Pamięci zewnętrzne mogą przechowywać wykorzystywane programy i dane, jak i pełnić rolę archiwum. Pamięci masowych (dysk twardy) używa się jako pamięci wirtualnej komputera. Łączna objętość używanych programów jest na tyle duża, że umieszczenie ich razem w pamięci operacyjnej nie jest możliwe. Muszą być one dostępne na każde żądanie użytkownika. Taką właśnie dostępność i łatwość sprowadzenia dowolnej ilości informacji do pamięci operacyjnej zapewniają pamięci zewnętrzne.

Pamięci te są znacznie tańsze od pamięci operacyjnej, wielokrotnie pojemniejsze, ale i także wolniejsze. Do najczęściej spotykanych pamięci zewnętrznych można zaliczyć:

- dyski magnetyczne:
 - dyski twarde
 - dyskietki twarde (np. SyQuest, ZIP)
 - „standardowe” dyskietki
- dyski magnetoptyczne (np. minidysk Sony)
- dyski optyczne:
 - CD-ROM
 - CD-R
 - CD-RW
 - DVD
 - DVD-RAM
- pamięci taśmowe:
 - streamery.

4.15. Na czym polega zasada działania pamięci z ruchomym nośnikiem magnetycznym?

Zasada działania wszystkich pamięci z nośnikiem magnetycznym jest podobna. Pod specjalną głowicą, zawierającą uzwojenia, przesuwa się materiał z naniesioną warstwą magnetyczną. Sygnał elektryczny, reprezentujący informację, jest doprowadzany do uzwojenia głowicy i wytwarza w jej szczeliny pole magnetyczne. Kierunek i natężenie pola zależy od prądu płynącego przez uzwojenie. Wytworzone pole powoduje przemagnesowanie przesuwanego się w nim nośnika i zarejestrowanie informacji, odpowiadającej sygnałowi doprowadzonemu do uzwojenia.

W celu odczytania tak zapisanej informacji należy namagnesowany nośnik przesunąć z odpowiednią szybkością pod głowicą. W uzwojeniu głowicy indukuje się wówczas napięcie, odpowiadające namagnesowaniu nośnika.

W urządzeniach elektroakustycznych, działających na identycznej zasadzie, sygnał wejściowy jest sygnałem analogowym. Najważniejszym problemem przy jego rejestrowaniu i odtwarzaniu jest możliwie wierne przeniesienie kształtu. W rozpatrywanych tutaj pamięciach zewnętrznych sygnał doprowadzony do głowicy reprezentuje informację cyfrową, kodowaną dwójkowo. Sytuacja jest więc pozornie łatwiejsza – chodzi bowiem o wierne przeniesienie tylko dwóch możliwych stanów sygnału. W rzeczywistości pojawiają się jednak inne, poważniejsze problemy.

Wierność odtwarzania zapisanej informacji cyfrowej musi być większa niż w układach analogowych, gdyż informacja zbyt mocno zniekształcona jest bezużyteczna. Innym ważnym problemem jest szybkość i gęstość zapisu informacji na nośniku, które muszą być wielokrotnie większe. Problemem jest także kodowanie informacji, ułatwiające jej identyfikację i zapewniające synchronizację. Problemy te zostaną omówione w następnym pytaniu.

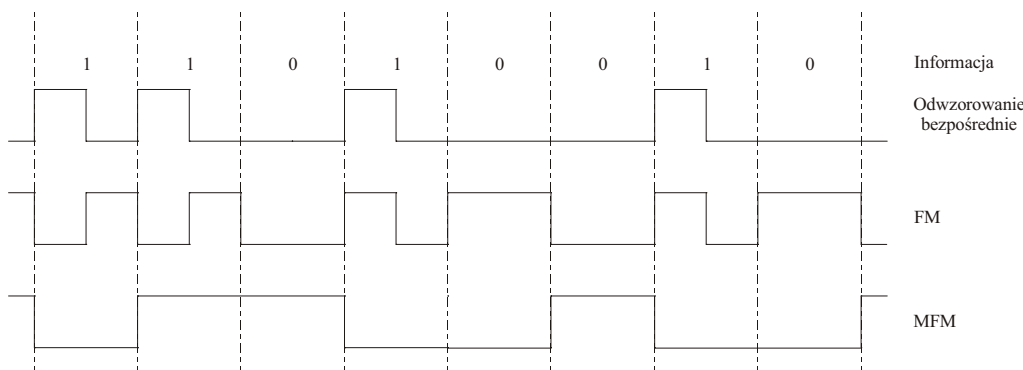
4.16. Jakie metody kodowania informacji stosuje się w pamięciach z ruchomym nośnikiem magnetycznym?

Podstawowym celem stosowania kodowania jest zapewnienie poprawnej identyfikacji zapisanej informacji przy jednoczesnej optymalizacji parametrów i kosztów pamięci.

Bezpośrednie odwzorowanie informacji w sygnał może powodować utratę informacji o początku i końcu danego bitu. Dzieje się tak, gdy kodowana informacja zawiera ciąg zer lub jedynek następujących po sobie.

Metodą, która umożliwi odtworzenie impulsów synchronizujących, jest kodowanie FM. W metodzie tej na początku każdego bitu następuje zmiana kierunku prądu przemagnesowującego. Dodatkowo, podczas trwania bitu reprezentującego jedynekę wprowadzany jest dodatkowy impuls przemagnesowujący. Wadą tej metody jest duża liczba przemagnesowań przypadających na dany ciąg danych.

Zmniejszenie liczby przemagnesowań można osiągnąć metodą MFM. Opiera się ona na analizie nie tylko pojedynczego bitu, lecz bitu bieżącego i poprzedniego. Impuls przemagnesowujący jest generowany na początku każdego bitu z wyjątkiem sytuacji, gdy następuje zmiana z jedynki na zero. Reguła generacji impulsu przemagnesowującego analizuje bit poprzedni i bieżący. Tak więc generowany jest maksymalnie jeden impuls na jeden bit informacji.



Rysunek 4.5. Kodowanie informacji w pamięciach z ruchomym nośnikiem magnetycznym

Istnieją jeszcze doskonalsze metody zmniejszające liczbę impulsów przemagnesowujących, przypadających na daną porcję informacji. Polegają one na jednoczesnym kodowaniu większej liczby bitów ciągu wejściowego (metoda RLL). Opis metody wykracza jednak poza ramy tego opracowania.

4.17. Jak działają dyski twarde (typu Winchester)?

Obecnie nazwy Winchester używa się już rzadko. Dyski tego typu są określane jako dyski twarde. Nośnikiem informacji jest zespół od dwóch do dziesięciu talerzy, umieszczonych na wspólnej osi. Talerze wirują z prędkością 3,6 do 10 tys. obrotów na minutę. Około 2,5 μm nad powierzchnią każdego z wirujących krążków zawieszona są głowice. Poduszka powietrzna, która tworzy się nad powierzchnią wirujących krążków, zapobiega uderzeniu głowicy w powierzchnię nośnika. Głowice są umocowane na ruchomym ramieniu. Pozwala to na ich przemieszczanie wzdłuż promienia nad powierzchnią nośników. Zespół talerzy, silnik napędzający oraz ruchome ramię wraz z głowicami są zamknięte w szczelnym pudełku. Do pudełka zawierającego część mechaniczną dysku przymocowany jest sterownik.

Nośnik magnetyczny jest podzielony na okręgi zwane ścieżkami, które dzielą się na sektory. W każdym sektorze zapisuje się blok danych. Odczytu i zapisu informacji dokonuje się w porcjach danych (blokach) o wielkości odpowiadającej sektorowi.

4.18. Jak działa napęd dysków elastycznych?

Pamięci z dyskami elastycznymi są najtańsze, najprostsze i oczywiście mają najgorsze parametry ze wszystkich pamięci dyskowych.

Nośnikiem informacji jest elastyczny dysk o średnicy 3,5" zamknięty w plastikowej obudowie. Konstrukcja obudowy umożliwia umieszczenie dyskietki w napędzie.

Po umieszczeniu jej w napędzie dochodzi do odsłonięcia nośnika magnetycznego, dosunięcia głowic do powierzchni nośnika oraz uruchomienia silnika poruszającego nośnik. Aby zwiększyć pojemność dyskietki, stosuje się zapis po obu stronach nośnika magnetycznego. Napęd ma wówczas dwie głowice, dosuwane z góry i z dołu powierzchni nośnika.

W obudowie dyskietki znajdują się otwory sygnalizujące zabezpieczenie przed zapisem oraz typ dyskietki.

W czasie pracy dysk wiruje z prędkością kilkuset obrotów na minutę, a głowica styka się bezpośrednio z powierzchnią nośnika. Zapis informacji jest realizowany na okrągłych ścieżkach. Zmiana ścieżki odbywa się poprzez prostopadły przesuw głowicy. Ścieżki są podzielone na sektory, w których zapisuje się bloki danych. W czasie zapisu przesyłany jest blok danych i zapisywany w określonym sektorze danej ścieżki. Zapisu i odczytu można dokonywać blokami o wielkości odpowiadającej wielkości sektora.

Typowa dyskietka o pojemności 1,44 MB zawiera dwustronny nośnik magnetyczny. Na każdej stronie znajduje się 80 ścieżek po 18 sektorów. Pojedynczy sektor służy do zapisania 512 bajtów, czyli na całej dyskietce fizycznie można zapisać $2 \cdot 80 \cdot 18 \cdot 512 = 1474560$ bajtów. Rozbieżność pomiędzy 1,44 MB a 1474560 B wynika z przeliczenia: $1474560:1024 = 1440$ kB, (tu kilobajty dziesiętnie zaokrąglono do 1,44 MB) co zostało nazwane 1,44 MB pomimo, że $1,44 \text{ MB} = 1,44 \cdot 1024 \cdot 1024 \text{ B} = 1509949,44 \text{ B}$ zamiast 1474560 B.

4.19. Co to są dyski wymienne?

Dyski wymienne, np. ZIP czy SyQuest, są odmianą pamięci z ruchomym nośnikiem magnetycznym. Do napędów dysków wymiennych, analogicznie jak do napędów dysków elastycznych, produkowane są specjalne dyskietki wymienne. Pojemność tych dyskietek jest wielokrotnie większa niż standardowych dyskietek elastycznych. Ich popularność jest jednak znacznie mniejsza, a cena wyższa.

Producentem napędów ZIP jest firma Iomega. Pojemność dyskietek do napędów ZIP wynosi 100 lub 250 MB. Napęd do komputera można podłączyć poprzez port równoległy, złącze SCSI, USB lub poprzez IDE (patrz rozdz. 5). Średnia prędkość transmisji wynosi 0,8 MB/s,

Dyski wymienne firmy SyQuest mają pojemności 230 MB, 1 GB lub 1,5 GB. Jeśli napęd jest podłączony przez port równoległy, średnia prędkość transmisji wynosi 1,25 MB/s, natomiast jeśli przez EIDE (patrz rozdz. 5), prędkość ta wynosi 5,6 MB/s.

4.20. Na jakiej zasadzie działają pamięci magnetoptyczne?

W pamięciach tych nośnikiem informacji jest specjalna warstwa magnetoptyczna.

Zapisu informacji na tego typu nośniku dokonuje się, podgrzewając określony punkt powyżej temperatury Curie i używając zewnętrznego pola magnetycznego (wytwarzanego przez głowicę zapisującą), przez co odwraca się bieguny magnetyczne nośnika. Podgrzanie wybranej plamki dokonuje się promieniem lasera mocy 18-20 mW. Ponieważ odwrócenie biegunów magnetycznych jest możliwe dopiero w temperaturze 150–200°C, informacja zapisana na takim dysku jest trwała, odporna na zewnętrzne pole magnetyczne. Trwałość zapisanych danych nie przekracza kilkudziesięciu lat.

Odczyt zapisanych informacji dokonywany jest optycznie. Moc lasera jest dziesięciokrotnie mniejsza niż przy zapisie. Przy odczycie używa się spolaryzowanego światła i wykorzystuje efekt Kerra (światło odbite jest spolaryzowane eliptycznie zgodnie lub niezgodnie z ruchem wskazówek zegara). Korzystając z odpowiedniego zespołu optycznego możliwe jest określenie zapisanej informacji.

Aby dokonać zapisu, należy najpierw skasować zawartość nośnika, nagrać właściwe informacje i dokonać weryfikacji zapisu. Powoduje to, że urządzenia te są niezbyt wydajne. W najnowszych modelach możliwe jest pominięcie fazy kasowania i weryfikacji zapisu, co podnosi znacznie wydajność dysków magnetoptycznych.

4.21. Na czym polega zasada działania pamięci z ruchomym nośnikiem optycznym CD, CD-R, CD-RW?

W nośnikach tego typu zarówno odczyt, jak i ewentualny zapis są wykonywane wyłącznie optycznie. Odstępstwem od tego jest wytwarzanie płyt CD-ROM poprzez tłoczenie.

Odtwarzanie płyt jest możliwe zarówno w urządzeniach odtwarzających, jak i nagrywających. Nagrywanie płyt możliwe jest tylko w specjalnych urządzeniach, które mają możliwość zapisu (są bardziej skomplikowane i droższe). Potocznie są one nazywane nagrywarkami. Nagrywanie płyt CD-R jest możliwe w nagrywarkach CD-R i CD-RW, natomiast nagrywanie płyt wielokrotnego zapisu wyłącznie w nagrywarkach CD-RW.

Nośnikiem informacji jest krążek o średnicy 12 lub 8 cm i grubości 1,2 mm. Na nośniku znajduje się spiralna ścieżka, której długość wynosi ok. 6 km. Krążek po umieszczeniu w napędzie wiruje z dużą prędkością. Odczyt danych prowadzony jest przez głowicę, wyposażoną w laser oraz fotodiodę i umieszczoną na ruchomym ramieniu.

Wytwarzanie płyt CD-ROM polega na sporządzeniu formy, którą tłoczone są krążki z poliwęglanu. Następnie na wytłoczony krążek jest napyłana cienka warstwa aluminium. Całość pokrywa się warstwą lakieru zabezpieczającego, a później umieszcza się napisy i inne ozdoby. Odczyt ścieżki odbywa się po stronie przeciwnej. Gdy promień lasera trafia na gładką powierzchnię, następuje odbicie i powrót promienia do fotodiody. Miejsce gładkie, od którego promień się odbija, to tak zwany land, natomiast wgłębienie w płycie to tak zwany pit. Jeśli promień lasera trafi na wgłębienie, następuje odbicie w innym kierunku i promień nie wraca do fotodiody. Ścieżka, na której są zapisane dane, składa się z kombinacji powierzchni gładkich i zagłębień.

Informacje nie są zapisane na płycie bezpośrednio. W celu zabezpieczenia danych przed błędami odczytu (spowodowanych zarysowaniami, zabrudzeniem płyty) stosuje się specjalne kodowanie. Informacje zapisywane na płycie są rozproszone. Uszkodzenie kilkuset kolejnych bajtów nie powoduje utraty danych. Jeśli uszkodzeniu ulegnie zbyt wiele danych, odczyt będzie niemożliwy (patrz pyt. 4.22).

Oprócz tłoczenia istnieje też inna metoda wytwarzania płyt. Można je nagrywać samemu, w specjalnych urządzeniach. Do nagrywania wykorzystuje się płyty zawierające warstwę specjalnego barwnika umieszczonego przed warstwą odbijającą. Czysta płyta CD-R na gładkim krążku poliwęglowym ma naniesiony fotoczuły barwnik. Na tę fotoczułą warstwę jest napyłana odbłaskowa warstwa złota, potem nakładany jest lakier zabezpieczający i nanoszone napisy. Laser zapisujący, o długości fali 780 nm i znacznie większej mocy niż odczytujący, rozgrzewa warstwę barwnika, tworząc powierzchnie pochłaniające promienie przy późniejszym odczycie. Powierzchnie pochłaniające są odpowiednikiem „dołków” na płytach tłoczonych.

W ostatnich latach pojawiła się również możliwość wielokrotnego zapisu na danym nośniku. Płyty umożliwiające wielokrotny zapis są określane jako CD-RW. Tutaj również przed warstwą odbijającą znajduje się warstwa barwnika. Barwnik stosowany w płytach CD-RW pod wpływem działania lasera o jednej długości fali staje się amorficzny i pochłania światło. Ten sam barwnik pod wpływem lasera o innej długości fali, staje się krystaliczny i przepuszcza światło. Zmiany, jakim jest poddawany barwnik, są odwracalne, więc płyta może być kasowana i nagrywana ponownie. Producenci płyt gwarantują trwałość około 1000 cykli kasowania i zapisu. Podobnie jak poprzednio, warstwa pochłaniająca światło odpowiada „dołkom” na płycie tłoczonej.

4.22. W jaki sposób koduje się informacje w pamięciach z nośnikiem optycznym?

W pamięciach optycznych każdy sektor oprócz właściwej informacji zawiera również informacje niezbędne do korygowania błędów. Liczba bajtów w sektorze i ich przeznaczenie są określane dla danego standardu nośnika oddzielnie. W standardowych dyskach CD-ROM sektor składa się z 2352 bajtów (12 – synchronizacja, 4 – nagłówek,

2048 – właściwe dane, 288 – korekcja błędów). W tak zapisanym sektorze można odtworzyć informacje zawierające do 450 kolejnych błędnych bajtów. Jest to również możliwe dzięki rozproszonemu zapisowi danych. Rozproszenie zapisu polega na tym, że informacje nie są zapisywane po kolei, a są poprzeplatane między sobą.

Ponieważ odczyt odbywa się drogą optyczną, najczęstszą przyczyną występowania błędów są zabrudzenia i zarysowania płyty. Niewielkie zanieczyszczenia powierzchni zewnętrznej nie są aż tak szkodliwe, ponieważ ognisko lasera jest oddalone od powierzchni płyty o ponad 1 mm. Większym problemem są zarysowania, które powodują załamywanie promienia lasera.

4.23. Co to jest DVD?

DVD (ang. *Digital Versatile Disk*) to nowy standard pozwalający na płycie o średnicy 12 cm i grubości 1,2 mm (wymiary identyczne jak w standardowej płycie CD-ROM) przechowywać informacje o wielkości do 17 GB. Tak duża pojemność umożliwia wykorzystanie DVD do zapisu filmów wysokiej jakości. Przykładowo film o czasie trwania 116 minut zajmuje około 4,5 GB przestrzeni dyskowej (szacunkowa pojemność jednostronnej płyty jednowarstwowej).

Format ten powstał w wyniku porozumienia rywalizujących ze sobą firm, poprzez połączenie największych zalet ich niezależnych opracowań. Napęd DVD jest przystosowany do odtwarzania zarówno płyt DVD, jak i starych CD-ROM.

W nowym standardzie zastosowano dwuwarstwowy i dwustronny zapis, zmniejszono rozmiary pitów i landów (z 0,6 μm do 0,4 μm) oraz odległość między ścieżkami (z 1,6 μm do 0,74 μm). Było to możliwe dzięki zmniejszeniu długości fali promienia lasera. Zwiększono także prędkość liniową.

Na dyskach DVD informacje przechowuje się na dwóch warstwach. Pierwsza warstwa jest pokryta substancją częściowo odbijającą, a częściowo przepuszczającą światło. Zdolność do odbijania światła zapewnia poprawne czytanie istniejących zagłębień, natomiast zdolność przepuszczania umożliwia odczyt danych z warstwy głębiej położonej. Zastosowanie dwuwarstwowej konstrukcji pozwala osiągnąć pojemność 8,5 GB. Ponieważ wykonuje się płyty dwustronne (czterowarstwowe), maksymalna pojemność krążka wynosi 17 GB.

4.24. Jak jest zbudowany streamer?

Streamer jest odmianą pamięci masowej, na której informacje są przechowywane na taśmie magnetycznej. Pamięć taśmowa jest rodzajem pamięci o dostępie sekwencyjnym. Oznacza to, że czas dostępu do danych zależy od położenia informacji na nośniku oraz od pozycji danych aktualnie czytanych. Pamięć taśmowa jest pamięcią sto-

sunkowo wolną. Głównym jej przeznaczeniem jest sporządzanie kopii informacji zapisanych na dyskach magnetycznych.

Konstrukcja i zasada działania w dużym uproszczeniu przypomina zasadę działania magnetofonu. Nośnikiem informacji jest taśma pokryta materiałem o właściwościach magnetycznych. Zasada zapisu i odczytu jest identyczna jak w innych pamięciach z nośnikiem magnetycznym.

Pod głowicą zawierającą uzwojenia znajduje się przesuwany z pewną prędkością nośnik. Sygnał elektryczny, reprezentujący informację, jest doprowadzany do uzwojenia głowicy i wytwarza w jej szczelinie pole magnetyczne. Kierunek i natężenie pola zależy od prądu płynącego przez uzwojenie. Wytworzone pole powoduje przemagnesowanie przesuwającego się w nim nośnika i zarejestrowanie informacji, odpowiadającej sygnałowi doprowadzonemu do uzwojenia.

W celu odczytania tak zapisanej informacji należy namagnesowany nośnik przesunąć z odpowiednią szybkością pod głowicą. W uzwojeniu głowicy zaindukuje się wówczas napięcie odpowiadające namagnesowaniu nośnika.

Napędy taśmowe mogą używać dwu sposobów zapisu. Pierwszy z nich, zapis liniowy, charakteryzuje się nieruchomą głowicą, pod którą przesuwana jest taśma ze stosunkowo dużą prędkością. Konstrukcja urządzenia oraz nośnik narażone są na duże obciążenia mechaniczne.

Drugim rodzajem zapisu, używanym w streamerach, jest zapis helikalny, przypominający nieco sposób zapisu i odczytu magnetowidów VHS. Głównym elementem konstrukcyjnym napędu jest zespół wirujących głowic. Zespół głowic umieszczony na bębnie jest usytuowany ukośnie w stosunku do nośnika. Równomierny przesuw taśmy oraz wirowanie głowic powodują powstawanie ukośnych ścieżek zapisu. Zastosowanie ruchomej głowicy pozwoliło zmniejszyć prędkość przesuwu taśmy przy większym upakowaniu danych na tej samej powierzchni. Zmniejszenie prędkości przy odczycie i zapisie przyczyniło się do skrócenia długości taśmy, a tym samym do skrócenia czasu przewijania w stosunku do urządzeń z zapisem liniowym. Helikalny standard zapisu DDS3 (ang. *Digital Data Storage*) jest stosowany w taśmach DAT.

4.25. Co to jest backup? W jaki sposób archiwizuje się informacje i dlaczego?

Backup czyli składowanie jest kopią informacji znajdujących się na dysku komputera. Jeżeli na dysku komputera znajdują się ważne informacje, to należy sporządzić ich kopię zapasową. W przypadku gdy nie wykona się kopii informacji i zostaną one utracone, ich odzyskanie nie będzie możliwe. Częstość wykonywania backupów zależy od ważności informacji, powinna być tak dobrana, aby utrata informacji od chwili wykonania poprzedniej kopii zapasowej nie była zbyt kosztowna. Oczywiście nie

trzeba za każdym razem kopiować wszystkich informacji. Po instalacji systemu komputerowego powinno się wykonać archiwizację „stanu wyjściowego”. Archiwizacja pełna (wykonanie kopii wszystkich plików) powinna być wykonywana regularnie z częstością zależną od polityki bezpieczeństwa firmy (np. raz w miesiącu). Trzecim rodzajem jest archiwizacja częściowa, (wykonanie kopii zapasowej tylko tych plików, które zostały zmienione od czasu ostatniej archiwizacji).

W zależności od tego, przed jakimi zdarzeniami losowymi chronimy nasze informacje, kopia może znajdować się w różnych miejscach. Jeśli dane mają być zabezpieczone na wypadek pożaru lub włamania, to kopia zapasowa powinna znajdować się w innym pomieszczeniu lub nawet budynku. Jeśli natomiast chce się zabezpieczyć informacje przed hackerem komputerowym, to kopia może znajdować się w tym samym pomieszczeniu.

Kopie zapasowe można wykonywać z wykorzystaniem wszystkich dostępnych rodzajów pamięci masowych (taśma, dysk twardy, dysk magneto-optyczny, CD-R/RW itp.). Najczęściej spotykane urządzenia do wykonywania kopii zapasowych to napędy taśmowe. Podczas wykonywania kopii zapasowych najczęściej informacje poddawane są kompresji. Zabieg ten pozwala zaoszczędzić miejsce na nośniku (patrz pyt. 4.27).

4.26. Co to jest mirroring?

Odbicie lustrzane (ang. *mirroring*) jest jedną z metod zabezpieczania się przed utratą informacji. W metodzie tej każdy dysk twardy ma swoją identyczną kopię. Zapis wykonywany jest jednocześnie na obu dyskach. W przypadku awarii zaletą jest natychmiastowa dostępność informacji, wadą jest wysoki koszt urządzeń.

Mirroring inaczej oznaczany jest jako RAID 1 (ang. *Redundant Array of Independent Disks*), czyli nadmiarowa grupa niezależnych dysków – macierz dyskowa.

Macierze dyskowe mogą pracować w konfiguracjach odpowiadających standardom RAID 0 do RAID 5:

- RAID 0 – informacje dzielone są na bloki i zapisywane na kolejnych dyskach, powoduje to przyspieszenie operacji dyskowych, nie gwarantuje bezpieczeństwa
- RAID 1 – wyżej opisany mirroring
- RAID 2 – zastosowanie do zapisu informacji samokorygującego się kodu Hamminga
- RAID 3 – informacje zapisywane w blokach na wielu dyskach, w taki sposób jak w RAID 0, dodatkowo zapisuje się bity parzystości na osobnym dysku, daje to wzrost prędkości i podnosi bezpieczeństwo
- RAID 4 – ulepszona wersja RAID 3, umożliwiająca równoczesny odczyt kilku plików, jednak nie ma możliwości jednoczesnego zapisu wielu plików

- RAID 5 – informacje są dzielone na bloki, nie ma oddzielnego dysku na bity parzystości, bit parzystości dodawany jest do każdego bloku.

4.27. Na czym polega kompresja archiwizowanych informacji?

Kompresja (ang. *compress*) archiwizowanych informacji stosowana jest w celu zmniejszenia obszaru zajmowanego przez oryginalne informacje, tak by po przeprowadzeniu operacji odwrotnej (dekompresji) uzyskać informacje identyczne jak przed kompresją.

W zależności od typu informacji stopień kompresji może sięgać nawet kilku razy (bazy danych czy pliki tekstowe).

4.28. Jakie środki ostrożności należy zachowywać przy przechowywaniu nośników magnetycznych?

Podstawowym i najczęściej spotykanym czynnikiem powodującym uszkodzenie nośników magnetycznych jest działanie zewnętrznych pól magnetycznych. Innymi czynnikami mogą być wysoka temperatura, wilgoć, eksplozje, kurz, brud, itp.

Jeśli na nośniku przechowuje się ważne informacje, to wskazane jest zabezpieczenie nośnika przed zapisem i złożenie go w miejscu odpornym na działanie wymienionych czynników.

5. System wejścia/wyjścia

5.1. Jakie znamy urządzenia wejścia/wyjścia?

W tabeli 5.1 przedstawiono najczęściej spotykane urządzenia wejścia/wyjścia. W oddzielnych kolumnach przedstawiono urządzenia służące wprowadzaniu informacji i urządzenia wyprowadzające informację z komputera. Jeśli urządzenie służy obu celom, zostało podane w obu kolumnach.

Tabela 5.1. Przykłady urządzeń wejściowych i wyjściowych.

Urządzenia wejściowe	Urządzenia wyjściowe
klawiatura	monitor
mysz	panel LCD
skaner	projektor
digitizer	drukarka
mikrofon	ploter
kamera	modem
modem	głośniki
przetwornik analogowo/cyfrowy	przetwornik cyfrowo/analogowy
czytnik kodu kreskowego	
touchpad	
trackball	
touchscreen	

5.2. Dlaczego w nazewnictwie urządzeń wejścia/wyjścia jest tak dużo nazw anglojęzycznych?

Nazewnictwo anglojęzyczne w urządzeniach wejścia/wyjścia zostało przejęte wraz z urządzeniami, a jednocześnie wystąpiły trudności w tworzeniu dla nich nazw polskojęzycznych. W urządzeniach wejścia/wyjścia jest najwięcej nowinek. Obecnie coraz częściej nowe urządzenia przyjmują oryginalne nazwy, nadane przez producenta.

Przyczyną stosowania nazw angielskich jest również powszechność języka angielskiego w informatyce. Niestety troska o czystość języka polskiego (w rozumieniu prof. Doroszewskiego) to w informatyce trochę walka z wiatrakami. Opór środowiska jest duży.

5.3. Jakie parametry techniczne i cechy funkcjonalne urządzeń wejścia/wyjścia są istotne z punktu widzenia komputera?

Podstawowym problemem przy współpracy urządzeń z komputerem jest standard złącz, jakimi będą połączone.

W różnych standardach stosowane są różne sposoby wymiany informacji (określony protokół transmisji, rodzaj zastosowanego kodowania przekazywanych informacji). Dla przykładu drukarkę podłącza się przez port równoległy, natomiast modem przez port szeregowy. Szeregowa transmisja charakteryzuje się przesyłaniem informacji bit po bicie pojedynczą linią (w danej chwili przesyłany jest jeden bit informacji). W transmisji równoległej jednocześnie przesyła się kilka bitów informacji oddzielnymi liniami.

Dla komputera bardzo ważne jest to, czy dane urządzenie jest jedno- czy dwukierunkowe (modem). Zarówno połączenie z drukarką (np. atramentową czy laserową), jak i z modemem jest dwukierunkowe. Dwukierunkowość połączenia z drukarką daje możliwość kontroli stanu, w jakim znajduje się drukarka (zacięcie papieru, brak tonera lub tuszu). Typowo jednokierunkowe jest połączenie pomiędzy kartą dźwiękową i głośnikami.

Do najważniejszych parametrów należą: rodzaj i prędkość transmisji. Transmisja może być synchroniczna (wymuszana zegarem taktującym) lub asynchroniczna. Dla przykładu: połączenie komputera z myszą realizuje się wolnym łączem asynchronicznym, natomiast przy połączeniu z modemem – łączem asynchronicznym o znacznie większej szybkości transmisji.

5.4. Co to jest system wejścia/wyjścia?

Przez system wejścia/wyjścia rozumie się zespół środków sprzętowych i programowych określających w jednolity sposób dla danego komputera metody i reguły przesyłania zarówno danych, jak i rozkazów między tym komputerem a jego otoczeniem. Podstawowymi czynnikami określającymi możliwości systemu wejścia/wyjścia są: szybkość transmisji, typy tej transmisji oraz sposób sterowania priorytetami urządzeń. Otoczeniem dla danego komputera może być także inny komputer np. w sieci komputerowej.

5.5. W jaki sposób przedstawia się informacje przesyłane do i z urządzeń wejścia/wyjścia?

Informacje są kodowane. Kodowaniem cyfrowym w informatyce zwykło się określać sposób jednoznacznego przyporządkowania informacjom ciągów zerojedynkowych. Spośród wielu teoretycznie możliwych rozwiązań, powszechnie stosowany jest standard kodowania znaków alfanumerycznych ASCII. Inne informacje, przesyłane pomiędzy urządzeniami wejścia/wyjścia, nie są kodowane według określonego standardu.

Zapis liter w standardowym kodzie ASCII odbywa się za pomocą 7 bitów. Rozszerzenie ASCII wykorzystuje 8 bitów (pierwszym 128 znakom odpowiadają znaki ze standardowego kodu ASCII, a pozostałe 128 symboli odpowiada znakom specjalnym). Rozszerzenie kodu ASCII zawiera m.in. znaki umożliwiające rysowanie tabel i ramek (semigrafika), znaki liter specyficznych dla języków innych niż angielski (ą, ę, ó, ł, ń, ć, itd.).

5.6. Jak w komputerze przedstawia się znaki narodowe?

Znaki narodowe, podobnie jak wszystkie informacje, są kodowane. Obecnie do zapisu znaków wykorzystuje się liczby zapisywane na ośmiu bitach. Pierwsze 128 znaków odpowiada kodom ASCII (ang. *American Standard Code for Information Interchange*). Kolejne kody reprezentują znaki dodatkowe. Ich znaczenie jest zależne od standardu, jaki przyjęto przy kodowaniu. Istnieje wiele sposobów umieszczenia znaków narodowych. Parę lat temu spotykano jeszcze około dwudziestu standardów zapisu znaków z „ogonkami” (ą, ć, ę, ł, ń, ś, ó, ź, ż, Ą, Ć, Ę, Ł, Ń, Ś, Ó, Ź, Ż). Do najczęściej spotykanych należały: ISO-8859-2, IBM (CP852), Mazovia, CSK, Cyfromat, Ventura.

Obecnie spotykane standardy kodowania „ogonków” to:

- CP852 – wprowadzony przez Microsoft i IBM
- CP1250 – wprowadzony do MS Windows
- ISO-8859-2 – proponowane przez standard międzynarodowy i Polską Normę.

Polska Norma (PN-93 T-42118), dotycząca sposobu kodowania polskich znaków, jest zgodna z ISO 8859-2. W tym standardzie kodowanie polskich znaków wygląda następująco:

Ą	Ć	Ę	Ł	Ń	Ó	Ś	Ź	Ż
161	198	202	163	209	211	166	172	175
ą	ć	ę	ł	ń	ó	ś	ź	ż
177	230	234	179	241	243	182	188	191

Istnieje już propozycja standardu, w którym znajdzie się miejsce na zakodowanie wszystkich znaków narodowych. Nosi on nazwę Unikod. Pełny Unikod jest standardem 32-bitowym. Aktualnie używa się 16 bitów. Standard ten jest niezastąpiony przy kodowaniu tekstów wielojęzycznych. W standardzie Unikod znaki odpowiadające kodom ASCII są przesyłane bez zmian, natomiast kody większe od 127 są modyfikowane. Taki sposób kodowania powoduje niewielkie zwiększenie objętości zapisywanych dokumentów.

5.7. W jaki sposób sprzęga się urządzenia wejścia/wyjścia z komputerem?

Urządzenia wejścia/wyjścia do komputera można dołączać poprzez interfejsy standardowe (RS-232, Centronix, PS2, USB) lub przez specjalizowane karty rozszerzeń, montowane w gniazdach płyty głównej komputera.

Cechą pierwszego rozwiązania jest łatwość dołączenia urządzenia zewnętrznego i możliwość jego szybkiego przeniesienia do innego komputera. Ten sposób podłączenia nakłada na producenta urządzenia zewnętrznego obowiązek dostosowania poziomów elektrycznych oraz sposobu kodowania danych do standardu interfejsu (patrz pyt. 5.10). Producent komputera nie określa rodzaju urządzenia wejścia/wyjścia jakie ma być do niego dołączone. Określa jedynie pewien standard konstrukcyjny (łączówki, konstrukcja mechaniczna), elektryczny (poziom i kształt sygnału), logiczny (format wprowadzanej i wyprowadzanej informacji) oraz programowy (tryb komunikowania się).

Dołączanie urządzeń zewnętrznych poprzez specjalizowane karty rozszerzeń jest bardziej skomplikowane. Montaż takiej karty wymaga rozebrania komputera. Producent urządzenia zewnętrznego ze specjalizowaną kartą ma większe możliwości niż przy wykorzystaniu sprzęgu standardowego. W tym przypadku producent sam określa standard konstrukcyjny, elektryczny, logiczny i programowy na styku karta – urządzenie zewnętrzne. W rozwiązaniu tym nie ma ograniczenia liczby linii wejścia/wyjścia, prędkości przesyłanych danych, poziomów i kształtów napięcia. Producent urządzenia musi jednak zachować standard na styku komputer – karta rozszerzeń.

Pierwsze rozwiązanie stosowane jest przy dołączaniu standardowego urządzenia zewnętrznego (mysz, drukarka, skaner, telefon komórkowy). Drugie rozwiązanie jest bardziej skomplikowane i wykorzystywane do podłączania nietypowych sterowników procesów przemysłowych oraz urządzeń wymagających większych prędkości przesyłania danych.

5.8. Jakie funkcje pełni BIOS w systemie komputerowym?

BIOS (ang. *Basic Input Output System*) jest częścią systemu operacyjnego (patrz rozdz. 9) i służy do obsługi standardowych urządzeń wejścia/wyjścia komputera. Zawiera podprogramy służące obsłudze tych urządzeń.

Zarówno system operacyjny, jak i inne programy komunikują się z urządzeniami zewnętrznymi poprzez podprogramy zawarte w BIOS-ie. Dzięki takiemu rozwiązaniu w komputerach różnych producentów mogą pracować te same programy i te same systemy operacyjne.

BIOS oprócz podprogramów obsługujących urządzenia zewnętrzne zawiera również program konfiguracyjny i testowy. Obecnie produkowane komputery umożliwiają zmianę BIOS-u. Jest to możliwe dzięki umieszczeniu go w pamięci stałej typu Flash (patrz pyt. 4.6).

5.9. W jaki sposób zmieniamy konfigurację BIOS-u?

Najczęściej konfigurację BIOS-u zmienia się poprzez program konfiguracyjny, uruchamiany przy starcie komputera. Program konfiguracyjny uruchamia się poprzez wciśnięcie odpowiedniego klawisza (najczęściej „Delete”).

Istnieją też inne rozwiązania. Do komputera może być dołączony program konfiguracyjny uruchamiany z dysku podczas pracy.

W BIOS-ie można zdefiniować hasło zabezpieczające komputer przed uruchomieniem przez osoby nieuprawnione. Hasło może ograniczać dostęp do samego BIOS-u (uniemożliwiając zmianę konfiguracji) lub dostęp do komputera (uniemożliwiając pracę bez podania hasła).

Ustawienia komputera zmieniane w BIOS-ie są zapisywane w pamięci CMOS. Jest to pamięć podtrzymywana bateryjnie. Rozwiązanie takie umożliwia usunięcie zawartości pamięci poprzez odłączenie baterii. Powoduje to ustawienie wszystkich wartości na domyślne (także usunięcie hasła zabezpieczającego).

5.10. Co to jest interfejs?

Interfejs (ang. *interface* – sprzęg) jest to zespół ustalonych reguł oraz środków technicznych łączenia komputera z urządzeniami zewnętrznymi. Na interfejs składają się:

- wymagania logiczne (liczba sygnałów i ich funkcje, sposób kodowania, sposób synchronizacji)
- wymagania elektryczne (kształt, poziom, czas przełączania sygnałów)

- wymagania konstrukcyjne (typ złącz, gniazd, styków, rodzaje i długości przewodów).

Interfejs może zapewniać przesyłanie danych w jednym lub w obu kierunkach. Interfejsy pozwalają na przesyłanie danych szeregowo lub równoległe.

5.11. Co to jest Centronix?

Centronix jest rodzajem interfejsu równoległego (patrz pyt. 5.3). Standardowy Centronix jest interfejsem jednokierunkowym. Istnieją jednak jego dwukierunkowe odmiany. W komputerach produkowanych obecnie można samemu wybrać pożądany tryb pracy portu równoległego (SPP, EPP czy ECP). Najczęstszym zastosowaniem tego interfejsu jest dołączenie drukarki do komputera. Sprzęg równoległy jest również wykorzystywany przy dołączaniu skanera, napędu dysków wymiennych ZIP lub kart sieciowych.

W standardzie Centronix złącze ma 25 styków (8 – dane, 9 – sygnały sterujące, 8 – masa). Przyjęte poziomy sygnały elektrycznego w złączu to 0 i +5 V.

5.12. Czym różnią się tryby pracy portu równoległego SPP, BPP, EPP, ECP?

Tryby te różnią się sposobem przesyłania danych oraz uzyskiwanymi szybkościami transmisji.

W trybie SPP (ang. *Standard Printer Port*) do transmisji dwukierunkowej wykorzystuje się przewody kontrolne i statusu. W czasie jednego cyklu zegara transmitowane są 4 bity. Transmisja w taki sposób jest wolna (do 90 kB/s).

Tryb BPP (ang. *Bidirectional Printer Port*) jest nowszą wersją trybu SPP i jest przystosowany do dwukierunkowego przesyłania danych. W ten sposób wyeliminowano użycie linii statusowych do przesyłania danych (jak to miało miejsce w SPP). Zmiana ta umożliwiła uzyskanie szybkości transmisji 180 kB/s.

Kolejną modyfikacją portu równoległego było wprowadzenie portu rozszerzonego EPP (ang. *Enhanced Parallel Port*). Wprowadzone ulepszenie (ang. *handshake* – transmisja z potwierdzeniem) pozwoliło zwiększyć szybkość transmisji do ok. 1 MB/s.

W trybie ECP (ang. *Enhanced Capability Port*) wykorzystano DMA, FIFO (ang. *First In First Out*) oraz kompresję danych. ECP jest rozszerzeniem trybu EPP, dzięki czemu uzyskano praktyczne szybkości transmisji do 2 MB/s.

5.13. Co to jest RS-232?

RS-232 (ang. *Recommended Standard*) jest standardem interfejsu szeregowego. W standardzie RS-232 zdefiniowano poziomy elektryczne sygnałów, fizyczny rodzaj złącza, rozmieszczenie przewodów w złączu oraz maksymalną długość przewodów. Standard przewiduje synchroniczny i asynchroniczny tryb transmisji. Najczęściej spotykany jest tryb transmisji asynchronicznej, polegający na przesyłaniu znaków w ściśle określonym formacie, ramce. Każda ramka zaczyna się bitem startu, następnie przesyłane są bity danych, bit kontroli parzystości (opcjonalny) oraz bit stopu. W ramce może występować bit kontroli parzystości, który umożliwia wykrycie nieparzystej liczby błędów.

Standard RS-232 definiuje dopuszczalne szybkości transmisji (150, 300, 600, 1.200, 2.400, 4.800, 9.600, 19.200, 38.400, 57.600, 115.200 bit/s).

Poziomy napięć elektrycznych zdefiniowane w standardzie są następujące: logicznej „1” odpowiada poziom napięcia od -3 V do -15 V , logicznemu „0” poziom napięcia od $+3\text{ V}$ do $+15\text{ V}$.

Według standardu RS-232 istnieją dwa rodzaje złącz: 9-stykowe lub 25-stykowe.

5.14. Jak procesor obsługuje urządzenia wejścia/wyjścia?

Obsługa urządzeń wejścia/wyjścia może być realizowana programowo, za pomocą przerwania lub poprzez bezpośredni dostęp do pamięci – DMA (ang. *Direct Memory Access*). W zależności od sposobu obsługi urządzenia zmienia się zaangażowanie procesora.

Najbardziej „zajmująca” procesor jest obsługa programowa. W tym przypadku, wykonywany program działa w pętli i odpytuje urządzenie zewnętrzne. Program główny jest zatrzymywany w czasie przesyłania i przetwarzania danych przez urządzenie. Metoda ta jest nieefektywna, procesor traci dużo czasu podczas oczekiwania na reakcję urządzenia.

Metoda obsługi za pomocą przerwania jest bardziej wydajna od poprzedniej. Zastosowanie przerwania zwalnia procesor z obowiązku ciągłego sprawdzania stanu urządzenia. W tej metodzie urządzenie, poprzez zgłoszenie przerwania, sygnalizuje procesorowi konieczność obsługi. Dzięki temu czas, w którym poprzednio procesor czekał na reakcję urządzenia, można teraz poświęcić na wykonywanie innych czynności. Obsługa za pomocą przerwania wymaga zaangażowania procesora do przesyłania danych między urządzeniami a pamięcią. Metoda ta staje się mało efektywna, jeśli przesyłamy dużo informacji do lub z szybkiego urządzenia. Wadę tę eliminuje kolejna metoda.

W metodzie z bezpośrednim dostępem do pamięci (DMA), informacje są przesyłane bezpośrednio z urządzenia do pamięci i z pamięci do urządzenia. W ten sposób

procesor jedynie nadzoruje przesyłanie informacji. Metoda ta jest najbardziej wydajna i zapewnia najszybszą komunikację z urządzeniem.

5.15. Dlaczego stosuje się przerwania?

W trakcie pracy komputera mogą występować zdarzenia mające źródło w jednostce centralnej bądź jej otoczeniu. Cechą charakterystyczną tych zdarzeń jest to, że pojawiają się w chwilach, których nie można przewidzieć. Zdarzenia te wymagają natychmiastowej obsługi. Przykładami zdarzeń mogą być: zdarzenia w komputerze (awaria zasilacza, dzielenie przez zero, przekroczenie zakresu liczb) lub zdarzenia z otoczenia komputera (gotowość urządzenia wejścia/wyjścia do wykonania transmisji, meldunek od urządzenia o zakończeniu transmisji, błąd transmisji). Oczywiście te specjalne sytuacje mogłyby być badane programowo. Podejście takie wymagałoby od programisty badania wskaźników świadczących o tych zdarzeniach. Wiązałoby się to z marnotrawstwem czasu procesora na czynności „dozorcy”. Dlatego wprowadzono tzw. przerwania. W system przerwań są wyposażone wszystkie obecnie produkowane komputery. Przerwanie jest sygnałem (wytwarzanym sprzętowo), sygnalizującym nadzwyczajne zdarzenie w komputerze lub jego otoczeniu i wymagające natychmiastowej reakcji procesora. Procesor musi mieć wbudowany mechanizm rozpoznawania przyczyny przerwania. Przerwanie jest zwykle obsługiwane przez procesor po zakończeniu cyklu wykonywanego rozkazu, powodując przejście do wykonywania podprogramu obsługi przerwania.

5.16. Jak klasyfikujemy przyczyny przerwań?

Przerwania można podzielić na dwie grupy: przerwania programowe i przerwania sprzętowe.

Przerwania programowe są wywoływane z wnętrza wykonywanego programu. Dzięki przerwaniam programista ma dostęp do wielu przydatnych procedur i funkcji. Programowanie w języku assemblerowym (patrz rozdz. 9) opiera się właśnie na umiejętnym wykorzystywaniu gotowych procedur, wywoływanych poprzez przerwania.

Przerwania sprzętowe są generowane przez „otoczenie” procesora. Źródłami przerwań sprzętowych mogą być np.: układy transmisyjne (niepoprawna transmisja, zakończenie transmisji), sterownik układów pamięci (błąd parzystości, uszkodzenie układu), układ klawiatury (naciśnięcie klawisza).

Może zaistnieć sytuacja, w której do procesora dotrze kilka zgłoszeń przerwań jednocześnie. Aby rozstrzygnąć, w jakiej kolejności należy je obsłużyć, wprowadzono priorytety przerwań.

5.17. Jakie wyróżniamy etapy obsługi przerwania?

Obsługę przerwania przez procesor można podzielić na kilka etapów:

- pojawienie się sygnału przerwania
- zapamiętanie aktualnego stanu rejestrów procesora, w celu umożliwienia późniejszego kontynuowania przerwanej programu
- zawieszenie pracy wykonywanego programu
- przejście do podprogramu obsługi przerwania
- ustawienie priorytetów przerwania
- wykonanie podprogramu obsługi przerwania
- odtworzenie stanu programu zawieszono i powrót do jego realizacji.

Obsługa przerwania zaczyna się zapamiętaniem stanu programu wykonywanego. Zapamiętywane zostają licznik rozkazów oraz stan rejestrów pomocniczych procesora.

Podczas obsługi przerwania należy zabronić procesorowi przyjmowania mniej ważnych przerw. Jest to realizowane przez ustawienie określonych bitów rejestru maski.

Po zakończeniu podprogramu obsługi przerwania, zostaje przywrócona zawartość rejestrów procesora i wykonywanie zawieszono programu jest kontynuowane.

5.18. Jak działa monitor graficzny wyposażony w kineskop?

Działo elektronowe wyrzuca elektrony, które odchylane przez pole elektromagnetyczne cewek uderzają w przednią część kineskopu zwaną luminoforem. Strumień elektronów jest odchylany przez cewki odchylające. Pole cewek jest tak zmieniane, aby strumień elektronów docierał kolejno do wszystkich punktów ekranu. Omiatanie zaczyna się od lewej strony ekranu do prawej. Dalej następuje powrót na początek linii niższej i kolejny przelot strumienia wzdłuż ekranu. Omiatanie kończy się w prawym dolnym rogu ekranu, skąd następuje powrót do lewego górnego punktu. Elektrony uderzając w określony punkt luminoforu powodują jego świecenie. Luminofor jest zbudowany z małych punktów pogrupowanych w triady. Każdy punkt triady świeci jednym z trzech kolorów RGB (czerwonym, zielonym, niebieskim). W kineskopach kolorowych wykorzystuje się trzy działa elektronów (osobne dla każdego z kolorów podstawowych). Przed luminoforem, na drodze elektronów, umieszczona jest specjalna maska, zapewniająca dokładne pozycjonowanie wiązki. Dzięki zastosowaniu maski koryguje się drobne rozbieżności w ogniskowaniu, co poprawia ostrość obrazu i czystość barw. Poszczególne punkty ekranu są ponownie zapalane z częstotliwością rzędu 75 Hz. Działanie to powoduje, że triady świecących punktów nie zdążą jeszcze zgasnąć i unika się migotania obrazu.

Sterowanie wyrzutniami elektronów oraz cewkami odchylającymi odbywa się za pośrednictwem układów elektronicznych. Układy te mają za zadanie przetworzenie sygnałów otrzymywanych z karty graficznej na sygnały sterujące kineskopem. Napięcia dostarczane z karty graficznej są rzędu 1 V, natomiast sterujące kineskopem rzędu kV. W skład układu elektronicznego wchodzi: wzmacniacze odchylenia poziomego i pionowego oraz wzmacniacz wideo. Poza wzmacniaczami niezbędne są również elementy umożliwiające dostrojenie jasności i kontrastu obrazu oraz zmianę geometrii.

5.19. Jakimi parametrami charakteryzuje się monitor graficzny?

Jakość monitora graficznego określają następujące parametry:

- długość przekątnej ekranu
- wielkość plamki
- dopuszczalne rozdzielczości
- częstotliwość odświeżania
- przeplot lub jego brak.

Długość przekątnej ekranu jest wyrażana w calach. W obecnie produkowanych monitorach długość przekątnej ekranu wynosi od czternastu do dwudziestu kilku cali. Im większy monitor, tym większy komfort pracy, ale i wyższa cena zakupu. Obecnie w zastosowaniach domowych najczęściej spotyka się monitory 15" i 17". Praktycznie nie kupuje się już monitorów mniejszych niż 15", a ceny większych niż 17" odstraszają kupujących.

Wielkość plamki jest związana z jakością kineskopu zastosowanego w monitorze. Im mniejsza plamka, tym lepsza jakość obrazu. Obecnie produkowane kineskopy mają plamkę wielkości 0,25 mm.

Rozdzielczość monitora określa liczba linii pionowych i poziomych wyświetlanego obrazu. Im większe rozdzielczości pozwala osiągnąć monitor, tym wyższy komfort pracy. Najczęściej spotykane rozdzielczości to: 640x480, 800x600, 1024x768, 1280x1024, 1600x1200. Są one wyrażone w tak zwanych pikselach, czyli najmniejszych elementach obrazu (ang. *picture elements*).

Częstotliwość odświeżania jest to liczba obrazów wyświetlanych w ciągu sekundy i wynosi od 60 do 120 Hz.

Przeplot jest zjawiskiem niepożądanym w monitorze. Jest to sposób wyświetlania obrazów wysokiej rozdzielczości na monitorze słabej jakości. Polega on na naprzemiennym wyświetlaniu dwóch różnych obrazów o mniejszej liczbie linii, co w efekcie

wywołuje złudzenie uzyskania większej rozdzielczości. Przeplot niekorzystnie wpływa na wzrok podczas pracy przy komputerze.

Z powyższego wynika, że kupując monitor należy dążyć do tego, aby miał jak największą przekątną ekranu, obsługiwał jak największe rozdzielczości przy wysokiej częstotliwości odświeżania, pracował bez przeplotu i miał jak najmniejszą plamkę.

5.20. Jakie wyróżniamy rodzaje drukarek?

Obecnie najczęściej spotyka się drukarki atramentowe, laserowe i igłowe. Czasami jeszcze można spotkać drukarki termiczne. Najwyższą jakość druku uzyskuje się za pomocą drukarki laserowej, najniższą dają drukarki igłowe.

5.21. Jakie parametry charakteryzują drukarkę?

Najważniejszymi parametrami charakterystycznymi drukarki są:

- szybkość druku – wyrażana w znakach/sekundę lub stronach/minutę
- rozdzielczość druku – wyrażana w cpi (ang. *char per inch*) lub lpi (ang. *line per inch*)
- posiadanie interpretera postscriptu lub jego brak.

Interpreter postscriptu jest to funkcja drukarki, która pozwala drukować obraz graficzny na podstawie jego tekstowego opisu. Do opisu wykorzystuje się język, który nazywany jest postscriptem. Informacje przesyłane do drukarki są w postaci tekstowej, gdzie są zamieniane na postać graficzną. Zaletami stosowania postscriptu są: lepsza jakość obrazu, skrócenie czasu transmisji informacji z komputera do drukarki, łatwość manipulacji obrazem.

Dodatkowo warto zwrócić uwagę na:

- sposób dołączenia do komputera
- poziom hałasu podczas drukowania
- pojemność zasobnika papieru
- rozmiar bufora pamięci
- tryb drukowania (tekst, grafika)
- zużycie energii.

Wymienione wcześniej drukarki różnią się sposobem drukowania, czego konsekwencją jest różna jakość, szybkość druku i poziom hałasu.

5.22. Jak działa drukarka igłowa?

Druk w tego rodzaju drukarce powstaje wskutek uderzania (najczęściej 9 lub 24) igieł w papier poprzez taśmę barwiącą. Wynikiem takich uderzeń są pojedyncze punkty ułożone w linie i wiersze. Z pojedynczych punktów są tworzone znaki alfanumeryczne lub grafika. Drukarka igłowa ma ruchomą głowicę (ruch w poziomie) oraz ruchomą rolkę przesuwającą papier (ruch w pionie).

Jakość wydruku w tej drukarce jest niska, a poziom hałasu duży. Zaletą tej drukarki jest możliwość druku na papierze samokopiującym (ważne przy drukowaniu rachunków) oraz duża pojemność zasobnika papieru (papier perforowany – składanka). Do zalet tej drukarki można też dodać niski koszt eksploatacji.

5.23. Na czym polega zasada działania drukarki atramentowej?

W tej drukarce druk powstaje podczas wyrzucania tuszu z dyszy głowicy. Podobnie jak w drukarce igłowej i tu stosuje się głowicę ruchomą w poziomie, a ruch w pionie jest realizowany poprzez przesuw papieru.

Jakość wydruku drukarki atramentowej jest znacznie wyższa niż igłowej (igły i uderzenie mechaniczne zastąpiono wyrzuceniem odrobiny tuszu). Drukarka ta charakteryzuje się bardzo małym poziomem hałasu, średnią szybkością druku (większą niż drukarka igłowa, lecz mniejszą niż drukarka laserowa). Drukarki atramentowe posiadają niezbyt duży pojemnik papieru, a druk odbywa się na pojedynczych kartkach. Do wad należy zaliczyć brak możliwości druku na papierze samokopiującym (ważne w firmach przy drukowaniu rachunków). Ważny jest rodzaj tuszu, tak by nie zasychał on w pojemniku, a szybko zasychał na papierze.

5.24. Jaka jest zasada działania drukarki laserowej?

Zasada działania tej drukarki opiera się na przenoszeniu naelektryzowanym włókien proszku barwiącego (tonera) na papier. Naelektryzowanie obracającego się bębna odbywa się promieniem światła laserowego. Do miejsc naelektryzowanych jest przyciągany proszek barwiący. Następnie toner z ruchomego bębna jest przenoszony na papier. Utrwalenie wydruku następuje, gdy naniesiony proszek barwiący jest termicznie wprasowywany w papier.

Jakość druku w drukarce laserowej jest wyższa niż w atramentowej. Drukarka ta charakteryzuje się również większą szybkością druku. Wadą drukarki jest konieczność przesłania obrazu całej strony do pamięci wewnętrznej (drukarka laserowa tworzy binarny obraz strony przed rozpoczęciem druku). Jest to szczególnie uciążliwe, gdy drukuje się stronę zawierającą złożoną grafikę lub tekst o różnych krojach pisma.

Wówczas potrzeba dużej ilości pamięci, a jeśli drukarka jej nie posiada, poprawne wydrukowanie strony nie będzie możliwe. Problemu tego nie ma, jeśli korzystamy z drukarki atramentowej lub igłowej (nie tworzą obrazu strony).

5.25. Jak działa drukarka termiczna?

W drukarce termicznej druk odbywa się na specjalnym papierze wrażliwym na ciepło. Obraz jest tworzony poprzez przypalenie papieru. Tak wykonany druk jest średniej jakości. Drukarki tego typu charakteryzują się małymi wymiarami i małym poborem mocy, stąd są powszechnie używane w sprzęcie przenośnym zasilanym bateryjnie oraz w faksach.

5.26. Czym różni się ploter od drukarki?

Ploter w inny sposób niż drukarka tworzy obraz otrzymywany z komputera. Obraz powstaje poprzez ruch pisaka po powierzchni papieru. Ploter jest urządzeniem elektromechanicznym. W odróżnieniu od drukarki ploter obsługuje znacznie większe formaty papieru (do formatu A0, podczas gdy drukarka najczęściej drukuje w formacie A4, czasami A3). Ze względu na sposób tworzenia obrazu i duże rozmiary rysunków plotery służą wykonywaniu rysunków technicznych i wykresów.

W ploterach istnieje możliwość wykonywania rysunków w wielu kolorach, poprzez użycie wielu pisaków.

Można również spotkać plotery służące do wycinania zadanych kształtów w metalu, drewnie czy tworzywach sztucznych. Aby to było możliwe, należy zamiast pisaka zastosować odpowiednią głowicę tnącą.

5.27. Jak działa skaner?

Skaner służy przetwarzaniu obrazu obiektu na postać cyfrową. Obraz w skanerze jest analizowany przez układ optyczny poruszany wzdłuż obiektu. Obraz z układu optycznego jest przetwarzany na postać cyfrową i przekazywany do komputera. Skaner może być ręczny (ang. *handy*) skaner jest przesuwany nad analizowanym obrazem lub stołowy (obiekt analizowany układa się na powierzchni skanera). Skanery wyższej jakości są wykonywane jako stołowe. W tych skanerach układ optyczny jest poruszany automatycznie pod analizowanym obiektem.

O jakości skanera świadczy otrzymywana rozdzielczość analizowanego obrazu – typowo 600 lub 1200 dpi (ang. *dot per inch* – punktów na cal). Oczywiście w skanerach ręcznych jest ona mniejsza niż w stołowych. Dla przykładu można podać, że na

płytcie CD-ROM można zapisać około 5000 zdjęć skanowanych z rozdzielczością 300 dpi i zapisanych w standardzie jpg.

Skaner do komputera podłącza się przez port równoległy drukarki lub port USB. Niektóre skanery posiadają jednak własną kartę, instalowaną bezpośrednio na płycie głównej komputera. Jeśli skaner jest podłączony poprzez port drukarki, wówczas drukarkę podłącza się do skanera. Takie rozwiązanie pozwala mieć dostęp do skanera i drukarki bez potrzeby przełączania kabli.

Istnieją też urządzenia będące połączeniem drukarki i skanera, a także drukarki, skanera, telefonu i faksu. Taki zestaw może pełnić funkcję kopiarki.

5.28. Co to są digitizery?

Digitizery są urządzeniami służącymi do zamiany obrazów, rysunków na zbiór współrzędnych. W digitizerze do komputera nie jest przenoszony obraz binarny obiektu (jak ma to miejsce w skanerze), a jedynie zbiór punktów, wskazanych pisakiem elektronicznym na specjalnym pulpicie. Dzięki temu otrzymany rysunek jest reprezentowany w pamięci komputera jako zbiór linii łatwych do skalowania i przetwarzania. Z tego wynika, że digitizer znakomicie nadaje się do przenoszenia rysunków technicznych i wykresów. Natomiast jest bezużyteczny przy przenoszeniu złożonej grafiki czy zdjęć.

Digitizer z punktu widzenia komputera jest odmianą standardowej myszy. Różnica między myszą a digitizerem polega na tym, że do komputera nie są przesyłane współrzędne względne (tak jak w myszy) lecz współrzędne bezwzględne. Pozwala to z dużą precyzją określać położenie wskazywanych punktów.

5.29. Czy do komputera można podłączyć kamerę cyfrową?

Oczywiście połączenie komputera z kamerą wideo jest możliwe. Kamera wideo staje się coraz częstszym wyposażeniem multimedialnego komputera. Kamerę najczęściej dołącza się poprzez złącze USB. Czasami posiada ona własną kartę instalowaną na płycie głównej komputera. Wzrost zainteresowania kamerami wideo współpracującymi z komputerem wiąże się z możliwością prowadzenia wideokonferencji. Stało się to możliwe po opracowaniu wydajnych algorytmów kompresji obrazu oraz standardów transmisji danych z odpowiednią szybkością.

5.30. Czy komputer może wytwarzać sygnały dźwiękowe i czy może rejestrować dźwięki?

Wytwarzanie dźwięków oraz rejestrowanie ich jest możliwe już od dawna. Karta dźwiękowa stanowi standardowe wyposażenie komputera. Umożliwia ona zarówno przetwarzanie sygnału cyfrowego na dźwięk w głośniku (wytwarzanie dźwięku), jak i dźwięku z mikrofonu na sygnał cyfrowy.

Karta dźwiękowa zawiera wyjście analogowe umożliwiające podłączenie słuchawek, zestawu głośniczków czy wzmacniacza mocy. Na karcie znajduje się kilka wejść sygnału analogowego. Najczęściej są to wejścia przeznaczone do podłączenia mikrofonu, CD-ROM-u oraz innego urządzenia zewnętrznego (np. telewizora). Wyjście i wejścia karty dźwiękowej pracują w trybie stereo.

Jakość dźwięku uzyskiwanego z karty dźwiękowej komputera jest porównywalna z jakością dźwięku na płytach CD.

5.31. Czym różnią się produkowane obecnie karty dźwiękowe?

Karty dźwiękowe różnią się jakością uzyskiwanego dźwięku. Większość z nich to karty 16-bitowe. Czasami można się jeszcze spotkać ze starymi kartami 8-bitowymi, oferującymi niską jakość dźwięku.

Karty mają różne częstotliwości próbkowania dźwięku. Typowe częstotliwości to: 8000, 11025, 16000, 22050, 32000, 32075, 44100 i 48000 próbek/sekundę. Im wyższa częstotliwość próbkowania, tym lepsza jakość dźwięku.

Należy pamiętać, że nie zawsze stosowanie zapisu stereo z najwyższą częstotliwością próbkowania jest sensowne. Jeśli rejestrujemy dźwięk niskiej jakości i zależy nam, aby powstający zapis nie zajmował zbyt wiele miejsca, należy zmniejszyć częstotliwość próbkowania i rozdzielczość próbek.

Karty dźwiękowe różnią się także dodatkowymi układami korekcji dźwięku oraz wytwarzania efektu echa. Jakość tych układów jest różna dla kart różnych producentów.

5.32. Czy można wydawać komputerowi polecenia głosowe?

Przy zastosowaniu standardowej karty dźwiękowej, mikrofonu i specjalnego oprogramowania jest to możliwe, ale tylko w ograniczonym zakresie. Sterowanie głosem ogranicza się do wydawania krótkich komend, które komputer porównuje z wprowadzoną wcześniej bazą słów wzorcowych. Program rozpoznawania głosu należy

najpierw nauczyć rozkazów dźwiękowych, jakie ma akceptować, i przypisać im określone zadania do wykonania.

5.33. Czy komputer analizuje obrazy?

Przeprowadzanie analizy obrazu przez komputer jest możliwe. System rozpoznawania obrazów identyfikuje obiekty na podstawie ich szczególnych cech fizycznych. Znalazło to zastosowanie np.: w rozpoznawaniu pisma, systemach wyszukiwania określonej twarzy w tłumie, systemach rozpoznawania typu samolotu na podstawie jego kształtu.

Pomimo opracowywania coraz to nowszych i bardziej złożonych algorytmów analizy obrazu systemy komputerowe nie dorównują umiejętnościom człowieka w tej dziedzinie.

5.34. Czy możliwe jest rozpoznawanie pisma?

Rozpoznawanie pisma jest możliwe. Aby to zrealizować, należy zamienić dokument papierowy na postać elektroniczną (służy do tego skaner). Następnie elektroniczny obraz dokumentu jest poddawany analizie przez specjalny program. Do analizy pisma drukowanego jak i odręcznego używa się programów OCR (ang. *Optical Character Recognition*). Działanie tych programów polega na zamianie znaków drukowanych lub pisanych ręcznie na ich odpowiedniki zapisane cyfrowo w pamięci komputera.

Programy rodziny OCR coraz częściej są przystosowane do uczenia się, gdy mają problem z analizą określonego znaku. Takie rozwiązanie pozwala uniknąć popełnienia błędów w dalszej części analizowanego tekstu.

5.35. W jaki sposób niewidomi odbierają wyniki pracy komputera?

Nowe systemy operacyjne (patrz rozdz. 9) i programy są przystosowane do współpracy także z osobami niepełnosprawnymi. Komunikacja z osobami niewidomymi odbywa się głosowo. Informacja wyświetlana na ekranie jest przetwarzana na postać mówioną z wykorzystaniem syntezy mowy. Specjalnie dla nich pojawiają się „udźwiękowane” wersje programów lub nakładek na istniejące już programy. Edytory tekstów umożliwiają odsłuchanie tego, co zostało wprowadzone oraz poprawienie błędów. Nie ma również problemu z odczytaniem zawartości poczty elektronicznej czy zawartości stron WWW.

Podobnie przedstawia się problem osób niepełnosprawnych ruchowo. To z myślą o nich właśnie tworzy się oprogramowanie analizy mowy, tak by było możliwe sterowanie pracą programów bez pomocy rąk.

5.36. Jakie znamy tryby transmisji danych?

Jeśli jako kryterium przyjmimy rytm czasowy, transmisję można podzielić na dwa tryby: asynchroniczną i synchroniczną. W transmisji asynchronicznej znaki są przesyłane ze zmiennymi w czasie odstępami. Kod znaku jest poprzedzany „bitem START”, który informuje odbiornik, że kolejne bity stanowią informację. „Bit STOP” kończy informację. Transmisja asynchroniczna jest stosowana przy niezbyt dużych szybkościach transmisji łącza, gdy informacje są przesyłane w nieregularnych odstępach. Przy transmisji synchronicznej nadajniki i odbiorniki są zaopatrzone w zegary taktujące i dlatego bity START i STOP są zbędne. Transmisję synchroniczną stosuje się przy dużych szybkościach.

Stosując inne kryterium podziału, można wyróżnić następujące rodzaje transmisji:

- jednokierunkowa (ang. *simplex*)
- naprzemienna (ang. *half duplex*)
- jednoczesna (ang. *duplex*).

W transmisji jednokierunkowej informacje można przysyłać tylko w jednym kierunku. Z tego sposobu korzysta się rzadko, gdyż odbiornik nie ma możliwości potwierdzenia nadajnikowi poprawności odbioru.

W transmisji naprzemiennych informacje można przysyłać w obu kierunkach, lecz nie jednocześnie. Role urządzeń końcowych zmieniają się naprzemiennie, tzn. nadajnik staje się odbiornikiem i odwrotnie.

Łącze jednoczesne może służyć do równoczesnej transmisji w obu kierunkach i jest równoważne parze łączy jednokierunkowych pracujących w przeciwnych kierunkach.

5.37. Jakimi parametrami określa się łącze transmisji danych?

Rozróżniamy następujące istotne parametry łącza transmisji danych:

- stopa błędów,
- szybkość modulacji
- szybkość transmisji.

Stopa błędów to stosunek liczby błędnie odebranych elementów do całkowitej liczby nadanych elementów. Stopa błędów nie powinna przekraczać 10^{-6} .

Szybkość modulacji określa maksymalną liczbę binarnych elementów sygnału, jakie mogą być przesłane w ciągu jednej sekundy i jest określana w bodach. Na przykład gdy czas trwania jednostkowego sygnału binarnego $\tau = 20$ ms szybkość modulacji jest równa:

$$V_m = \frac{1}{\tau} = \frac{1}{20 \cdot 10^{-3} s} = 50 \text{ bodów}$$

Szybkość transmisji jest to liczba bitów informacji przesyłanych w ciągu sekundy. Dla rozpatrywanego przypadku wysyłanie sygnału odpowiadającego jednemu znakowi trwa 150 ms (START – 20 ms, 5 bitów po 20 ms, impuls STOP 20 ms). W ciągu jednej sekundy może być przesłane $1/0,150$ znaków, lecz każdy znak zawiera 5 bitów informacji, a więc:

$$V_t = 5 \cdot \frac{1}{0,150} \frac{\text{bit}}{s} = 33,3 \frac{\text{bit}}{s}$$

Różnica między V_m oraz V_t wynika z występowania w sygnale dodatkowych informacji sterujących.

5.38. Co to jest modem i jak się porozumiewa z komputerem?

Modem jest urządzeniem umożliwiającym transmisję danych cyfrowych w łączu analogowym. Słowo modem powstało przez połączenie słów modulacja i demodulacja. Obecnie modemem nazywa się powszechnie urządzenie służące do transmisji danych, przy użyciu standardowego łącza telefonicznego.

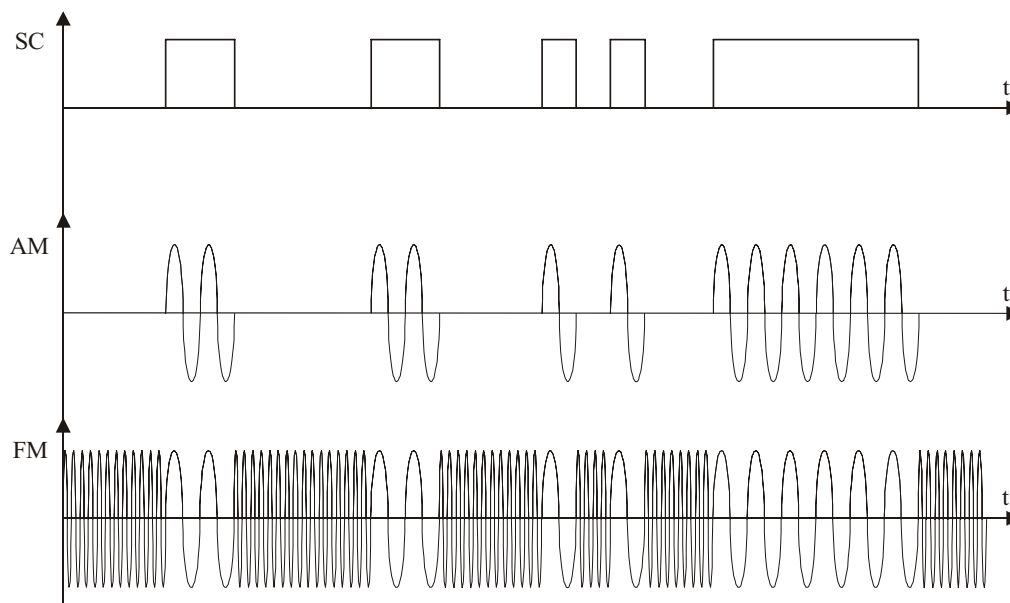
Połączenie modemu z komputerem jest wykonywane wg standardu RS-232. Modem może być również wykonany jako urządzenie wewnętrzne, w postaci karty rozszerzeń, instalowanej bezpośrednio w slotcie płyty głównej. Konfigurowanie, ustalanie parametrów transmisji oraz inicjowanie połączenia jest wykonywane poprzez standardowe polecenia AT (ang. *attention command*).

5.39. Jakie środki techniczne są niezbędne, aby można korzystać z modemu?

Aby skorzystać z modemu, należy podłączyć go do komputera oraz do publicznej sieci telefonicznej. W komputerze, który ma współpracować z modemem, musi być zainstalowane odpowiednie oprogramowanie komunikacyjne, które umożliwi nawiązanie połączenia i zapewni transmisję danych. W obecnych systemach niezbędne oprogramowanie jest już standardowo zainstalowane, umożliwiając współpracę z modemem. Jedyne, co należy zainstalować, to sterownik (program) do konkretnego typu modemu, stanowiący połączenie między modemem danego producenta, a uniwersalnym oprogramowaniem systemu.

5.40. Dlaczego do połączeń na duże odległości używa się modemów?

Wynika to z właściwości linii wykorzystywanych do transmisji np. linii telefonicznych. Linie te są przeznaczone do transmisji sygnałów analogowych o ograniczonej częstotliwości. Linia o dużej długości przedstawia sobą pewną rezystancję, ma określoną indukcyjność i pojemność. Sygnał cyfrowy, po przejściu przez nią jest niezrozumiały dla odbiornika. Taki kanał transmisyjny nie nadaje się do bezpośredniej transmisji sygnałów cyfrowych. Dlatego stosuje się modulację amplitudy AM i modulację częstotliwości FM.



Rysunek 5.1. Przykłady modulacji

5.41. W jaki sposób modemy różnych producentów współpracują ze sobą?

Współpraca różnych modemów jest możliwa dzięki opracowaniu standardów V.xx. Standardy te definiują sposoby transmisji danych z określonymi prędkościami, a także sposoby korekcji błędów i metody sprzętowej kompresji.

Dla przykładu można podać, że standard V.32 opisuje transmisję danych z maksymalną szybkością transmisji 9600 bps (bitów/sekundę), V.32bis – 14400 bps, V.34 – 28800 bps, V.90 – 56000 bps.

Standard V.42 definiuje sposób korekcji błędów, zapewniający poprawną transmisję danych, zaś standard V.42bis – sprzętową kompresję danych.

Dzięki dostosowaniu się wielu producentów do standardów międzynarodowych istnieje możliwość współpracy różnych modemów. Dodatkową ważną zaletą jest to, że nowe modemy, pracujące w nowszych standardach, są przystosowane do współpracy z modemami wolniejszymi. W takim wypadku szybkość transmisji łącza jest ograniczana przez wolniejszy z modemów.

5.42. Jakie stosuje się metody zabezpieczeń przed zniekształceniem przesyłanej informacji?

Aby zabezpieczyć informację przed zniekształceniem, stosuje się urządzenia detekcyjne i korekcyjne. W urządzeniach detekcyjnych błąd jest odnotowywany w umowny sposób – znacznik przy sygnale błędnym. Systemy detekcyjne są uzależnione od inwencji człowieka. W urządzeniach korekcyjnych do informacji właściwej dodaje się bity kontrolne. Jeśli wystąpi błąd podczas transmisji, to na podstawie bitów kontrolnych, błąd ten może być skorygowany. Przy odbiorze informacji bity kontrolne są pomijane.

Urządzenia zabezpieczające mogą pracować ze sprzężeniem decyzyjnym bądź ze sprzężeniem informacyjnym. Decyzja zwrotna może nakazać nadajnikowi powtórzenie błędnie nadanego znaku (gdy błąd nie może zostać skorygowany przez odbiornik) bądź kontynuowanie transmisji (błąd jest skorygowany przez odbiornik). W przypadku sprzężenia informacyjnego do nadajnika zostaje przesłana błędnie wysłana informacja. Identyfikacja błędu przez porównanie następuje wówczas w nadajniku. Systemy ze sprzężeniem zwrotnym stosuje się w łączach dwukierunkowych.

5.43. Jakie stosuje się metody kompresji danych w modemach?

Kompresję w modemach stosuje się w celu usunięcia nadmiarowości przesyłanej informacji. Pozwala to przyspieszyć transmisję danych (np. jeżeli dany znak powtarza się kilkukrotnie, można przesłać kod tego znaku wraz z liczbą powtórzeń).

W modemach korzysta się z algorytmu kompresji, opisanego w standardzie V.42bis. Algorytm kompresji V.42bis jest uzupełnieniem protokołu korekcji błędów V.42. Algorytm V.42bis używa kompresji LZW (od nazwisk autorów: Lempel, Ziv, Welch). Najczęściej kompresja jest prowadzona sprzętowo, przez sam modem, nie obciążając komputera.

Nie wszystkie przesyłane informacje opłaca się kompresować. Dobrze kompresują się np. pliki tekstowe, natomiast nie nadają się do tego pliki poddane już wcześniej kompresji np. ZIP czy MP3. Próba poddania takich danych ponownej kompresji odnosi odwrotny skutek, gdyż objętość ich wzrasta. Dzieje się tak, ponieważ algorytm kompresji nie jest w stanie bardziej skompresować danych, a jedynie wprowadza nowe, dodatkowe informacje. Algorytm V.42bis kontroluje przesyłane informacje i decyduje o włączeniu lub wyłączeniu kompresji sprzętowej.

5.44. Co podłączamy do złącza PS2?

Złącze PS2 jest przeznaczone do podłączania klawiatury i myszy. Złącze to zostało wprowadzone do komputerów IBM w 1987 roku. Jednak dopiero w ostatnich latach inni producenci zaczęli je powszechnie stosować. Wtyk w złączu PS2 jest wykonany jako sześciostykowy MINI-DIN. PS2 jest synchronicznym portem szeregowym, pracującym w standardzie TTL (ang. *Transistor-Transistor Logic*).

W chwili obecnej, w nowych komputerach, zarówno mysz, jak i klawiatura są przyłączane złączem PS2. Coraz trudniej znaleźć klawiaturę wyposażoną w stare złącze AT czy mysz dołączaną przez port szeregowy RS-232.

5.45. Jak działa i jak się komunikuje z komputerem mysz?

Zadaniem myszy jest zamiana ruchu ręki operatora na ruch kursora na ekranie. Sposób wykrywania ruchu zależy od typu myszy.

W myszach mechaniczno-optycznych (najczęściej spotykanych) ruch ręki wraz myszą jest przekazywany poprzez specjalną kulkę, która tocząc się po podkładce, porusza dwiema rolkami znajdującymi się wewnątrz myszki. Rolki są ułożone prostopadłe do siebie i wykrywają ruch w poziomie i w pionie. Ruch rolek powoduje przerywanie promienia świetlnego, co z kolei jest zamieniane na zmiany sygnału elektrycznego. Sygnał elektryczny podlega dalszej obróbce i poprzez łącze szeregowe jest przesyłany do komputera. Mysz pozwala określać przemieszczenie względne, nie zaś pozycję bezwzględną, jak ma to miejsce w digitizerze.

W myszach optycznych nie ma elementów mechanicznych (kulki i rolek). Mysz optyczna musi być poruszana po specjalnej podkładce. Ruch jest wykrywany poprzez detektory analizujące światło odbijane od pokratkowanej podkładki. Następnie sygnał z detektorów podlega przetworzeniu i jest przesyłany łączem szeregowym do komputera. Mysz tego typu, ze względu na brak elementów mechanicznych, jest trwalsza i wygodniejsza w obsłudze.

Połączenie myszy z komputerem jest wykonane przez standardowe złącze szeregowe RS-232 lub PS2.

5.46. Czym można zastąpić mysz?

Oprócz myszy, do sterowania ruchem kursora, używany jest trackball, touchpad, touchscreen czy digitizer.

Trackball jest zwany także manipulatorem kulkowym lub „kotem”. Zasada działania jest podobna jak w myszy z kulką. Różnica polega na tym, że kulka jest poru-

szana bezpośrednio palcami, a nie poprzez ruch po podkładce. Jest to szczególnie wygodne, gdy nie ma zbyt dużo miejsca na poruszanie myszą. Tracball znalazł głównie zastosowanie w komputerach przenośnych oraz jako mysz wbudowana w klawiaturę w komputerach stacjonarnych.

Touchpad to specjalna płytką rezystancyjna wrażliwa na dotyk. Przekazywanie ruchu polega na wykrywaniu zmian rezystancji na jej powierzchni. Jest to urządzenie bardzo wygodne w użyciu. Touchpad znalazł zastosowanie, podobnie jak tracball, w komputerach przenośnych. W najnowszych komputerach tego typu stosuje się touchpad zamiast tracballa. Coraz częściej można spotkać klawiatury do komputerów stacjonarnych z wbudowaną płytką dotykową emulującą mysz.

Touchscreen jest nazwą monitora wrażliwego na dotyk, który wykonuje się poprzez naklejenie na jego powierzchni specjalnej folii. Jest ona wrażliwa na dotyk i przekazuje do komputera bezwzględną pozycję wskazanego punktu oraz siłę nacisku.

5.47. Dlaczego monitory LCD stają się coraz bardziej popularne?

Monitory LCD (ang. *Liquid Crystal Display*) są kolejną generacją monitorów komputerowych. W porównaniu z popularnymi monitorami wyposażonymi w kineoskop, CRT (ang. *Cathode Ray Tube*), monitory LCD mają wiele zalet:

- wysoką jakość obrazu (kontrast i jasność)
- doskonałą geometrię obrazu
- brak emisji szkodliwego promieniowania
- brak migotania
- znacznie mniejsze rozmiary (głębokość) i waga
- mniejszy pobór prądu
- odporność na zakłócenia elektromagnetyczne.

Jedyną wadą jest cena. Monitory LCD mogą być nawet kilka razy droższe od monitorów CRT. Z tej też przyczyny głównym zastosowaniem ekranów LCD są komputery przenośne i urządzenia specjalizowane.

5.48. W jakim celu stosujemy ekrany dotykowe?

Ekran dotykowe służą ułatwieniu pracy użytkownika. Ze względu na znacząco powierzchnię styku palca z ekranem nie nadają się do prac wymagających precyzji. Poprzez dotyk obszaru ekranu, który odpowiada wybranemu elementowi, można sterować nim i uruchamiać aplikacje tak jak za pomocą myszy. Ekran dotykowy posiada

specjalną czułą na dotyk lub ciepło folię. Do komputera dodatkowo musi być podłączone urządzenie, które odbiera sygnały z folii, przetwarza je i przesyła do komputera. Na podstawie przesłanych współrzędnych określany jest obiekt, który znajdował się na ekranie i podejmowana jest decyzja o rozpoczęciu zadania.

Ekran dotykowy najczęściej można spotkać na lotniskach, w bankach, w sterownikach urządzeń przemysłowych.

5.49. Jakie zalety mają kody kreskowe?

Kod kreskowy składa się z równoległych ciemnych i jasnych pól o ściśle określonej szerokości. Odpowiednie złożenia tych kresek tworzy ciąg cyfr, który odpowiada kodowi produktu.

Zalety wynikające z zastosowania kodu kreskowego są następujące:

- skrócenie czasu identyfikacji produktu
- eliminacja użycia klawiatury
- oszczędność czasu przeznaczonego na inwentaryzację
- eliminacja omyłności człowieka,

Najczęściej kod kreskowy spotkać można na produktach handlowych, jest to kod 13-cyfrowy, a dla towarów o małych rozmiarach bywa to kod 8-cyfrowy.

Oprócz kodów jednowymiarowych istnieją też kody dwuwymiarowe, które tworzą unikatowy wzór składający się z kilkudziesięciu wierszy i kolumn. Kody tego typu umożliwiają zapisanie i szybki odczyt do kilku tysięcy znaków.

5.50. Jak są przesyłane dane z klawiatury do komputera?

Klawiatura jest połączona z komputerem łączem szeregowym, pracującym w trybie synchronicznym. Wewnątrz klawiatury znajduje się mikrokomputer jednokładowy, wykrywający naciskanie, trzymanie i puszczenie klawisza. Mikroprocesor przetwarza informacje o wciśniętym klawiszu na specjalne kody (ang. *scan codes*) i przesyła je łączem szeregowym do komputera. Przy wciskaniu klawiszy wysyłany jest do komputera przeważnie jeden bajt. Występują też klawisze, po naciśnięciu których wysyłane są dwa bajty. Istnieje też jeden klawisz (Pause/Break), po wciśnięciu którego przesyła się aż osiem bajtów.

5.51. Do czego służy i czym się charakteryzuje złącze PCMCIA?

Złącze PCMCIA (ang. *Personal Computer Memory Card International Association*) służy do przyłączania do komputera kart rozszerzeń. Jako karty PCMCIA wykonuje się moduły pamięci, karty sieciowe, modemowe, karty SCSI. Zaletą kart PCMCIA jest możliwość instalacji podczas pracy komputera, a także łatwa ich wymiana. W złącza PCMCIA są wyposażone praktycznie wszystkie komputery przenośne i niektóre stacjonarne. Karty PCMCIA są kartami 16-bitowymi lub 32-bitowymi.

PCMCIA zawiera 68-stykowe złącze. PCMCIA może obsłużyć karty wymagające dużych szybkości transmisji (100 Mb/s karty Ethernet). Karta PCMCIA jest zasilana poprzez złącze napięciem 3,3 V. Istnieją trzy typy kart PCMCIA: Typu I o grubości 3,3 mm, Typu II o grubości 5,0 mm oraz Typu III o grubości 10,5 mm. Karty wszystkich typów mają standardowe wymiary 85,6 mm na 54,0 mm.

5.52. Co to jest USB?

USB (ang. *Universal Serial Bus*) jest interfejsem szeregowym, przeznaczonym do podłączania np. drukarek, myszy, głośników, skanerów, kamer cyfrowych, pamięci zewnętrznych czy szybkich modemów. Jest to więc złącze uniwersalne. Do złącza USB można dołączać i odłączać urządzenia podczas pracy komputera. Standard ten przewiduje przepływ danych z szybkością transmisji do 12 Mb/s (full speed z ekranowanym kablem). Jeśli nie jest wymagana tak wysoka szybkość transmisji, można w celu zmniejszenia kosztów wykonania urządzenia ograniczyć ją do 1,5 Mb/s (low speed). Do portu można podłączyć jednocześnie do 127 urządzeń, stosując koncentratory USB. Maksymalna długość kabla w tym standardzie wynosi pięć metrów. Jeśli potrzebna jest większa długość, wówczas co pięć metrów należy stosować koncentratory. Złącze USB może mieć 4, 5, 8, 10 lub 16 wyprowadzeń.

5.53. Jakie wyróżniamy standardy magistral i złączy kart rozszerzeń?

Karta rozszerzeń jest to moduł elektroniczny, który umieszcza się w specjalnych złączach płyty głównej. Służy ona do zwiększenia funkcjonalności komputera. Spotkać można złącza typu: ISA, EISA, VLB, PCI oraz AGP. Jednak rozwiązaniami powszechnie spotykanymi są ISA, PCI i AGP.

Standard ISA (ang. *Industry Standard Architecture*) opiera się na 16-bitowej magistrali, pracującej z częstotliwością taktowania 8 MHz. Magistrala PCI (ang. *Peripheral Component Interconnect*) jest 32-bitowa i umożliwia transfer informacji z prędkością do 132 MB/s, przy taktowaniu 33 MHz.

Magistrala ISA nie nadaje się do szybkiego przesyłania dużych ilości informacji, ponieważ jest zbyt wolna. Nowym standardem, zapewniającym większe szybkości transmisji jest magistrala PCI. Jej powstanie było spowodowane koniecznością szybkiej komunikacji urządzeń i procesora. Komunikacja procesora z nową szyną PCI odbywa się przez specjalny mostek. Rozwiązanie takie pozwala na komunikację urządzeń PCI bez zabierania czasu procesorowi.

5.54. Jakie zadania spełnia sterownik sprzętowy dysku?

Sterownik sprzętowy (ang. *controller*) służy do dołączenia dysku do magistrali, a następnie poprzez magistralę i adapter do komputera. Jest on zintegrowany z dyskiem. Najczęściej spotykane są dwa rodzaje magistral dyskowych: IDE oraz SCSI.

IDE (ang. *Integrated Drive Electronics*) to zintegrowany sterownik dysku. Standard ten zakłada przeniesienie całej „inteligencji” na elektronikę wbudowaną w napęd dyskowy. Ze sterownikiem IDE współpracuje adapter, wbudowany w płytę główną komputera lub zainstalowany jako dodatkowa karta rozszerzeń. Standardowy sterownik IDE miał ograniczenie pojemności dysków do 528 MB. Powstała nowa odmiana standardu – EIDE (ang. *Enhanced IDE* – rozszerzony IDE). Do magistrali IDE można dołączyć jedno lub dwa urządzenia ze sterownikami IDE. Urządzeniami tymi, oprócz dysków twardych, mogą być: napędy CD-ROM, napędy taśmowe, napędy dysków wymiennych. Jeśli do magistrali dołączamy dwa urządzenia, to jedno z nich jest nadrzędne (ang. *master*), a drugie podrzędne (ang. *slave*). We współczesnych komputerach są zintegrowane z płytą główną dwa takie adaptory. Standardowo więc, do komputera można dołączyć cztery takie urządzenia.

SCSI (ang. *Small Computer System Interface*) rozpowszechnił się jako standard przy budowie serwerów i obsłudze dużych dysków. Poprzez magistralę SCSI dołącza się również inne urządzenia, przekazujące duże ilości danych np. streamery, napędy CD-ROM, skanery. Do jednego adaptera SCSI można dołączyć do siedmiu urządzeń.

Urządzenia wyposażone w sterownik SCSI są znacznie droższe niż pracujące pod kontrolą IDE. Wynika to ze znacznej złożoności sterownika. Zaletą urządzeń pracujących pod kontrolą SCSI jest ich większa wydajność w porównaniu z IDE.

5.55. Co należy rozumieć pod pojęciem chipset?

Chipset umożliwia komunikację poszczególnych elementów komputera i jest pośrednikiem pomiędzy procesorem, a współpracującymi z nim urządzeniami. Chipset jest to zestaw specjalizowanych układów scalonych (od jednego do czterech układów) o bardzo wysokim stopniu scalenia. Chipset pełni najczęściej następujące funkcje sterownika: procesora lub procesorów, pamięci, przerwań, kanałów DMA, magistrali

PCI, mostka PCI do magistrali ISA, czasomierza, głośnika systemowego, zarządzania energią, klawiatury, myszy, napędów dysków twardych i elastycznych, portów szeregowych i równoległych. Płyta główna jest miejscem skupiającym wszystkie te elementy. Każdy z nich wymaga specjalizowanego sterownika. Dawniej część sterowników znajdowała się na kartach rozszerzeń, część na płycie głównej.

Obecnie dąży się do wbudowania w chipset jak największej liczby sterowników tak, aby maksymalnie zmniejszyć liczbę układów scalonych montowanych na płycie głównej komputera.

5.56. Dlaczego magistrala PCI jest lepsza od ISA?

Omówienie magistrali ISA i PCI znajduje się w pytaniu 5.53. Szyna PCI jest taktowana z częstotliwością 33 MHz, a ISA – 8 MHz. PCI jest magistralą 32-bitową, natomiast ISA jest 16-bitowa. Szyny te różnią się także sposobem komunikacji z procesorem i pamięcią. Magistrala ISA znacznie obciąża procesor. W magistrali PCI zastosowane są bufony pamięci izolujące procesor od urządzeń zewnętrznych, które mogłyby spowolnić jego pracę.

5.57. Czym różnią się magistrale IDE i SCSI?

Problem ten został już poruszony w pytaniu 5.54. Do różnic należy zaliczyć:

- protokół komunikacji urządzeń z adapterem
- fizyczną budowę złącza
- liczbę dołączanych jednocześnie urządzeń
- złożoność
- wydajność i uzyskiwane szybkości transmisji
- obciążenie procesora
- koszt urządzeń.

Urządzenia SCSI są znacznie droższe i bardziej złożone od IDE, lecz zapewniają większą wydajność, większą elastyczność i mniej obciążają procesor. Magistrale SCSI są powszechnie używane w komputerach wymagających dużych szybkości transmisji, gdzie należy dołączać więcej urządzeń do wspólnej magistrali. W standardzie SCSI urządzenia mogą przekazywać informacje pomiędzy sobą, bez obciążania procesora. Standard IDE znalazł zastosowanie w komputerach przeznaczonych do zastosowań biurowych i domowych, natomiast SCSI głównie w serwerach i wydajnych komputerach.

5.58. Do czego służy port AGP?

AGP (ang. *Accelerated Graphics Port*) jest nowym interfejsem kart graficznych. AGP powstał, ponieważ szybkość transmisji oferowana przez PCI okazała się niewystarczająca dla grafiki. W standardzie AGP dodano szybką, bezpośrednią magistralę między chipsetem a sterownikiem grafiki. W ten sposób zmniejszono obciążenie szyny PCI. Poprzez zastosowanie interfejsu AGP zwiększyła się szybkość transmisji ze 132 MB/s, dla PCI, do 528 MB/s, dla AGPx2, przy częstotliwości taktowania 66 MHz. Dodatkową zaletą standardu AGP jest umożliwienie karcie grafiki bezpośredniego pobierania danych z pamięci komputera, bez potrzeby ich kopiowania do pamięci karty.

5.59. Jaka rolę pełnią w systemie komputerowym sterowniki programowe?

Sterownik programowy urządzenia (ang. *driver*) to niewielki program, pełniący rolę pośrednika pomiędzy urządzeniem wejścia/wyjścia, a systemem operacyjnym (patrz rozdz. 9). Sterowniki są stosowane do takich urządzeń, jak: karty sieciowe, karty graficzne, karty muzyczne, napędy dysków twardych i elastycznych, CD-ROM, klawiatura, mysz, monitory, drukarki, skanery itd.

Dzięki zastosowaniu sterowników do konkretnych urządzeń, system operacyjny odwołuje się w standardowy sposób do sterownika, a nie bezpośrednio do urządzenia danego producenta, z którym komunikacja może odbywać się w inny, niestandardowy sposób. Stosowanie specjalizowanych sterowników, dostarczanych przez producentów urządzeń, pozwala na pełne wykorzystanie ich możliwości.

6. Architektura systemów komputerowych

6.1. Co spowodowało, że zamiast określenia „komputer” zaczęto używać terminu „system komputerowy”?

Ciągły wzrost zastosowań komputerów stymulował ich rozwój, zarówno w aspekcie funkcjonalnym, jak i technicznym. W ten sposób komputer przekształcał się z urządzenia elektronicznego o dość jednolitej i regularnej strukturze w zespół wzajemnie powiązanych ze sobą urządzeń. Podzespoły wchodzące w skład komputera są wyposażane przez producenta w oprogramowanie, zapewniające jego efektywne wykorzystanie.

Ten zbiór urządzeń wraz z oprogramowaniem zaczęto nazywać systemem komputerowym. Zmiana terminologiczna wynikała nie tylko ze wzrostu stopnia złożoności struktury komputera. Wiązała się także z faktem pojawienia się nowej dziedziny nauki tzw. techniki systemów, a zatem z próbą nowego kompleksowego (systemowego) spojrzenia na komputer.

Pojęciem system komputerowy określa się sprzęt (ang. *hardware*) wraz z niezbędnym oprogramowaniem (ang. *software*).

6.2. Co należy rozumieć przez określenie „architektura” w odniesieniu do systemu komputerowego?

W pierwszym odruchu termin „architektura”, rozumiany tradycyjnie jako „sztuka projektowania, wznoszenia budowli...” [44], może dziwić i być mylący. Można bowiem pytać, co ma z tym wspólnego system komputerowy. Przecież nie chodzi tu chyba o „wystrój”, kształty i kolorystykę komputera. Słusznie – nie w tym znaczeniu używamy terminu „architektura”.

Za twórcę terminu „architektura systemu komputerowego” uważa się W. Bucholza redaktora pracy „Planning a Computer System” opublikowanej w 1962 roku. Architekturę systemu komputerowego definiują następujące cechy:

- lista rozkazów
- budowa słowa rozkazowego
- format danych
- sposób adresowania pamięci
- sposób współpracy z urządzeniami zewnętrznymi

- sposób reagowania na przerwania
- tryby użytkowania komputera.

Architektura systemu komputerowego to po prostu organizacja jego elementów.

System komputerowy można opisywać na różnych poziomach abstrakcji: funkcjonalnym, struktury i technologii. Architektura to sposób opisu systemu komputerowego na poziomie strukturalno-funkcjonalnym, a nie technologicznym.

6.3. Jak klasyfikujemy systemy komputerowe?

Systemy komputerowe można porównywać stosując różne kryteria. Jeśli jako kryterium przyjmuje się przeznaczenie systemu, wówczas wyróżnia się komputery:

- ogólnego przeznaczenia (konfiguracja tak dobrana, aby komputer mógł służyć wielu celom)
- specjalizowane (przeznaczone do wykonywania określonego zadania lub klasy zadań).

Kryterium klasyfikacji może być tryb przetwarzania informacji. Tryb ten wynika z powiązań funkcjonalnych, istniejących w systemie komputerowym. Pod tym względem systemy komputerowe można sklasyfikować następująco:

- systemy jednoprogramowe – jednozadaniowe (jednocześnie jest wykonywany tylko jeden program, wykonanie kolejnego programu jest możliwe po całkowitym zakończeniu programu poprzedniego)
- systemy wieloprogramowe – wielozadaniowe (wiele programów może być wykonywanych pozornie jednocześnie, wykonywanie wielu programów następuje na przemian)
- systemy wieloprocessorowe (zadania są dzielone na fragmenty, wykonywane równolegle przez oddzielne procesory)
- systemy wielodostępne – wieloużytkownikowe (możliwe jest korzystanie z systemu jednocześnie przez wielu użytkowników).

Kolejnym kryterium podziału systemów komputerowych, wiążącym się z gabarytami komputera, może być sposób jego użytkowania. Według tego kryterium wyróżnić można:

- komputery osobiste (biurowe, domowe, używane w jednym miejscu – stacjonarne)
- komputery przenośne (ang. *laptop*, *notebook*)
- komputery kieszonkowe (ang. *palmtop*).

6.4. Co to jest skalowalność?

Skalowalność (ang. *scalability*) systemu komputerowego to cecha, ważna zwłaszcza przy konstruowaniu systemów rozproszonych tzn. zbiorów komputerów połączonych za pomocą sieci i wyposażonych w odpowiednie oprogramowanie systemowe (rozproszone), polegająca na możliwości harmonijnego rozrastania się systemu w miarę upływu czasu i zwiększania liczby jego użytkowników bez konieczności rewolucyjnych zmian projektowych. Skalowalność systemów jest przedmiotem intensywnych badań. Podstawowy sposób osiągania skalowalności polega na zastosowaniu zwielokrotnień (wyposażania systemów w dodatkowe egzemplarze poszczególnych zasobów) usług systemu oraz nadmiarowości jego składowych. Rozproszony system plików ze zwielokrotnionymi serwerami plików jest tego dobrym przykładem. Problem skalowalności występuje w zagadnieniach z pozoru nie związanych bezpośrednio z informatyką. Przykładem źle skalowalnych dużych systemów są wielkie aglomeracje miejskie, a w nich drożność układów komunikacyjnych lub kwestie skutecznego usuwania śmieci. Tak jak w przypadku nadmuchiwania balonika, skalowalność każdego systemu informatycznego ma swoje granice.

6.5. Co oznacza pojęcie: wielowątkowość?

Wielowątkowość (ang. *multithreading*) systemu komputerowego jest to właściwość nowoczesnych systemów operacyjnych, w których poszczególne procesy mogą być wykonywane wielotorowo w postaci wątków. Wątek dzieli z innymi równorzędnymi wątkami przestrzeń adresową procesu, a w szczególności jego kod i dane, otwarte pliki i sygnały. Znaczne dzielenie zasobów przez wątki przyspiesza przełączanie kontekstu między wątkami tego samego procesu i obniża koszt tworzenia wątków w porównaniu z przełączaniem kontekstu między tradycyjnymi procesami.

Wielowątkowość pozwala wykonywać prace kooperatywne ekonomiczniej niż przy udziale wielu osobnych procesów, ponieważ jest realizowana we wspólnej przestrzeni adresowej, co ułatwia komunikację. Najczęściej pracę komputera organizuje się w ten sposób, że jeden procesor dzieląc systematycznie swój czas, wykonuje jednocześnie więcej niż jeden proces użytkowy.

6.6. Na czym polega wieloprogramowość (wielozadaniowość)?

W jednoprogramowych systemach komputerowych program zaczęty musi być wykonany do końca (np. pod kontrolą MS-DOS – rozdz. 9). W systemach takich nie jest możliwe przeplatanie się fragmentów różnych programów. Wieloprogramowość to taki tryb pracy systemu komputerowego, w którym poszczególne programy tworzą zbiór i nie muszą być wykonywane w nieprzerwany sposób. Wykonywanie tych pro-

gramów następuje etapami, według określonych zasad. Z punktu widzenia obserwatora ma się do czynienia z pozorną jednoczesnością wykonywania programów. Systemem takim jest np. Windows 95, 98.

Zasady określające etapowość, wynikają z chęci uzyskania maksymalnej wydajności systemu. Popularnym trybem pracy jest system z ustalonymi priorytetami. Pod pojęciem priorytetu rozumie się stopień ważności zadania. Wieloprogramowość może być także realizowana na zasadzie podziału czasu (ang. *time sharing*) dla poszczególnych użytkowników.

Zastosowanie systemu wieloprogramowego pozwala w większym stopniu wykorzystać zasoby sprzętu niż system jednoprogramowy. Podczas gdy jeden z programów korzysta z urządzeń wejścia/wyjścia, inny może używać procesora. Przy zastosowaniu takiego trybu pracy wykorzystanie zasobów procesora jest możliwe prawie w 100%. Niestety występuje tu pewne negatywne zjawisko. Część czasu procesora jest zużywana przez system operacyjny na przełączanie pomiędzy poszczególnymi programami.

System ten poprawia stopień wykorzystania sprzętu, zwiększa jednak niewygodę użytkownika, w porównaniu z systemem jednoprogramowym.

6.7. Jakie powinny być racjonalne reguły przydzielania priorytetów programom użytkowym w systemie wielozadaniowym?

Zadania, jakie są realizowane, to ciągi żądań obliczeń i transmisji wejścia/wyjścia. Mogą istnieć zadania o przewadze żądań obliczeń, jak i zadania o przewadze żądań transmisji. Ważne jest ich rozmieszczenie na osi czasu tak, by optymalnie wykorzystać zasoby systemu.

Jeśli założy się sytuację, w której najwyższy priorytet przydzielono zadaniu, które przede wszystkim wymaga obliczeń i tylko kilku transmisji, to zadanie takie będzie zajmowało procesor i nie dopuści innych programów do obliczeń. Stanie się tak dlatego, że zadania o niższym priorytecie będą miały możliwość skorzystania z zasobów procesora tylko wtedy, gdy program o wyższym priorytecie będzie dokonywał transmisji. Wynika z tego, że lepiej przyznać wyższy priorytet zadaniom, które zawierają więcej żądań transmisji. Najniższy zaś powinien być przyznany tym programom, które głównie dokonują obliczeń. Wówczas obliczenia w procesorze będą wykonywane w czasie, gdy inne programy dokonują transmisji wejścia/wyjścia.

6.8. Czy jest możliwe, aby w systemie wielozadaniowym dwa programy zostały wykonywane w takim samym czasie, jaki jest potrzebny do wykonania każdego z nich w systemie jednozadaniowym?

Przy wyjątkowo szczęśliwym zbiegu okoliczności czynność ta jest prawie wykonalna. Taki szczególny przypadek ma miejsce wówczas, gdy dwa programy korzystają z zasobów procesora w różnym czasie. Jest to możliwe, gdy jeden z programów prowadzi transmisję wejścia/wyjścia w czasie, gdy drugi program wykonuje obliczenia.

Sytuacja taka jest prawie możliwa, ponieważ należy uwzględnić jeszcze czas procesora potrzebny programowi koordynującemu na podział zasobów systemu pomiędzy oba te programy.

6.9. Jak działa system wielodostępny?

System wielodostępny pozwala korzystać z zasobów jednego komputera wielu użytkownikom. Przykładem systemu operacyjnego zapewniającego wielodostępność jest UNIX (patrz rozdz. 9). W systemie wielodostępnym każdemu użytkownikowi cyklicznie, na pewien czas są przydzielane zasoby komputera. Czas odpowiedzi systemu wielodostępnego na żądanie użytkownika powinien być na tyle krótki, aby użytkownik nie odczuwał istnienia innych współużytkowników. Z punktu widzenia użytkownika korzystanie z systemu wielodostępnego jest takie samo jak z systemu jednoprogramowego. Zaletą systemu wielodostępnego jest lepsze wykorzystanie zasobów, niż przy korzystaniu przez wielu użytkowników z wielu komputerów oddzielnie.

Systemy wielodostępne najczęściej pracują jako serwery sieciowe, pod kontrolą systemu UNIX. Aby korzystać z zasobów serwera sieciowego, należy mieć konto robocze. Po „zalogowaniu” (inaczej mówiąc – zameldowaniu) się na takie konto użytkownik może wykonywać obliczenia bądź korzystać z zasobów serwera.

Serwery pocztowe, www, ftp, systemy wyszukiwania informacji, bazy danych, np. Oracle, itp. (rozdz. 8) pracują właśnie jako systemy wielodostępne i umożliwiają dostęp do swych zasobów dużej liczbie użytkowników.

6.10. Co to są systemy komputerowe czasu rzeczywistego?

Systemami czasu rzeczywistego (ang. *real-time*) nazywa się systemy komputerowe przeznaczone do współpracy z urządzeniami wymagającymi bezzwłocznej reakcji ze strony komputera. Typowym przykładem może być system sterowania komputerowego procesem technologicznym lub pociskiem raketowym. Rolę urządzeń wejścio-

wych pełnią czujniki, mierzące parametry obiektów fizycznych, a rolę urządzeń wyjściowych układy przetwarzające informację na czynności regulujące działanie obiektów.

Szybkość wykonywania operacji w procesorze musi być wystarczająca do tego, by reakcje programu nadążały za zdarzeniami zewnętrznymi. W systemach czasu rzeczywistego (zwanymi także nadążnymi) istnieje rozbudowany, wielopoziomowy system przerwań. Systemy te posiadają zegary służące cyklicznemu generowaniu przerwań.

Niemalże „klasycznym” systemem komputerowym czasu rzeczywistego (lecz wyspecjalizowanym) może być współczesna cyfrowa centrala telefoniczna komutująca swoich abonentów (użytkowników).

6.11. Co spowodowało wprowadzenie systemów wieloprocesorowych?

Systemy wieloprocesorowe zostały stworzone w celu podniesienia niezawodności systemów komputerowych oraz zwiększenia szybkości obliczeń. Zwiększenie niezawodności wymagało wprowadzenia do systemów rezerwowych procesorów. Procesory te wykonywały zadania w przypadku awarii procesora głównego. Ponieważ szybkość obliczeniowa systemu jednoprocessorowego jest ograniczona (patrz pyt. 3.24, 3.25), sposobem na jej zwiększenie było umieszczenie kilku procesorów w jednym systemie. W taki sposób uzyskano zwiększenie szybkości, poprzez podział zadań na fragmenty, które są wykonywane równolegle przez oddzielne procesory.

Równoległe (współbieżne) wykonywanie zadań odbywa się w ten sposób, że zostają one podzielone na takie fragmenty, by jeden fragment zadania nie czekał na wyniki innego.

6.12. Co to są systemy wieloprocesorowe i jakie znamy ich struktury?

System nazywa się wieloprocesorowym, jeśli:

- posiada co najmniej dwa procesory
- wszystkie procesory mają dostęp do wspólnej pamięci i urządzeń wejścia/wyjścia
- istnieje system operacyjny przystosowany do obsługi wielu procesorów.

Można wyróżnić kilka odmian systemów wieloprocesorowych. W takich systemach poszczególne procesory mogą mieć własną pamięć podręczną lub własne urządzenia wejścia/wyjścia. Niektóre z nich mogą mieć dedykowane obsłudze konkretnych urządzeń.

Najprostszym rodzajem systemu wieloprocesorowego jest taki, który zawiera jedną magistralę. W tym systemie wszystkie procesory, pamięć i urządzenia wejścia/wyjścia są dołączone do wspólnej magistrali. Wadą tego rozwiązania jest mała skalowalność i w przypadku kilku procesorów może nastąpić przepełnienie. Efekt ten można zredukować przez wyposażenie każdej jednostki centralnej w pamięć podręczną.

Inną budową charakteryzują się pierścieniowe systemy wieloprocesorowe. W systemach tych procesory mają własną pamięć, podzieloną na część prywatną i część wspólną. Z części wspólnej mogą korzystać wszystkie procesory, zaś część prywatna jest wykorzystywana jako stos i pamięć zmiennych lokalnych. W odróżnieniu od systemów ze wspólną magistralą, systemy pierścieniowe nie mają scentralizowanej pamięci. Są one określane jako systemy z rozproszoną pamięcią wspólną.

Systemy wieloprocesorowe pierścieniowe i ze wspólną magistralą pracują w konfiguracjach do 64 procesorów. Wynika to z tego, że w pewnym momencie przepustowość magistrali osiąga wartość graniczną i wzrost liczby procesorów nie powoduje już wzrostu wydajności systemu. Systemy z większą liczbą procesorów można zbudować, zmniejszając liczbę komunikatów wysyłanych magistralą lub zwiększając jej przepustowość. Zwiększenie przepustowości może polegać na zbudowaniu kilku lub nawet całej siatki magistral.

Procesory z pamięcią współdzieloną można łączyć przełącznikiem krzyżowym. W tej architekturze używa się wielu modułów pamięci dołączanych do procesorów poprzez matrycę przełączników. Zaletą przełączników krzyżowych jest to, że procesory mogą korzystać z modułów pamięci równolegle. Do zalet należy zaliczyć również mniejszą liczbę konfliktów przy dostępie do magistrali. Wadą systemów z przełącznikiem krzyżowym jest koszt sprzętu.

6.13. Jakie są wady systemów wieloprocesorowych?

Wady systemów wieloprocesorowych wynikają ze znacznego stopnia złożoności systemów i konieczności dzielenia zadań na mniejsze fragmenty. Do głównych wad należą:

- złożone, kosztowne i trudne do zaprojektowania oprogramowanie
- czasochłonne i trudne sprawdzanie oprogramowania
- nieliniowy wzrost wydajności systemu, względem przyrostu kosztów elementów
- konflikty przy dostępie do magistral i do pamięci, wymagające stosowania układów arbitrażu.

6.14. Co to są komputery sterowane danymi?

Komputery sterowane danymi (ang. *dataflow*) mają niestandardową architekturę równoległą. W tradycyjnych komputerach von Neumanna, sterowanie wykonywaniem programu odbywa się poprzez użycie licznika rozkazów. W komputerach sterowanych przepływem danych nie ma licznika rozkazów, a polecenia są wykonywane, gdy dostępne staną się argumenty rozkazów.

Poszczególne etapy programu są wykonywane w niezależnych węzłach (operacja jest wykonywana po dostarczeniu wszystkich argumentów), a po wykonaniu zadania wynik jest przekazywany do kolejnego węzła. Liczba operacji, jakie mogą być jednocześnie wykonywane (liczba węzłów, w których są prowadzone obliczenia), zależy od konstrukcji sprzętu.

6.15. Do obsługi jakich architektur przystosowany jest system operacyjny Windows 95, Windows NT, a do jakich UNIX?

System operacyjny Windows 95 jest przystosowany do pracy na komputerach o 32-bitowej architekturze. Jest to system wielozadaniowy, który pozwala na wielowątkowe wykonywanie programów 32-bitowych. Windows 95 zapewnia zgodność z poprzednimi wersjami (poprawnie wykonuje programy 16-bitowe). Osiągnięto to jednak kosztem obniżenia wydajności systemu. System ten jest przygotowany do pracy w sieci jako stacja robocza.

Windows NT (ang. *New Technology*) jest również systemem 32-bitowym. Może pracować na komputerach wieloprocesorowych (jednak tylko do kilkudziesięciu procesorów). System ten pozwala nie tylko rozwiązywać różne zadania w różnych procesorach, lecz także dzielić jedno zadanie na wiele procesorów. Windows NT może służyć jako system operacyjny stacji roboczych lub dla serwerów.

UNIX jest przeznaczony do pracy na każdym sprzęcie (od stacji roboczych do superkomputerów). Potrafi wykorzystać zasoby komputerów wieloprocesorowych. UNIX jest systemem wielodostępnym i wieloprogramowym, najczęściej używanym jako system operacyjny serwerów sieciowych. Z punktu widzenia firm produkujących np. duże bazy danych zaletą jego jest skalowalność (patrz pyt. 6.4). Najczęściej spotykanymi odmianami tego systemu są:

- SCO Unix (dla procesorów Intel)
- Solaris (dla stacji roboczych Sun)
- Linux (bezpłatna wersja systemu UNIX, początkowo przeznaczona dla procesorów Intel, obecnie także dla procesorów Sparc i Alpha)
- HP-UX (dla komputerów Hewlett-Packard)
- Irix (dla komputerów Silicon Graphics).

6.16. Jakie przyczyny spowodowały powstanie sieci komputerowych?

Rozwój komputerów doprowadził do powstania wielu niezależnych ośrodków obliczeniowych. Istnienie dużych, oddzielnych ośrodków powodowało nieoptymalne wykorzystanie zasobów komputerów. Dążenie do udostępnienia możliwości obliczeniowych i zasobów komputerowych, rozmieszczonych na pewnym obszarze, doprowadziło do rozpoczęcia badań nad współpracą odległych systemów komputerowych. Taka współpraca umożliwiałaby korzystanie wielu użytkowników z zasobów wielu komputerów. Doprowadziło to do powstania koncepcji sieci komputerowych.

7. Sieci komputerowe – podstawy

7.1. Co to jest sieć komputerowa i jakie są cele jej użytkowania?

Pod pojęciem sieć komputerowa należy rozumieć zbiór komputerów i urządzeń końcowych połączonych medium transmisyjnym. Medium transmisyjne musi zapewniać bezpieczny dostęp użytkowników do wspólnych zasobów, takich jak informacje lub urządzenia wejścia/wyjścia.

Sieć komputerowa pozwala na łatwy przepływ informacji między komputerami. Każda podłączona jednostka może korzystać z zasobów lub też udostępniać swoje zasoby. Zasobami mogą być: bazy danych, aplikacje (programy użytkowe), drukarki, skanery, modemy itd. Zastosowanie specjalnych aplikacji pozwala na jednoczesne wykorzystanie mocy obliczeniowej wszystkich komputerów.

Biorąc pod uwagę obszar, na którym mogą znajdować się komputery wyróżnia się:

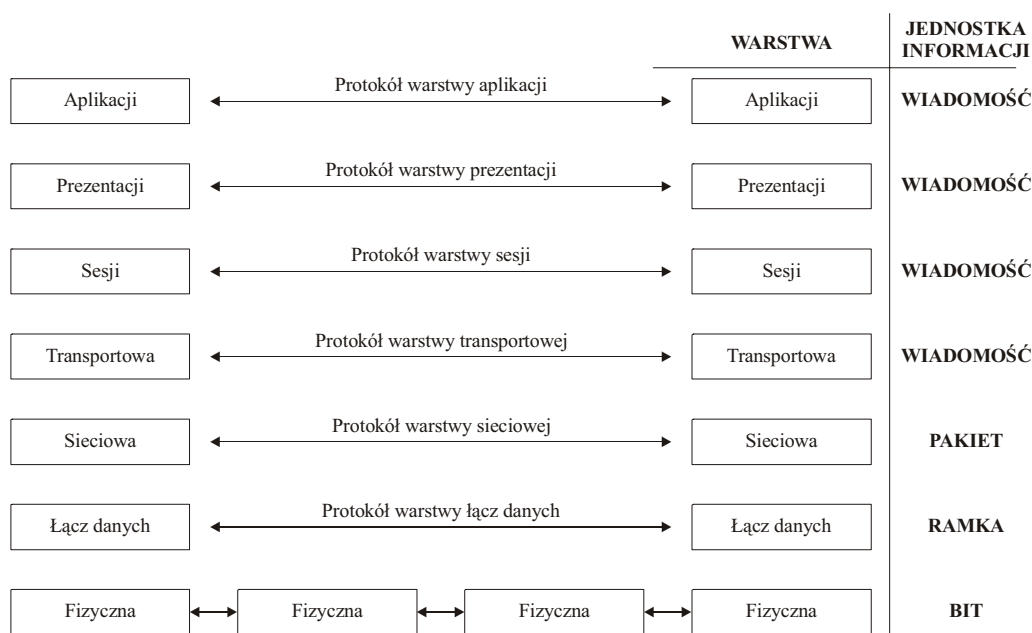
- sieci lokalne LAN (ang. *Local Area Network*)
- sieci miejskie MAN (ang. *Metropolitan Area Network*)
- sieci rozległe WAN (ang. *Wide Area Network*) np.: Internet.

7.2. Co to jest model warstwowy ISO-OSI?

Model warstwowy ISO-OSI (ang. *Open System Interconnection*) stworzony został przez organizację ISO (ang. *International Organization for Standardization*). Zawiera on zbiór zasad, według których komunikują się urządzenia sieciowe. Model ten opracowano dla sieci przełączających pakiety. Pakiet jest porcją informacji, składa się z nagłówka, informacji właściwej i informacji zabezpieczającej przed powstawaniem błędów w trakcie transmisji. Pojęcie „pakiet” ściśle związane jest z warstwą sieciową modelu ISO-OSI. W obrębie warstwy łącz danych operuje się pojęciem „ramka”. Podobnie jak pakiet posiada ona nagłówek i informacje.

Spośród wielu złożonych zagadnień komunikacji można wyodrębnić określone zadania, które mogą być rozwiązywane przez osobne moduły oprogramowania lub osobne urządzenia. Obiekty rozwiązujące podobne zadania tworzą warstwę. Model przewiduje siedem warstw, z których każda zbudowana jest na bazie warstwy poprzedniej i może komunikować się tylko z warstwami przylegającymi bezpośrednio do niej. Proces komunikacji realizowany jest na poziomie odpowiadających sobie warstw. Każda z warstw ma swój własny protokół komunikacyjny (patrz pyt. 7.9).

Proces komunikacji między dwoma programami odbywa się za pośrednictwem warstw modelu OSI. Informacje w postaci pakietów przekazywane są z warstwy wyższej do warstwy niższej. Warstwa niższa dodaje do odebranych informacji nagłówek, który ją identyfikuje. Następnie sytuacja powtarza się do momentu, kiedy pakiety dotrą do warstwy najniższej. Rzeczywista komunikacja odbywa się na poziomie warstwy najniższej. Nosi ona nazwę warstwy fizycznej. Warstwa fizyczna odbiorcy odbiera i interpretuje dane. Jeżeli zachodzi taka konieczność, są one pozbawiane nagłówka warstwy fizycznej i przekazywane do warstwy wyższej. Dzięki temu warstwa wyższa dostaje pakiet w takiej samej postaci, w jakiej został on wysłany z drugiego komputera.



Rysunek 7.1. Warstwowy model ISO/OSI

7.3. Jakie zadania spełniają poszczególne warstwy modelu ISO-OSI?

Model zawiera siedem warstw. Warstwa fizyczna odpowiada za transmisję sygnałów w sieci i konwertuje bity informacji na sygnały analogowe.

Warstwa łącz danych odbiera i konwertuje strumień bitów odbierane z urządzeń transmisyjnych. W tej warstwie następuje eliminacja zakłóceń powstających w kanale, a jeżeli zachodzi konieczność następuje retransmisja ramki. Warstwa łącz danych syn-

chronizuje prędkość przesyłania danych i umożliwia transmisję jednoczesną (ang. *duplex*) (patrz pyt. 5.36).

Warstwa sieciowa odpowiada za sterowanie działaniem podsieci transportowej. Przesyła informacje między węzłami i wyznacza trasy, którymi są one przesyłane. Proces znajdowania drogi w sieci nazywa się trasowaniem lub routingiem. Nie jest wymagane, aby pakiety pomiędzy ustalonymi punktami poruszały się za każdym razem po tej samej drodze. Warstwa sieciowa odpowiada za formowanie informacji w ramki, a po ich przesłaniu rozdziela je.

Warstwa transportowa zapewnia niezawodną komunikację pomiędzy abstrakcyjnymi portami komunikacyjnymi odległych komputerów, gwarantuje też poprawną kolejność dostarczania pakietów. Jeżeli odebrane informacje dotarły uszkodzone, żąda ich retransmisji.

Warstwa sesji pełni kontrolę nad nawiązywaniem i zrywaniem połączenia przez aplikacje. Po nawiązaniu stosownego połączenia warstwa pełni szereg funkcji zarządzających. Praktyczne implementacje warstwy sesji powinny zapewniać bezpieczeństwo przesyłanych danych.

Warstwa prezentacji odpowiada za obsługę formatów danych, kodowanie i dekodowanie zestawów znaków, kodowanie w celu utajnienia itp.

Warstwa aplikacji zapewnia komunikację programom użytkowym. Jej zadania są zróżnicowane i polegają na świadczeniu usług końcowych.

7.4. Dlaczego stosujemy model warstwowy?

Każda warstwa w modelu ma wyodrębnione zadania. Model nie narzuca fizycznej budowy poszczególnych warstw. Skupia się na zasadach współpracy pomiędzy nimi. Rozwiązanie takie sprawia, że fizyczna implementacja warstwy nie musi być identyczna u każdego producenta. Każdy może stosować własne rozwiązania firmowe.

Natomiast ściśle określone są zasady wzajemnej komunikacji między warstwami. Jeżeli są one zachowane, urządzenia różnych producentów mogą ze sobą poprawnie współpracować.

Twórcy oprogramowania sieciowego mogą również dzielić się między sobą zadaniami (warstwami), co jest ważne we współczesnej metodologii tworzenia oprogramowania w dużych zespołach programistów i współpracy z twórcami sprzętu komunikacyjnego. Wprowadza to w każdy twórczy proces pewien porządek (patrz pyt. 9.31 – dla porównania warstwowy model systemu operacyjnego).

7.5. Jakie zasoby można udostępnić w sieci komputerowej?

Rodzaj zasobu, jaki może być udostępniony w sieci komputerowej jest zależny od możliwości oprogramowania, które to zapewnia. Praktycznie wszystkie zasoby, jakie ma komputer mogą zostać udostępnione. Zasoby te można podzielić na trzy główne grupy:

- dane
- urządzenia
- moc obliczeniowa.

7.6. Co to jest segment sieci?

Segment sieci jest to odcinek przewodu wraz z dołączonymi do niego komputerami, ograniczony przez urządzenia aktywne separujące ruch w sieci lub terminatory. Urządzenia takie jak repeatery lub huby dzielą sieć na osobne odcinki, ale nie osobne segmenty. Natomiast switch, bridge lub router dzieli sieć na segmenty. Więcej informacji o tych urządzeniach w dalszej części rozdziału 7.

7.7. Jakie wyróżnia się rodzaje sieci komputerowych?

Ze względu na obszar, który obejmują swoim działaniem, sieci komputerowe dzielimy na: LAN (ang. *Local Area Network*), MAN (ang. *Metropolitan Area Network*), WAN (ang. *Wide Area Network*).

Sieci lokalne – LAN (mające na ogół jednego właściciela) – mogą łączyć ze sobą od kilku do kilkuset komputerów. Teren zajmowany przez LAN jest niewielki, zazwyczaj zamyka się w jednym budynku. Sieci lokalne mogą też łączyć komputery pracujące w kilku budynkach położonych blisko siebie. Połączenie wielu sieci LAN za pomocą szybkiego medium transmisyjnego daje w rezultacie sieć MAN.

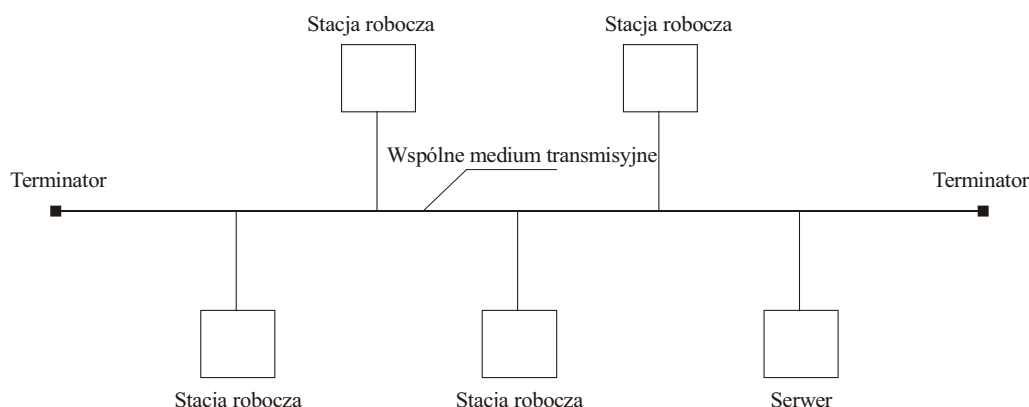
Sieci metropolitalne korzystają z bardzo szybkich magistral światłowodowych, łączą ze sobą wiele sieci lokalnych.

Połączenie wielu sieci metropolitalnych tworzy sieć WAN.

7.8. Jakie są topologie lokalnych sieci komputerowych?

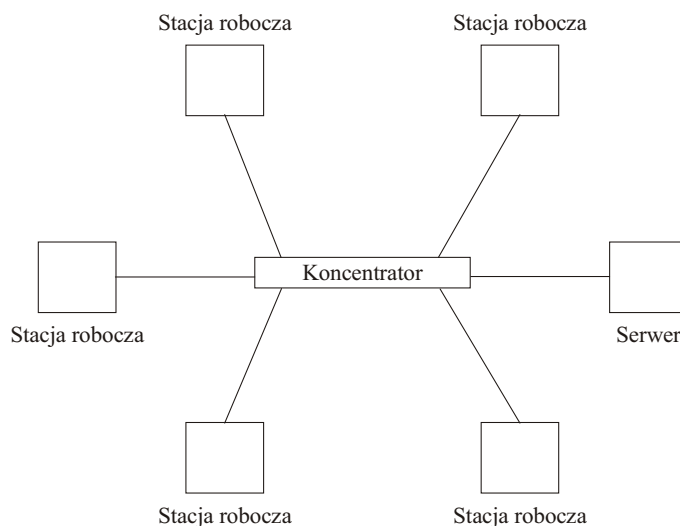
Sposób rozmieszczenia elementów w sieci komputerowej oraz rodzaj połączeń między nimi jest określany przez zastosowaną topologię. Lokalne sieci komputerowe są zazwyczaj elementami sieci złożonych.

Topologia liniowa (magistralowa) przedstawiona na rysunku 7.2 oznacza konfigurację, w której do pojedynczego medium transmisyjnego podłączone są wszystkie stacje robocze. Długość medium transmisyjnego w topologii liniowej jest ściśle określona i zależy od typu kabla zastosowanego do budowy. Nadawane sygnały docierają do wszystkich podłączonych stacji, ale w tej samej chwili może nadawać tylko jedna stacja. Zaletą tej topologii jest prosta budowa, niski koszt urządzeń i niewielka ilość zużytego kabla. Wadą jest mała koncentracja urządzeń, co utrudnia nadzór i naprawę uszkodzeń. W przypadku uszkodzenia kabla w jakimkolwiek punkcie cała sieć przestaje funkcjonować.



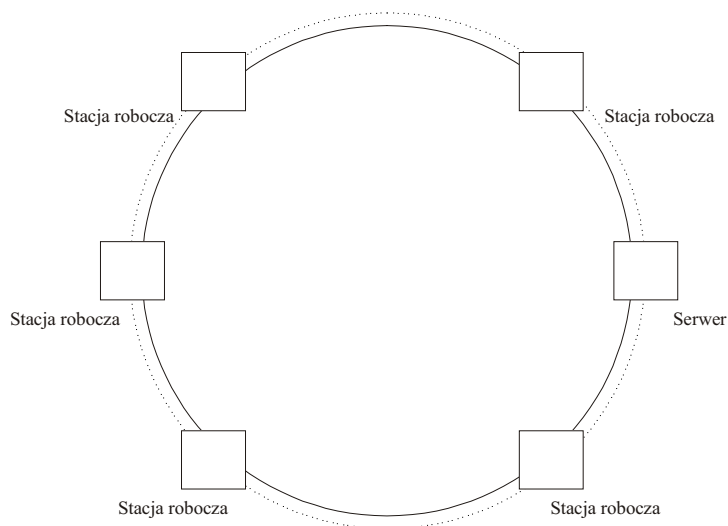
Rysunek 7.2. Topologia liniowa (magistralowa)

Topologia gwiazdy (rys. 7.3) zawiera centralne urządzenie, do którego podłączone są wszystkie elementy sieci. Wszystkie połączenia odbywają się za pośrednictwem koncentratora. W zależności od tego, czy urządzenie koncentrujące jest „inteligentne”, sygnały wysyłane przez jedną stację mogą trafiać do wszystkich końcówek lub tylko do jednej, dla której są przeznaczone. Odległości stacji roboczych uzależnione są od typu zastosowanego okablowania. Zaletą jest łatwość zarządzania, monitorowania i naprawy. Awaria jednej stacji roboczej lub jej okablowania nie powoduje uszkodzenia całej sieci. Uszkodzenie koncentratora unieruchamia całą sieć.



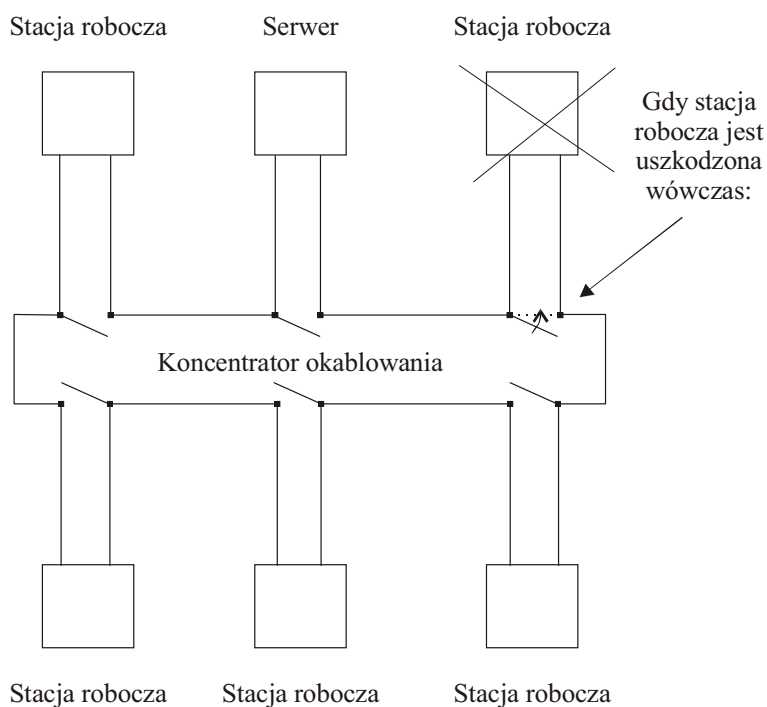
Rysunek 7.3. Topologia gwiazdy

Topologia pierścienia (rys. 7.4) polega na wzajemnym połączeniu stacji roboczych za pomocą jednego medium w układzie zamkniętym, który przypomina okrąg. Długości połączeń między komputerami są ograniczone i zależą od typu zastosowanego okablowania. Wysyłane sygnały, zanim dotrą do adresata, poruszają się w jednym kierunku przez poszczególne węzły. Awaria stacji lub łącza może spowodować unieruchomienie całej sieci. Trudne jest diagnozowanie uszkodzeń, a modyfikacja połączeń wymaga wyłączenia całej sieci.



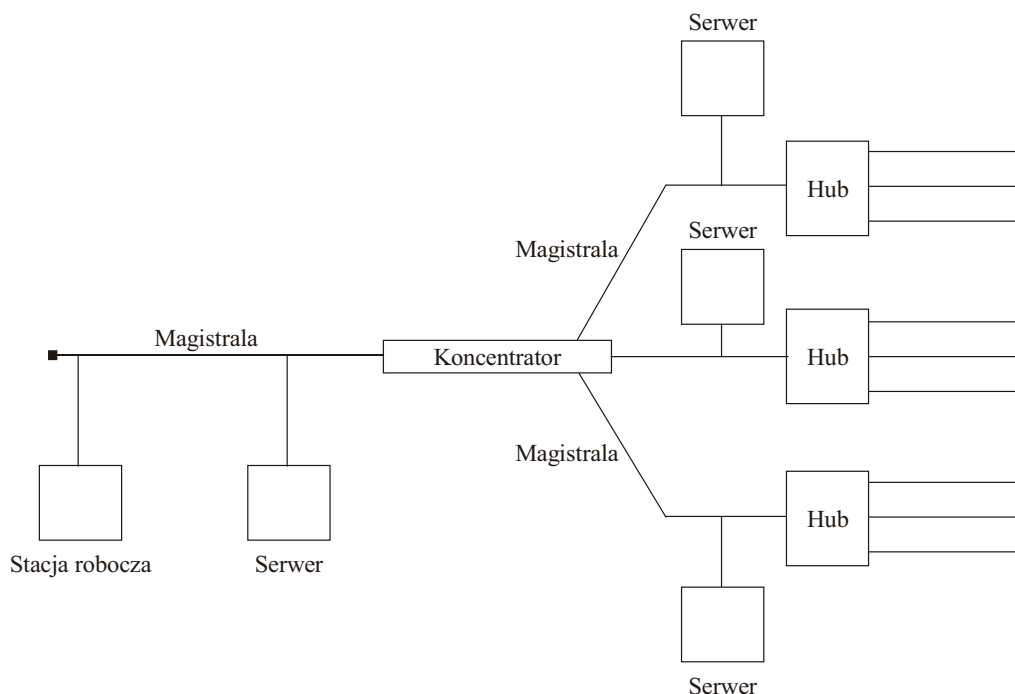
Rysunek 7.4. Topologia pierścienia

Zastosowanie awaryjnych obejść (rys. 7.5) może być przydatne w przypadku awarii stacji roboczej. W takim przypadku możliwa jest szybka rekonfiguracja sieci. Polega ona na zwarceniu połączeń prowadzących do uszkodzonego komputera.



Rysunek 7.5. Topologia pierścienia z awaryjnymi zabezpieczeniami

Topologia drzewa (rys. 7.6) powstaje w wyniku połączenia wielu magistral. Pierwsza magistrala podłączona jest do koncentratora, który dzieli ją na dwie, trzy lub więcej magistral. Proces dzielenia może zachodzić również w następnych magistralach. Liczba poziomów, które są utworzone przez podział, jest ograniczona. Zaletą topologii jest łatwy nadzór i rozbudowa sieci, również lokalizacja uszkodzeń nie sprawia trudności.



Rysunek 7.6. Topologia drzewa

Przedstawiony na rysunku 7.6 koncentrator jest urządzeniem opisanym w pytaniu 7.22.

Topologia pierścień-gwiazda powstała z połączenia topologii pierścienia i topologii gwiazdy. Zaletą konfiguracji jest możliwość dołączania nowej stacji bez przerywania pracy sieci, również odłączenie węzła nie powoduje awarii sieci. Centralnym elementem struktury jest pierścień, do którego podłączone są struktury w topologii gwiazdy.

Topologia gwiazda-magistrala reprezentuje sytuację, w której stacje robocze są bezpośrednio połączone w gwiazdę, a utworzone gwiazdy połączone są odcinkami kabli głównych.

7.9. Co to jest protokół sieciowy (komunikacyjny)?

Protokołem w sieci komputerowej nazywamy zbiór ściśle określonych zasad, według których komunikują się urządzenia sieciowe.

Protokoły odgrywają podstawową rolę w procesie realizacji połączenia w sieciach komputerowych.

7.10. Co to jest Ethernet ?

Standard Ethernet definiuje rodzaj okablowania sieci oraz specyfikuje sygnały przesyłane przez sieć. Odnosi się on do pierwszych dwóch warstw modelu OSI. Został stworzony przez firmę Xerox w latach siedemdziesiątych. Na początku lat osiemdziesiątych firmy Xerox, DEC i Intel rozpoczęły promocję Ethernetu. Obecnie jest to najpopularniejszy standard używany przy budowie lokalnych sieci komputerowych.

7.11. Jakie wyróżniamy rodzaje sieci Ethernet?

Pierwsze sieci Ethernet pracowały z prędkością 10 Mb/s. Stosowano przewód koncentryczny, skrętka lub światłowód.

Gwałtowny rozwój sieci komputerowych wymusił na konstruktorach stworzenie nowego standardu. Nazwano go Fast Ethernet, a prędkość transmisji w tym standardzie wynosi 100 Mb/s. Zwiększenie szybkości pracy sieci wprowadza większe wymagania w stosunku do okablowania. W sieci Fast Ethernet nie można stosować przewodów koncentrycznych, okablowanie powinno mieć piątą kategorię. Początkowo Fast Ethernet używany był w szkieletach sieci LAN.

Obecnie sieci 10 Mb/s są wypierane przez technologię Fast Ethernet, natomiast jako szkielet LAN stosuje się Gigabit Ethernet o przepustowości 1000 Mb/s. Jest to najnowszy standard sieci Ethernet wymaga stosowania skrętki kategorii piątej, w której wykorzystywane są cztery pary przewodów. Odległości większe niż 100 m wymagają zastosowania przewodów światłowodowych.

7.12. Jakie rodzaje okablowania stosowane są przy budowie sieci Ethernet?

Okablowanie stosowane przy budowie sieci Ethernet jest uzależnione od topologii oraz od prędkości przesyłania w tej sieci.

Dla sieci do 10 Mb/s istnieją następujące standardy okablowania:

- 10BASE5 – przewód koncentryczny, maksymalna długość segmentu 500 m, tzw. gruby Ethernet
- 10BROAD36 – przewód koncentryczny, maksymalna długość segmentu 3600 m
- 10BASE2 przewód koncentryczny, maksymalna długość segmentu 185 m, tzw. cienki Ethernet
- 10BASE-T – skrętka o maksymalnej długości segmentu 100 m
- 10BASE-F – kabel światłowodowy.

Dla sieci do 100 Mb/s wyróżniamy następujące rodzaje okablowania:

- 100BASE-TX – skrętka o maksymalnej długości segmentu 100 m (wykorzystane dwie pary)
- 100BASE-FX – kabel światłowodowy
- 100BASE-T4 – skrętka kategorii 5 o maksymalnej długości segmentu 100 m (wykorzystane cztery pary).

Dla sieci do 1 Gb/s istnieją następujące rodzaje okablowania:

- 1000Base-SX – kabel światłowodowy wielomodowy z transmisją laserową o długości fali 850 nm
- 1000Base-LX – kabel światłowodowy jedno- lub wielomodowy z transmisją laserową o długości fali 1300 nm
- 1000Base-CX – ekranowana skrętka
- 1000Base-T – skrętka kategorii 5 o długości segmentu od 25 m do 100 m (wykorzystane cztery pary).

7.13. Jakie urządzenia stosuje się do budowy i łączenia sieci LAN?

Przy budowie małych sieci komputerowych najczęściej wykorzystuje się wzmacniacze i koncentratory. Urządzenia te pracują w warstwie fizycznej, wzmacniają i regenerują sygnały bez ingerencji w przesyłane informacje.

Jeżeli w sieci znajduje się dużo komputerów, należy zastosować urządzenia separujące ruch. Przełączniki i mosty należą do urządzeń pracujących w warstwie łącz danych lub w warstwie sieciowej. W odróżnieniu od wzmacniaczy, przełączniki analizują adresy fizyczne komputerów. Na podstawie stworzonych tabel, ramki docierające do urządzenia kierowane są tylko do tego segmentu, w którym znajduje się komputer docelowy. Przełączniki w znaczący sposób obniżają ruch generowany w sieci.

7.14. Jakie protokoły sieciowe są najczęściej używane w sieciach LAN?

Dzięki szybkiemu rozwojowi sieci Internet najczęściej wykorzystywanym protokołem w sieciach LAN jest protokół IP (ang. *Internet Protocol*) (część zestawu protokołów TCP/IP). Równie popularnym protokołem jest IPX (część zestawu protokołów SPX/IPX firmy Novell) oraz NetBEUI (firmy Microsoft). Protokół AppleTalk DDP (ang. *Datagram Delivery Protocol*) jest wykorzystywany tylko w sieciach komputerowych firmy Apple.

7.15. Jakie media używane są do połączeń sieciowych?

W transmisji przewodowej wykorzystuje się:

- przewody elektryczne:
 - skrętkę
 - kabel koncentryczny
- przewody światłowodowe:
 - światłowody jednomodowe
 - światłowody wielomodowe.

Transmisja bezprzewodowa może odbywać się:

- drogą radiową
- w paśmie podczerwieni.

7.16. Jakie są kategorie okablowania?

Mnogość rozwiązań firmowych, które pojawiały się na rynku, wymusiła proces normalizacji. Normalizacja polegała na stworzeniu oficjalnych dokumentów zawierających pewne ogólne ustalenia, pozwalające na współpracę różnych producentów kabli, sprzętu aktywnego oraz innych elementów okablowania. Dzięki temu można łączyć ze sobą elementy różnych producentów i mieć pewność ich prawidłowego współdziałania.

Wyróżniamy następujące, zatwierdzone kategorie okablowania:

- kategoria 1 – tradycyjna, nieekranowana skrętka telefoniczna, odpowiednia do przesyłania głosu, nie przystosowana do transmisji danych
- kategoria 2 – nieekranowana skrętka, służąca do przesyłania danych z prędkościami do 4 Mb/s, kable tej kategorii zbudowane są z dwóch par skręconych przewodów
- kategoria 3 – kable tego typu pozwalają na transmisję z szybkością do 10 Mb/s, kable tej kategorii zbudowane są z czterech par skręconych przewodów
- kategoria 4 – kable z maksymalną szybkością transmisji określoną na 16 Mb/s, kabel jest zbudowany z czterech par przewodów
- kategoria 5 – miedziana skrętka o rezystancji 100 ohm pozwalająca na przesyłanie danych z szybkością 100 Mb/s.

Propozycje nowych kategorii, które jeszcze nie zostały zatwierdzone są następujące:

- kategoria 5E (ang. *enhanced*) – ulepszona

- kategoria 6 – do 200 (250) MHz na złączu RJ45
- kategoria 7 – do 600 MHz na nowym rodzaju złącza kompatybilnym „w dół” z RJ45.

7.17. Co rozumiemy pod pojęciami inteligentny budynek, okablowanie strukturalne?

Początkowo idea budynku inteligentnego zakładała tylko integrację instalacji przeciwpożarowych, oświetlenia i klimatyzacji.

Gwałtowny rozwój teleinformatyki, powszechne wprowadzanie elektronicznej wymiany informacji spowodowały, że do budynków zaczęto wprowadzać sieci komputerowe oraz skomplikowane systemy sterujące i zabezpieczające. Zastosowanie komputerów pozwala sterować instalacjami znajdującymi się w budynku, co w znacznym stopniu poprawia jego bezpieczeństwo oraz zapewnia użytkownikom wysoki komfort pracy. Inteligentny budynek jest obiektem, w którym wszystkie systemy są ze sobą powiązane, potrafią automatycznie reagować na zmieniające się warunki i minimalizować zagrożenia. Stanowi to znaczne ułatwienie dla człowieka. Nie wymaga stałego nadzoru i zmniejsza koszty utrzymania.

Pojęcie budynku inteligentnego jest ściśle związane z okablowaniem strukturalnym, które jest jednym z jego elementów. Głównymi częściami takiej struktury są systemy zarządzania:

- energią
- ogrzewaniem
- klimatyzacją
- bezpieczeństwem
- kontrolą dostępu
- ochroną przeciwpożarową
- oświetleniem
- telekomunikacją.

Okablowanie strukturalne stanowi pewne rozwiązanie, które może służyć jako sieć komputerowa, telekomunikacyjna, sygnałowa i alarmowa. Jest ono realizowane za pomocą przewodów miedzianych i światłowodów. W jego skład wchodzi: krosownice, gniazda, kable łączące, panele montażowe i szafy rozdzielcze.

7.18. Jakie wyróżnia się prędkości przesyłania informacji w sieciach komputerowych?

Prędkość przesyłania informacji w sieciach komputerowych jest uzależniona od zastosowanego standardu. Najczęściej spotykanymi prędkościami transmisji w sieciach komputerowych są: 1, 2, 10, 34, 100, 155, 622 Mb/s oraz 1 i 2,4 Gb/s.

Prędkość wyraża się w bitach na sekundę, natomiast wielkość plików podawana jest w bajtach. Wiedząc, że jeden bajt ma osiem bitów można w przybliżeniu porównać czasy, jakie potrzebne będą na przesłanie wzorcowego pliku o wielkości 500 MB przez sieci o różnych prędkościach.

Tabela 7.1. Porównanie czasów przesyłania pliku o wielkości 500 MB przez sieci o różnych przepływnościach.

Czas [s]	Prędkość [Mb/s]
4000	1
400	10
4	1000
1,7	2400

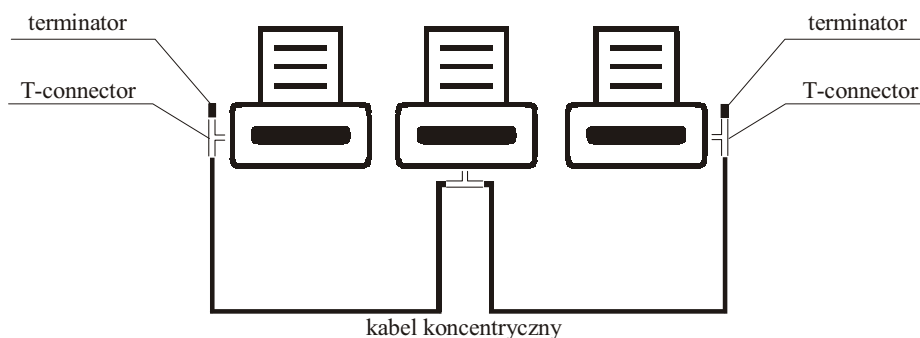
7.19. Czy informacje przesyłane w sieci są zabezpieczane przed błędami?

Sieci komputerowe mają mechanizmy zabezpieczające przesyłane w nich informacje przed powstawaniem błędów. Protokoły sieciowe odpowiedzialne za transmisję informacji potrafią wykrywać i naprawiać pojawiające się błędy. Skuteczność zabezpieczeń oraz umiejętność ich eliminacji uzależniona jest od właściwości protokołu obsługującego sieć.

7.20. Czy można w prosty sposób połączyć siecią Ethernet dwa komputery bez dodatkowych urządzeń?

Najprostszym sposobem połączenia komputerów w sieć lokalną jest zastosowanie kabla koncentrycznego, tzw. cienkiego Ethernetu. Sieć zbudowana z wykorzystaniem kabla koncentrycznego ma topologię magistralową (różnice między cienkim i grubym Ethernetem, patrz pyt. 7.12).

Łączone komputery powinny zawierać karty sieciowe, które są wyposażone w złącza BNC (ang. *Bayonet Neil-Concelman*) i odpowiednie rozgałęźniki – T-connectory. Topologia magistralowa wymaga podłączenia dwóch kabli do jednej karty sieciowej, na końcach kabla muszą być zamontowane terminatory.



Rysunek 7.7. Połączenie komputerów w standardzie Ethernet przy użyciu kabla koncentrycznego

7.21. Jaka jest różnica między routerem a bramą?

Brama (ang. *gateway*) jest urządzeniem, które pozwala na przekazywanie i translację informacji między sieciami, obsługuje i umożliwia współpracę różnych protokołów, a także usuwa różnice w formacie danych, prędkości ich przesyłania oraz poziomie sygnałów. Brama może działać na dowolnej warstwie modelu OSI.

Routery są urządzeniami łączącymi fizycznie oddzielone sieci w logiczną całość. Pracują na trzeciej (sieciowej) warstwie modelu OSI. Routery mają możliwość uczenia się, zapamiętywania i wyboru drogi o konkretnych właściwościach, np.: najkrótszej, najtańszej lub najszybszej.

7.22. Czym różni się wzmacniacz od przełącznika?

Wzmacniacze (ang. *repeater*) i koncentratory (ang. *hub*) wzmacniają i regenerują sygnały, należą do tej samej grupy urządzeń sieciowych, a różnią się liczbą gniazd i topologią sieci, w jakiej znajdują zastosowanie. Mają od kilku do kilkudziesięciu gniazd. Sygnały docierające do koncentratora są wzmacniane i przekazywane do wszystkich portów urządzenia.

Przełączniki (ang. *switch*) i mosty (ang. *bridge*) w porównaniu ze wzmacniaczami są o wiele bardziej skomplikowane. Potrafią analizować i modyfikować ruch w sieci tak, by przesyłane informacje kierowane były tylko do portu, do którego podłączony

jest komputer docelowy. Zastosowanie przełączników umożliwia rozbudowę sieci bez znaczącego pogorszenia jej parametrów.

7.23. Co to jest serwer?

Pojęcie serwer może odnosić się do komputera lub konkretnego programu komputerowego, który udostępnia pewne usługi. Z usług oferowanych na serwerze korzystają inne komputery lub programy komputerowe. Strona inicjująca połączenie to klient, natomiast jednostka odpowiadająca na zapytanie – to serwer (architektura klient – serwer).

Funkcja serwera lub klienta nie jest związana na stałe z komputerem lub programem, ale jest pojęciem zmieniającym się w zależności od czasu i wykonywanego zadania.

Może wystąpić taka sytuacja, w której serwer-komputer jest jednocześnie klientem. Dotyczy to systemów wielozadaniowych, w których równocześnie uruchomionych jest kilka programów.

Komputery obsługujące usługi internetowe, takie jak: poczta elektroniczna, strony www, transfer plików itp., nazywa się serwerami. O tych samych komputerach można powiedzieć, że są na nich uruchomione serwery usług internetowych.

7.24. Co to jest EDI?

EDI (ang. *Electronic Data Interchange*) jest to międzynarodowy standard elektronicznej wymiany dokumentów. EDI to próba przekształcenia społeczeństwa żyjącego w świecie „papierowym” na żyjące w świecie „dokumentów elektronicznych”.

Procesy zmiany formatu i transportu dokumentów wykonywane są przez komputery. Zastosowanie systemu elektronicznego obrotu dokumentami minimalizuje ryzyko powstania błędu. Przy wykorzystaniu EDI praca człowieka jest zminimalizowana. Dotychczasowe metody wymiany informacji w formie papierowej były czasochłonne i wymagały zaangażowania ludzi, którzy musieli ręcznie wypisywać wiele dokumentów, np.: faktury, przelewy, zezwolenia, dokumenty celne itp.

Korzystanie z elektronicznej wymiany dokumentów wymaga połączenia z siecią obsługującą standard EDI. Jeżeli nadawca ma własny, wewnętrzny format dokumentów elektronicznych, przed wysłaniem muszą być one przekonwertowane na format zgodny z EDI. Po stronie odbiorcy zachodzi proces odwrotny. Zmiana formatu umożliwia współpracę instytucji, które używają różnych systemów zapisu dokumentów elektronicznych.

8. Internet

8.1. Jak powstał Internet?

Korzenie sieci Internet sięgają lat sześćdziesiątych. Wtedy w Advanced Research Project Association (ARPA) zaczęły się prace nad siecią, która mogłaby połączyć komputery znajdujące się w kilku ośrodkach naukowych w USA. Już w 1969 powstała taka sieć. Nazywała się ARPANET. Połączyła ona cztery placówki naukowe – Uniwersytety Los Angeles, Santa Barbara, Uniwersytet Stanowy Utah oraz Instytut Stanforda.

W ciągu następnych kilku lat sieć rozszerzyła się o następnych 20 placówek, a w niedługim czasie sięgnęła także przez ocean do Anglii.

Na początku lat osiemdziesiątych pojawiły się pierwsze komercyjne zastosowania sieci, która w 1982 roku zyskała miano Internetu. W sieci Internet w 1984 roku znajdowało się około 1000 serwerów.

Koniec lat osiemdziesiątych to początek złotej ery Internetu. Był on wtedy używany jako jedno z podstawowych narzędzi umożliwiających komputerom komunikację ze sobą.

W 1989 roku liczba komputerów podłączonych do Internetu wzrosła do 100 tys., a w 1990 roku było ich już trzykrotnie więcej.

W chwili obecnej (styczeń 2000 r.) liczba zarejestrowanych komputerów to ponad 72 mln (dane z <http://www.isc.org>).

8.2. Co to jest Internet i jak działa?

Internet, czasami nazywany infostradą, jest zbiorem tysięcy sieci komputerowych tworzących globalne medium komunikacyjne.

Sieci te, połączone ze sobą, tworzą to, co potocznie nazywa się Internetem.

Komputery podłączone do sieci Internet, w przeważającym stopniu nie komunikują się ze sobą bezpośrednio. Informacje przepływają przez wiele urządzeń pośrednich, które nazywane są węzłami. Taka organizacja połączeń umożliwia wzajemną komunikację komputerów, które znajdują się w różnych sieciach, na różnych kontynentach. Od nadawcy wymagana jest znajomość internetowego adresu odbiorcy (adresu IP).

Adres IP jest unikatowym identyfikatorem komputera, identyfikuje również sieć, w której pracuje ten komputer. Jest on zapisywany za pomocą 32 bitów. Dla łatwiejszego zapamiętania adres IP zapisywany jest jako liczba składająca się z czterech bajtów oddzielonych kropkami, np. zapis adresu: 212.122.222.222 jest równoważny zapisowi w trybie szesnastkowym D47ADEDE lub dziesiętnym 3564822238. Każda z oddzielonych kropkami pozycji adresu przyjmuje wartość od 0 do 255.

Adres komputera w sieci składa się z dwu części. Pierwsza to adres sieci, druga to adres komputera w tej sieci. Hierarchiczna budowa adresów jest analizowana przez specjalne urządzenie (router) pracujący w węźle sieci. Router podejmuje decyzję, dokąd dalej skierować przychodzące do niego informacje. Pakiet pokonuje w ten sposób do kilkudziesięciu (domyślnie trzydziestu) przystanków zanim dotrze do celu. Droga, którą przekazywane są informacje, nie jest ściśle określona. W przypadku awarii jednej z dróg, następuje rekonfiguracja połączeń między węzłami i informacje dostarczane są do innego komputera węzłowego z nadzieją, że ten będzie w stanie ominąć uszkodzenie. Zwykły użytkownik nie jest w stanie przewidzieć, którędy pakiet dotrze do adresata. Możliwa jest sytuacja, że pakiety należące do tej samej wiadomości zostaną dostarczone różnymi trasami. Z uwagi na dowolność wyboru drogi nie można gwarantować, że pakiety dotrą do celu w takiej samej kolejności, w jakiej były wysyłane. Odbiorca musi posortować pakiety i ułożyć je we właściwej kolejności.

Zarówno nadawca, jak i odbiorca, nie mają żadnego wpływu na sposób działania sieci.

8.3. Kto administruje Internetem?

Ogólną opiekę nad Internetem sprawuje organizacja ICAN (ang. *The Internet Corporation for Assigned Names and Numbers*). Jest to niedochodowa organizacja, która została utworzona w celu przejęcia odpowiedzialności za:

- przyznawanie adresów IP
- przyznawanie parametrów protokołów
- zarządzanie systemem nazewnictwa domen
- administrację głównych serwerów sieci Internet.

Instytucje wspomagające ICAN, zajmujące się przydzielaniem adresów IP, to: RIPE NCC (Europa), ARIN (Ameryka Pn. i Pd., część Afryki), APNIC (Azja, Australia i Pacyfik).

Istnieją dwa typy domen najwyższego poziomu: domeny ogólne oraz domeny z kodami państw. Ogólne domeny zostały utworzone na użytek społeczności Internetu, podczas gdy domeny z kodami krajów utworzono do użytku odpowiadających im państw, jeżeli będą tego potrzebowały. Ogólne domeny (ang. *Generic Domains*) są

trzyliterowe np.: *.com*, *.org*, *.net*, *.edu*, *.gov*, *.mil*, *.int*. Domeny państwowe są dwuliterowe np.: *.pl*, *.us*, *.sk*. Rejestracją domen *.com*, *.net*, *.org* zajmują się komercyjne firmy zrzeszone przez InterNIC (<http://www.internic.net/alpha.html>). Domeny *.gov*, *.mil*, *.edu* zarezerwowano dla USA, domena *.int* jest przeznaczona dla instytucji międzynarodowych nie posiadających przynależności państwowej, np. Międzynarodowa Unia Telekomunikacyjna (<http://www.itu.int>).

Rejestracją domen państwowych zajmują się uprawnione instytucje (<http://www.iana.org/cctld/cctld-whois.htm>).

Sprawa administracji całej sieci Internet rozkłada się na poszczególne mniejsze sieci, które mają swoich właścicieli. Każdy administrator mniejszej sieci dba tylko o wycinek światowego Internetu, który należy do jego kompetencji.

8.4. Co to jest etykieta sieciowa?

Etykieta sieciowa, często nazywana netykietą (ang. *net* – sieć), jest zbiorem zasad, jakimi powinien kierować się użytkownik Internetu. Przypomina trochę zasady dobrych manier, z tą tylko różnicą, że dotyczy korzystania z Internetu.

Zawartość netykiety zależy od usługi, której ona dotyczy. Grupy dyskusyjne rządzą się innymi zasadami niż pogawędki na IRC-u (ang. *Internet Relay Chat*), również poczta elektroniczna ma swoją etykietę.

Pomimo tego, że netykietę można spotkać w bardzo różnych odmianach, zawsze zasady w niej zawarte mają na celu ułatwienie i uprzyjemnienie pracy zarówno sobie, jak i innym internautom.

8.5. W jaki sposób można uzyskać dostęp do Internetu?

Wśród wielu różnych sposobów uzyskania dostępu do sieci Internet można wyróżnić dwie grupy.

Pierwsza z nich to łącza stałe (dedykowane, dzierżawione). Charakteryzują się tym, że nie wymagają ciągłego łączenia i rozłączania w chwili, gdy istnieje potrzeba skorzystania z zasobów sieci Internet. Zestawienia takiego łącza dokonuje się raz, w momencie instalacji i konfiguracji. Do tej grupy należy sieć lokalna połączona z Internetem, linia dzierżawiona lub bezprzewodowe łącze dzierżawione.

Druga grupa to połączenia komutowane, zestawiane bezpośrednio przed rozpoczęciem pracy w Internecie, a po jej zakończeniu rozłączane. Należą do nich połączenia modemowe analogowe i cyfrowe, wykorzystujące linie telefoniczne.

8.6. Jakie podstawowe protokoły transmisji dostępne są w Internecie i jak działają?

Każdy komputer podłączony do Internetu ma swój własny adres IP, niepowtarzalny, jednoznacznie identyfikujący go w sieci.

Aby dwa komputery znajdujące się w jednej sieci mogły się ze sobą komunikować, muszą znać nawzajem swoje adresy fizyczne. W tym celu należy przekształcić adres IP na adres fizyczny tak, aby informacja mogła być poprawnie przesłana. W sieci Ethernet stosuje się 48-bitowe adresy fizyczne, które przypisywane są w trakcie procesu produkcyjnego urządzeń sieciowych. 32-bitowy adres IP nie zawiera fizycznego 48-bitowego adresu ethernetowego.

Przekształcenia adresu IP na adres fizyczny dokonuje protokół odwzorowania adresów ARP (ang. *Address Resolution Protocol*). Protokół ARP zapewnia dynamiczne odwzorowanie, nie wymaga przechowywania tablic adresowych.

Podstawowym protokołem, zapewniającym przenoszenie pakietów bez użycia połączenia (pakiety przenoszone są za pośrednictwem kilku do kilkudziesięciu węzłów, znajdujących się pomiędzy nadawcą i odbiorcą, bez wcześniejszego ustalania drogi), jest Internet Protocol w skrócie IP. Protokół IP jest zdefiniowany jako zawodny system przenoszenia pakietów bez użycia połączenia. Nie zapewnia on poprawności przenoszonych informacji. Wysyłane pakiety mogą podróżować różnymi ścieżkami, mogą zostać zgubione, zduplikowane, zatrzymane lub dostarczone z błędem. System nie sprawdza błędów, nie powiadamia nadawcy ani odbiorcy. Jeżeli wystąpią błędy, nadawca musi zostać poinformowany, aby podjąć działania w celu uniknięcia skutków takiej sytuacji. W tym celu powstał protokół komunikatów kontrolnych internetu ICMP (ang. *Internet Control Message Protocol*). Protokół ICMP jest traktowany jako wymagana część IP, każda implementacja protokołu IP musi wspierać ICMP. Gdy wystąpi błąd, protokół ICMP powiadamia o tym nadawcę, który musi zadbać o korekcję.

UDP (ang. *User Datagram Protocol*) jest protokołem nadrzędnym. Przesyłane komunikaty, oprócz danych, zawierają informację o porcie nadawcy i odbiorcy. Oprogramowanie UDP odbiorcy potrafi rozróżnić dla jakiego procesu przeznaczone są przesyłane informacje. UDP podobnie jak protokół IP jest zawodny, program użytkowy korzystający z UDP musi sam zadbać o poprawność przesyłania danych. Programy oparte na UDP dobrze sobie radzą w sieciach o dużej niezawodności i małych opóźnieniach, ale gorzej w większych intersieciach TCP/IP.

Za niezawodne przesyłanie danych w sieciach odpowiedzialny jest protokół TCP (ang. *Transmission Control Protocol*), który również zawiera informację o porcie nadawcy i odbiorcy. TCP jest bardzo skomplikowanym protokołem i jest całkowicie odpowiedzialny za bezbłędne dostarczenie pakietów. Wraz z IP tworzy rodzinę protokołów TCP/IP, która realizuje usługę niezawodnego, bezpołączeniowego przekazywania danych w Internecie.

8.7. Jak są rozróżniane poszczególne komputery w Internecie?

Podstawą rozpoznawania komputera w Internecie jest protokół IP. Na jego podstawie każdy komputer podłączony do sieci otrzymuje unikatowy adres, który jest jego reprezentacją. Adres IP reprezentuje nie tylko sam komputer, ale również sieć, do której jest on podłączony.

Zapamiętywanie adresu IP w postaci czterech liczb oddzielonych kropkami nie jest trudne dla komputera, ale jeżeli człowiek ma zapamiętać taki adres, staje się to kłopotliwe. W celu ułatwienia zapamiętywania adresów IP stworzony został system DNS (ang. *Domain Name System*). Odzworowuje on cyfrowy sposób zapisu adresu na wieloczkłowy, łatwy do zapamiętania adres alfanumeryczny, np. adresowi 212.122.192.34 przypisana jest nazwa *sun1.atr.bydgoszcz.pl*. Przyjęta jest zasada, że sposób nadawania nazw związany jest z przynależnością organizacyjną i administracyjną danego adresu. W podanym przykładzie komputer nazywa się *sun1*, należy do Akademii Techniczno-Rolniczej w Bydgoszczy w Polsce.

Jeden adres IP może mieć przypisanych kilka nazw, które określają dostępne serwisy, natomiast nazwa może mieć przypisany tylko jeden adres.

8.8. Według jakiego klucza są nadawane nazwy i adresy komputerom pracującym w ramach Internetu?

Adres składa się z nazwy komputera lub usługi oraz domeny rozdzielonych kropkami.

- *www.wp.pl*
- *www.man.bydgoszcz.pl*
- *www.mzios.gov.pl*
- *pit.atr.bydgoszcz.pl*
- *www.nas.gov*
- *www.ny.com*

Domeny mają budowę hierarchiczną. Mogą składać się z kilku członów. Pierwszy z nich jest nazwą komputera, ostatni oznacza domenę ogólną lub domenę państwa. Domeny dwuliterowe wskazują na przynależność państwową, np.: *.pl*, *.de*. Trzyliterowe domeny ogólne oznaczają rodzaj instytucji np.:

- komercyjna – *.com*
- związana z siecią – *.net*
- rządowa – *.gov* (zarezerwowana dla USA)
- naukowa – *.edu* (zarezerwowana dla USA)

- wojskowa – *.mil* (zarezerwowana dla USA)
- międzynarodowa – *.int*.

Hierarchiczne ułożenie nazw domen pozwala jednoznacznie zidentyfikować przynależność komputera, np. adres: *www.man.bydgoszcz.pl* oznacza, że komputer ten znajduje się w Polsce, w Bydgoszczy, należy do sieci MAN i dodatkowo wiadomo, że jest to serwer stron WWW.

8.9. Jakie jest odniesienie modelu OSI do protokołu TCP/IP?

Podobnie jak model OSI, również protokół TCP/IP ma strukturę warstwową. Istnieją jednak między nimi pewne różnice. Protokoły TCP i IP odpowiadają za ustalenie zasad komunikacji, szczegółów formatu komunikatów i sposobu odpowiadania na nie. Muszą też obsługiwać pojawiające się błędy i nienormalne sytuacje. Działają niezależnie od sprzętu sieciowego, zapewniają działanie podstawowych usług internetowych: ftp, telnet, poczta itp.

Najwyższy poziom to warstwa programów użytkowych. Następuje tam wywołanie programów, które korzystają z usług TCP/IP. Programy użytkowe wymieniają informacje z protokołami warstwy transportowej w postaci pojedynczych komunikatów lub strumienia bajtów. Warstwa transportowa zapewnia pewność komunikacji między programami użytkownika. W przypadku wystąpienia błędów może ponownie wysłać utracone przez nadawcę pakiety. Warstwa intersieci obsługuje komunikację między odległymi serwerami. Warstwa ta sprawdza poprawność przysyłanych pakietów i stwierdza, czy należy je przesłać dalej, czy przetwarzać na miejscu. Interfejs sieciowy odbiera i transportuje pakiety IP przez sieć.

8.10. Dlaczego powstał protokół IPv6?

Jednym z istotnych problemów w Internecie jest brak wolnych adresów IP. Adres IP jest unikatowym identyfikatorem komputera, identyfikuje również sieć, w której pracuje ten komputer.

Na częściowe rozwiązanie tego problemu pozwala translacja adresów – NAT (ang. *Network Address Translation*). Translacji adresów dokonują programy, które są zainstalowane na komputerach stanowiących połączenie sieci lokalnej i sieci Internet. Komputer – translator posiada minimum dwa adresy IP: zewnętrzny (istniejący w sieci Internet) i lokalny (reprezentujący komputer w sieci wewnętrznej). Cała sieć wewnętrzna w Internecie będzie widziana jako jeden komputer o adresie takim jak adres internetowy komputera – translatora.

Pełnym rozwiązaniem braku adresów IP jest wprowadzenie protokołu IPv6. W odróżnieniu od swojego poprzednika adresy w protokole IPv6 składają się ze 128 bitów,

a nie 32 bitów, jak to było wcześniej. Oprócz większej liczby adresów protokół IPv6 wprowadza kilka nowych elementów, jakich brakowało protokołowi IPv4.

8.11. Jak sprawdzić działanie połączenia z Internetem?

Najpopularniejszym programem, który służy do sprawdzania poprawności połączenia z Internetem, jest ping. Jest to program sprawdzający poprawność połączenia TCP/IP pomiędzy dwoma komputerami: lokalnym i odległym. Występuje na wszystkich platformach systemowych (Windows/Unix/Linux).

Aby sprawdzić istnienie połączenia z Internetem powinno się najpierw sprawdzać poprawność połączenia z najbliższym węzłem, który jest naszą bramą, następnie z dalszymi węzłami sieci.

Działanie programu ping polega na wysłaniu do zdalnego węzła, raz na sekundę, prośby o odpowiedź. Przychodzące odpowiedzi są potwierdzeniem poprawności połączenia. Jego jakość określają czasy oczekiwania podawane przez program w trakcie pomiarów.

8.12. Jakie podstawowe serwisy są dostępne w Internecie?

Obecnie najpopularniejszym serwisem, jaki oferuje Internet jest przesyłanie poczty elektronicznej. Elektroniczne listy mogą wędrować w ciągu kilku minut, a czasem nawet sekund, na drugi koniec kuli ziemskiej. Daje to ogromną przewagę nad tradycyjną pocztą, ponieważ w bardzo krótkim czasie można nie tylko wysłać wiadomość, ale i dostać odpowiedź.

Równie popularne w sieci jest przeglądanie witryn internetowych, popularnie nazywanych WWW (ang. *World Wide Web*). Istnieje bardzo dużo programów służących do przeglądania tych stron. Większość dużych firm i organizacji posiada swoje witryny w Internecie. WWW jest dobrym sposobem na znalezienie wiadomości, które w tradycyjny sposób nie są dostępne. Niezaprzeczalnym atutem korzystania z witryn WWW jest możliwość przeszukiwania zasobów. Istnieją w tym celu specjalne strony, tzw. wyszukiwarki. Serwery takie posiadają ogromne bazy danych, w których znajdują się zapisane informacje. Wyszukiwanie odbywa się przez wpisanie do formularza znajdującego się na stronie WWW frazy, którą chcemy odszukać. Następną zaletą, jaką dysponuje WWW, jest fakt, że czas powstania witryny jest o wiele krótszy niż wydanie publikacji lub książki. Trzeba jednak uważać na wiarygodność informacji zawartych w Internecie, bowiem szybkość powstawania stron WWW nie zawsze idzie w parze z dokładnością i wiarygodnością.

FTP (ang. *File Transfer Protocol*), czyli transfer plików w Internecie, jest następnym serwisem, o którym należy wspomnieć. Ułatwia on przenoszenie plików między

odległymi komputerami. Serwis ten działa na zasadzie klient–serwer. Oznacza to, że serwer udostępnia swoje zasoby, a klient może się z nim połączyć i z nich skorzystać. Często serwis ftp istnieje w powiązaniu z serwisem WWW, który umożliwia przeszukiwanie zasobów ftp i ma ładny interfejs graficzny.

Serwisy, telnet i ssh, które umożliwiają pracę na odległym komputerze, są bardzo rozpowszechnione. Pozwalają one korzystać z zasobów zdalnego komputera.

Grupy dyskusyjne, potocznie nazywane newsami, również należą do popularnych serwisów, podobnie jak różnego rodzaju pogawędki internetowe lub wideokonferencje.

8.13. Co to jest DNS i jak działa?

DNS, czyli serwer nazw internetowych (ang. *Domain Name Server*), jest rozproszoną bazą danych, która umożliwia translację nazwy serwerów z formatu symbolicznego na numeryczny i odwrotnie (ang. *RevDNS*). DNS działa korzystając z sieci serwerów nazw, przechowujących i przesyłających między sobą informacje adresowe. Każdy komputer podłączony do Internetu, korzysta z usług przynajmniej jednego serwera nazw lub jest sam takim serwerem.

System DNS ma formę hierarchiczną, podobnie jak nazwy domen. Każda domena ma swój, przynajmniej jeden, serwer nazw. Nosi on wtedy nazwę primary i jest autorytatywnym źródłem informacji. Jeżeli w obrębie domeny działają dwa serwery DNS, to jeden z nich musi być serwerem zapasowym, przechowującym kopię zawartości serwera podstawowego.

Komputer podłączony do sieci Internet komunikuje się z systemem DNS, wysyłając zapytanie do serwera, który jest podany jako jej serwer nazw. Serwer nazw działa w sposób rekurencyjny. Oznacza to, że po otrzymaniu zapytania realizuje odpowiednią sekwencję zapytań innych serwerów i zwraca odpowiedź. Taka odpowiedź jest przez pewien czas przechowywana w pamięci podręcznej serwera, a przy ponownym zapytaniu o ten sam adres serwer podaje go nie komunikując się z innymi. Buforowanie zapytań zmniejsza ruch w sieci. Jedną z kilku możliwych konfiguracji serwera DNS opiera się tylko na buforowaniu zapytań wychodzących z komputerów podłączonych do sieci, którą ten serwer obsługuje.

8.14. Co to jest serwis WWW i jak on działa?

Przeglądarka stron Internetowych korzysta z multimedialnego systemu informacyjnego WWW, czasami nazywanego światową pajęczyną informacyjną (ang. *World Wide Web*). Przeznaczony jest on do przedstawiania i udostępniania danych oraz informacji multimedialnych.

Udostępniane materiały znajdują się na serwerach. Mogą zawierać najróżniejsze formy danych: od tekstów, przez grafikę, animacje, dźwięki, sekwencje wideo, po duże bazy danych.

Serwery WWW znajdują się na całym świecie i połączone są przez Internet. Taka lokalizacja i wzajemne połączenie umożliwia korzystanie z zawartych na nich danych wszystkim podłączonym do sieci. Informacje między serwerem a klientem przesyłane są przez sieć za pomocą protokołu HTTP (ang. *Hypertext Transfer Protocol*).

Identyfikatorem serwera jest jego adres internetowy. Składa się on z dwóch części. Pierwsza to nazwa serwera, druga to domena. Przykładowo: w adresie `http://www.iweb.pl` nazwa serwera to `www`, a domena to `iweb.pl`. Ciąg znaków `http://` oznacza, że do protokołu HTTP będzie używany w transmisji między serwerem a klientem.

8.15. W jaki sposób można przysyłać pliki w Internecie?

Przesyłanie plików, popularnie nazywane FTP (ang. *File Transfer Protocol*), działa podobnie jak WWW, z tą tylko różnicą, że informacje są przesyłane w postaci plików. Klient nawiązuje połączenie z serwerem, następnie musi podać nazwę użytkownika i hasło. Wtedy dopiero następuje zalogowanie i transmisja. Możliwe jest przesyłanie plików w obie strony, do serwera i do klienta. Można przeglądać zawartość katalogów serwera i wybrać sobie to, co nas interesuje. Jeśli nie posiada się identyfikatora, należy użyć: `anonymous`, a jako hasło podać adres poczty elektronicznej. Mamy wtedy ograniczony dostęp do zasobów serwera, możemy je przeglądać i odczytywać, ale zapis jest zwykle możliwy tylko w katalogu `upload` lub `incoming`.

8.16. Czy można pracować na zdalnym komputerze podłączonym do Internetu? Jeśli tak, to jak to zrobić?

Praca na zdalnym komputerze jest możliwa, muszą jednak być spełnione pewne warunki. Oba komputery powinny być podłączone do sieci Internet lub lokalnej sieci Intranet za pośrednictwem protokołu TCP/IP. Na komputerze zdalnym powinna być dostępna usługa `telnet` (transmisja nieszyfrowana) lub `ssh` (transmisja szyfrowana), na komputerze lokalnym musi być program, który jest klientem usługi `telnet` lub `ssh`.

Usługa zdalnej pracy jest najłatwiejsza do zrozumienia, jeśli ktoś już zetknął się z systemem UNIX. Jest to system, w którym większość pracy przebiega na terminalach działających w trybie tekstowym. Użytkownik zalogowany do systemu dostaje do dyspozycji powłokę systemową, tak zwany `shell`. Z jego poziomu może obsługiwać system przez wpisywanie poleceń i podawanie parametrów. Przed rozpoczęciem pracy na zdalnym komputerze wywoływany jest program `login`, który autoryzuje użytkownika i chroni przed niepowołanym dostępem do serwera za pomocą hasła (ang. *password*).

Program wykonujący tę usługę jest odpowiednikiem lokalnego terminala. Z punktu widzenia systemu nie jest bowiem istotne, czy użytkownik pracuje lokalnie, czy też zdalnie. Jediną różnicą może być szybkość odpowiedzi systemu na polecenia. Pod wszystkimi innymi względami terminal emulowany (terminal naśladowujący) zachowuje się dokładnie tak samo, jak lokalny terminal serwera, czyli najpierw daje nam do obejrzenia login, a po wpisaniu identyfikatora i hasła udostępnia jego zasoby.

8.17. Jak działa poczta w Internecie?

Poczta elektroniczna (ang. *electronic mail*, *e-mail*) jest to forma przesyłania informacji między użytkownikami Internetu. Jej działanie jest bardzo zbliżone do pracy normalnej poczty. Przekazywane informacje mogą obejmować tekst, grafikę itp.

Nadawca musi znać adres poczty elektronicznej odbiorcy listu. Adres taki składa się z nazwy użytkownika i sieci, w której użytkownik ma konto. Są one oddzielone znakiem @ (znak @ w żargonie zwany „małpą”, „zawijaszem” lub „embrionem” oznacza ang. *at* – przy). Przykładowy adres może wyglądać tak: *pit@iweb.pl*.

Każdy elektroniczny list składa się z dwóch części: nagłówka i właściwej treści. W nagłówku zawarte są informacje sterujące, potrzebne do prawidłowego dostarczenia wiadomości. Są to między innymi: kto ją wysłał (From), do kogo (To:), kiedy (Date:), jaki ma temat (Subject:), do kogo zostały wysłane kopie listu (Cc:). Nagłówek może też zawierać nazwę programu, jaki służył do wysłania wiadomości, oraz nazwy serwerów, które pośredniczyły w przekazywaniu listu.

Procedura wysłania wiadomości rozpoczyna się od redakcji jego treści, zaadresowania i przesłania jej do serwera pocztowego, który obsługuje nasze konto poczty elektronicznej. Serwer pocztowy musi rozpoznać, do kogo przeznaczona jest przesyłka. W tym celu analizuje adres e-mail w poszukiwaniu nazwy sieci, w której znajduje się konto adresata. Następnie znajduje adres internetowy serwera pocztowego odbiorcy i rozpoczyna proces przesyłania poczty. Serwer odbierający przechowuje pocztę do czasu, aż adresat jej nie odczyta. Wiadomości można odczytać bezpośrednio na serwerze lub połączyć się z nim z innego komputera za pomocą programu pocztowego.

8.18. Do czego służy serwis news?

Grupy dyskusyjne, które potocznie nazywane są newsami, służą do wymiany poglądów między użytkownikami sieci.

Organizacja serwisu opiera się na hierarchicznym podziale tematycznym. Zgodnie z tym podziałem skonstruowane są nazwy grup. Są one trochę podobne do nazw domen komputerów w Internecie, np.: *pl.comp.os.linux.sieci*, *pl.soc.prawo*, *pl.rec.humor*. Nazwa grupy podzielona jest na kilka członów oddzielonych kropkami w ten sposób,

że pierwszy człón jest najbardziej znaczący. W podanym przykładzie *pl.* decyduje o języku, w jakim prowadzone są dyskusje, drugi człón *comp* oznacza, że grupa ta jest poświęcona komputerom. Następne człony *os.linux.sieci* zacieśniają tematykę do rozważań na temat użytkowania i konfiguracji sieci systemu operacyjnego Linux.

Za pomocą odpowiedniego programu obsługującego serwis grup dyskusyjnych, każdy użytkownik może odczytać wiadomości wysyłane przez innych Internautów oraz umieścić swoją wiadomość. Należy dodać, że serwery usług grup dyskusyjnych udostępniają je tylko określonym użytkownikom, np. w obrębie określonej domeny lub instytucji. Firmy zapewniające dostęp do Internetu posiadają serwery grup dyskusyjnych i udostępniają je swoim klientom. Korzystanie z lokalnego serwisu grup dyskusyjnych minimalizuje obciążenie sieci. Aby umożliwić wymianę poglądów między klientami różnych serwerów, wymieniają się one między sobą nowymi wiadomościami. Między serwerami przesyłane są tylko niektóre grupy (zadeklarowane przez administratora serwera), których liczba może wahać się od 3000 do ponad 10000.

8.19. Jak można prowadzić pogawędki w Internecie?

Wielka sieć komputerowa, jaką jest Internet, służy do przekazywania informacji w najróżniejszej formie. Uniwersalność sprzyja powstawaniu wielu ciekawych programów, wykorzystujących Internet jako medium transmisyjne. Jednym z nich jest IRC (ang. *Internet Relay Chat*), czyli rodzaj pogawędki na żywo w Internecie. Do tego celu należy mieć odpowiedni program. Program łączy się z serwerem, do którego podłączeni są inni użytkownicy aktualnie przebywający w sesji IRC-a.

Rozmowa polega na przesyłaniu krótkich (do kilku linijek naraz) wiadomości między użytkownikami – docierają one do rozmówcy zwykle po kilku sekundach. Takie krótkie linijki wysyła się do innego użytkownika sieci IRC, podając jego pseudonim (ang. *nick*).

Transmisja pomiędzy użytkownikami na IRC-u odbywa się za pośrednictwem serwerów. Klient wysyła informacje do serwera, a serwer rozsyła otrzymane informacje do innych użytkowników, będących na tym samym kanale (wirtualnym miejscu, gdzie może rozmawiać wielu użytkowników jednocześnie). Na IRC-u możliwe jest zdefiniowanie wielu kanałów, na których się rozmawia. Internetowe rozmówki można prowadzić używając różnych serwerów IRC. Serwery te są ze sobą zsynchronizowane. Oznacza to, że uaktualniają wzajemnie wiadomości, jakie otrzymują od klientów. Daje to możliwość rozmawiania osobom, które znajdują się daleko od siebie.

8.20. Czy można słuchać radia lub oglądać telewizję przez Internet?

Jednym z wielu zastosowań Internetu jest przesyłanie dźwięku lub obrazu z dźwiękiem. RealAudio i RealVideo to technologie realizujące to zastosowanie.

Transmisja odbywa się w czasie rzeczywistym, nie ma konieczności wcześniejszego kopiowania danych na dysk lokalny.

Umożliwia to słuchanie audycji radiowych lub oglądanie programów telewizyjnych na żywo. Wiele stacji radiowych i telewizyjnych ma własne serwery umożliwiające odbieranie programów przez Internet.

Technologie RealAudio i RealVideo nie ograniczają się tylko do transmisji audycji w czasie rzeczywistym, źródłem informacji mogą być również pliki zapisane na dysku twardym.

8.21. Czy możliwe jest połączenie wideokonferencyjne z innym użytkownikiem Internetu?

Połączenia wideokonferencyjne można realizować za pośrednictwem Internetu. Potrzebne jest do tego specjalne oprogramowanie oraz kamera i mikrofon podłączone do komputera.

Przy połączeniach wideokonferencyjnych w Internecie i sieciach lokalnych przyjęty jest jeden standard kompresji. Umożliwia to współpracę programów różnych producentów.

Choć transmisja pomiędzy użytkownikami odbywa się bezpośrednio, bez użycia serwera, to przy nawiązywaniu połączenia potrzebna jest informacja o adresie internetowym drugiego rozmówcy. Służy do tego serwer katalogowy. Przed rozpoczęciem połączenia użytkownik loguje się do serwera i pozostawia tam informacje o swoim aktualnym adresie internetowym.

Jakość dźwięku i obrazu uzależniona jest od prędkości połączenia internetowego. Im większa prędkość – tym lepsza jakość. Do użytku domowego wystarczy modem analogowy o prędkości 56 kb/s. Studyjna jakość połączenia wymaga prędkości transmisji rzędu 30 Mb/s.

8.22. Jak wysłać faks przez Internet?

Wysyłanie faksów przez Internet odbywa się za pośrednictwem serwera, który odbiera zgłoszenia przychodzące z Internetu i dalej przekazuje je w formie faksu do

adresata. Serwer pośredniczący nazywany jest bramką. Najczęściej spotykane rodzaje bramek to bramki wykorzystujące pocztę elektroniczną lub strony WWW.

Dostarczenie faksu polega na wypełnieniu odpowiedniego formularza na stronie www lub wysłaniu wiadomości e-mail. W treści takiej wiadomości umieszcza się zawartość faksu, jako temat należy wpisać numer telefonu odbiorcy.

8.23. Czy można korzystać z Internetu przez telefon komórkowy?

Korzystanie z Internetu przez telefon komórkowy jest możliwe. Oprócz telefonu użytkownik musi posiadać specjalny modem GSM i kabel łączący telefon z komputerem. Niektóre telefony mają wbudowany wewnątrz modem, np.: NOKIA. W takim przypadku nie jest wymagany dodatkowy modem, wystarczy kabel łączący telefon z komputerem. Wraz z kablem dostarczane jest oprogramowanie obsługujące modem GSM.

Przed rozpoczęciem użytkowania należy zainstalować oprogramowanie i skonfigurować parametry połączenia. Jeżeli w planie taryfowym telefonu nie ma usługi transmisji danych, należy wykupić dodatkową usługę. Wszyscy operatorzy telefonii komórkowej w Polsce udostępniają transmisję danych.

Prędkości osiągane w transmisji przez telefon komórkowy zazwyczaj nie przekraczają 9,6 kb/s. Takie połączenia nie dają komfortu pracy. Sieć Plus GSM wprowadziła usługę pozwalającą na dostęp do sieci z prędkością 43200 b/s, co wydaje się być optymalnym rozwiązaniem dla telefonów komórkowych. Zaletą łączności bezprzewodowej jest jej mobilność i możliwość wykorzystania wszędzie tam, gdzie jest zasięg sieci komórkowych.

Korzystanie z Internetu jest możliwe również bez użycia dodatkowego osprzętu (modem, kabel, komputer). Możliwe jest to jednak tylko w niektórych telefonach obsługujących WAP (ang. *Wireless Application Protocol*). WAP specyfikuje zestaw protokołów komunikacyjnych standaryzujących sposób, w jaki bezprzewodowe urządzenia, takie jak telefony komórkowe i radiotelefony, mogą korzystać z Internetu (poczta elektroniczna, WWW, grupy dyskusyjne i pogawędki IRC).

WAP został opracowany przez cztery firmy: Ericsson, Motorola, Nokia i Unwired Planet.

8.24. Czy połączenia w Internecie są bezpieczne?

W większości połączenia w Internecie nie są szyfrowane. Oznacza to, że potencjalnie każdy, kto ma dostęp do przesyłanych informacji, może dowiedzieć się co one za-

wierają. Niezaszyfrowana forma przesyłania informacji przez sieć Internet nie jest bezpieczna.

Aby podnieść bezpieczeństwo przesyłanych informacji, stosuje się połączenia szyfrowane z kluczem 128-bitowym – SSL (ang. *Secure Sockets Layer*). Chronią one informacje przed niepożądanym dostępem osób trzecich.

Podstawowe usługi, takie jak: poczta elektroniczna, przeglądanie stron WWW, transfer plików ftp, czy telnet początkowo nie były szyfrowane. Dynamiczny rozwój Internetu i wzrost ważności przesyłanych informacji wymusiły stworzenie szyfrowanej implementacji podstawowych usług internetowych.

8.25. Czy można robić zakupy i operacje bankowe przez Internet?

Korzystając z Internetu możemy dokonywać zakupów bez konieczności wychodzenia z domu.

Sklepy internetowe działają poprzez witryny WWW. Strony sklepów internetowych umożliwiają wirtualnie przeglądanie oferowanych towarów. Spośród oferowanych produktów można wybrać i zamówić interesujący towar. Oprócz towaru można wybrać formę płatności oraz sposób i termin doręczenia. Wirtualny sklep podliczy dokonane zakupy i powiadomi o kwocie, jaką należy zapłacić.

Najczęściej stosowane formy płatności wymagają posiadania karty kredytowej, ale spotykane są również inne formy płatności.

Niewątpliwym minusem elektronicznego handlu jest brak możliwości targowania się ze sprzedawcą, choć biorąc pod uwagę dynamikę rozwoju Internetu, nie jest wykluczone, że w niedługiej przyszłości także ten problem zostanie rozwiązany.

8.26. Co to jest SSL?

SSL (ang. *Secure Sockets Layer*) jest to protokół bezpiecznej komunikacji pomiędzy serwerem a klientem. Podstawowym zadaniem protokołu jest zapewnienie prywatności i wiarygodności pomiędzy dwoma aplikacjami komunikującymi się ze sobą.

SSL jest niezależny od protokołu aplikacji. Dla protokołów wyższej warstwy jest przezroczysty.

Komercyjne wykorzystywanie protokołu SSL wymaga wygenerowania specjalnego certyfikatu. Podpisywaniem wiarygodności certyfikatów SSL zajmuje się grupa instytucji certyfikujących CA (ang. *Certifying Authorities*). Uzyskanie podpisu od CA nie jest proste i wymaga przedstawienia szeregu dowodów tożsamości. W ten sposób mamy pewność, że serwery identyfikujące się podpisem od CA są rzeczywiście tymi, za które się podają.

9. Oprogramowanie systemów komputerowych

9.1. Co to jest algorytm?

Algorytmem nazywamy zbiór reguł postępowania (przepis), mający na celu, w skończonej liczbie kroków, przetworzenie informacji wejściowych (danych) w informacje wyjściowe (wyniki). Inaczej mówiąc, algorytm jest informatycznym planem rozwiązania zadania. Podane określenie ma charakter opisowy i tym samym niezbyt ścisły.

W potocznym znaczeniu algorytmem może być np.: przepis w książce kucharskiej, recepta na wykonanie leku w aptece itp.

Raz opracowany algorytm dla danego problemu może służyć do rozwiązywania wszystkich problemów tej samej klasy, dla której został opracowany, różniących się jedynie doбором konkretnych danych wejściowych. Na przykład algorytm rozwiązywania równania kwadratowego jest przydatny dla dowolnych wartości parametrów a , b , c , które stanowią tutaj dane wejściowe.

Przy opracowaniu algorytmu zakłada się pewien poziom szczegółowości, na którym formułuje się dany algorytm. Odbiorcą (wykonawcą) algorytmu może być zarówno komputer, jak i człowiek, np. studiujący publikację naukową. Przed opracowaniem algorytmu należy stwierdzić, czy zadanie jest rozwiązywalne, a jeśli tak, to czy rozwiązanie to jest jednoznaczne. Algorytm powinien uwzględniać wszystkie możliwe teoretycznie warianty przebiegu obliczeń zaistniałe z powodu różnego doboru danych wejściowych.

9.2. Jaki jest źródłosłów terminu algorytm?

Termin algorytm wiąże się z matematykiem perskim o nazwisku Muhamed Ibn Musy Al Chopezmi (z Chopezmu), który około roku 820 n.e. opisał system pozycyjnego kodowania dziesiętnego liczb i sposoby liczenia w tym systemie. System ten stosowany był wyłącznie w Indiach. Książka Al Chopezmiego została na początku XII wieku przetłumaczona na łacinę i w Europie zainicjowała rozwój metod wykonywania obliczeń pisemnych. Zwolennicy metod proponowanych w tej książce nazywali się algorytmistami i konkurowali z tymi, którzy w obliczeniach posługiwali się kamyczkami (kalkulatorzy) oraz tymi, którzy używali liczydła (abacyści). Pojęcie algorytmu, czyli recepty na postępowanie rachunkowe na piśmie, zostało później rozszerzone na sformalizowane obliczenia matematyczne.

9.3. Jakie są typowe konstrukcje algorytmiczne?

Wyróżnia się następujące konstrukcje algorytmiczne:

- proste
- z rozwidleniem
- z pętlą
- złożone.

W prostych konstrukcjach algorytmicznych używa się jedynie poleceń (instrukcji) przetwarzania i pomocniczych, nie można użyć instrukcji warunkowych. W takiej konstrukcji kolejność realizacji poszczególnych operacji jest ściśle określona, żadna z nich nie może być pominięta ani powtórzona.

Algorytm z rozwidleniem zawiera w sobie wybór jednej z kilku możliwych dróg realizacji danego zadania. W takiej konstrukcji istnieje przynajmniej jedno polecenie warunkowe.

Często w wielu algorytmach zdarza się, że występują wielokrotne powtórzenia niektórych operacji lub ich ciągów, różniących się jedynie zestawem danych, na których działają. Pętla obejmuje tę część instrukcji, która ma być powtarzana. Organizację pętli zapewniają:

- polecenia, w których określa się informacje dotyczące liczby obiegów pętli, określane licznikiem
- polecenia modyfikacji licznika, w których określa się sposób zmiany licznika po każdym obiegu pętli
- polecenia warunkowe, w którym następuje sprawdzenie, czy licznik osiągnął wartość graniczną określającą zakończenie pracy w pętli, i przejście do następnych bloków.

9.4. Jaka postać będzie miał przykładowy algorytm dla zadania poszukiwania maksymalnego elementu z ciągu danych n liczb?

Przykładowy algorytm poszukiwania maksymalnego elementu z ciągu danych n ($n \geq 2$) liczb mógłby wyglądać tak:

1. Ustaw zmienną *licznik* równą jeden.
2. Wczytaj wartość zmiennej *A*.
3. Wartość zmiennej *A* przypisz zmiennej *M*.

4. Zwiększ licznik o 1.
5. Wczytaj wartość zmiennej A .
6. Jeżeli A jest większe od M , to wartość zmiennej A przypisz zmiennej M . Jeżeli nie, to przejdź do punktu 7.
7. Jeżeli licznik nie jest równy n , to wróć do punktu 4. Jeśli tak, to przejdź do punktu 8.
8. Wyświetl wartość zmiennej M .

Przykładami prostych konstrukcji algorytmicznych są z osobna polecenia nr 1, 2, 3, 4, 5, 8. Przykładami poleceń z rozwidleniem są z osobna polecenia nr 6, 7. Sekwencja zaś poleceń nr 4, 5, 6, 7 (gdy licznik $\leq n$) tworzy konstrukcję algorytmiczną zwaną pętlą.

9.5. Co rozumiemy pod pojęciem język programowania, program i programowanie?

Językiem programowania nazywamy język, za pomocą którego można zapisywać algorytmy przetwarzania informacji przeznaczone dla komputera. Programem nazywamy algorytm przetwarzania informacji zapisany w języku programowania. Aby zdefiniować dowolny język programowania, należy określić:

- alfabet, czyli zbiór dopuszczalnych symboli, z których będą tworzone słowa i zdania w tym języku
- składnię (syntaktykę), czyli reguły tworzenia poprawnych słów i zdań (zwanych tu instrukcjami) tego języka
- reguły znaczeniowe (semantykę) interpretujące znaczenie poszczególnych zdań i konstrukcji języka.

W znaczeniu strukturalnym program ma postać ciągu instrukcji. Nie jest to zbyt ściśle określenie, gdyż w programach występują także tzw. opisy (deklaracje), których zadaniem jest dostarczenie komputerowi informacji o tym, jakie zmienne (np. liczby, wektory, macierze), jakiego typu (np. rzeczywiste, całkowite, logiczne, znakowe) i w jakiej liczności będą w programie występować. Na podstawie tych informacji komputer dokonuje rezerwacji odpowiednich obszarów pamięci przeznaczonych na te obiekty.

Znane jest sławne sformułowanie Wirtha (a jednocześnie tytuł książki) „STRUKTURY DANYCH + ALGORTM = PROGRAM”.

Alfabet i składnię definiuje się zazwyczaj ściśle i formalnie, reguły znaczeniowe zaś określa się opisowo. Program zapisany w jakimś języku programowania (z wyjątkiem wewnętrznego) nazywamy programem źródłowym. Po to, aby mógł być wykonany przez komputer, trzeba go przetłumaczyć na język wewnętrzny. Tłumaczenia dokonuje translator.

Programowaniem (w szerokim rozumieniu) nazywa się jednoznaczne formułowanie zadań oraz sposobu ich rozwiązywania na komputerze, prowadzące do wykonania szeregu czynności, takich jak:

- a) sformułowanie problemu
- b) określenie metody rozwiązania
- c) dyskusja warunków istnienia rozwiązania
- d) zapis algorytmu w wybranym języku programowania
- e) uruchomienie testowe programu
- f) opracowywanie dokumentacji.

Często programowaniem (w wąskim rozumieniu) nazywa się etap d).

9.6. Jak klasyfikujemy języki programowania?

Jednym z podstawowych kryteriów klasyfikacji jest stopień złożoności możliwych konstrukcji, które można utworzyć w danym języku. Ten stopień złożoności i wynikająca z niego wygoda programowania określa się umownie, definiując kilka poziomów języków programowania. Zwykle wyróżnia się języki:

- wewnętrzne
- assemblerowe (symboliczne)
- wysokiego poziomu (proceduralne).

Z punktu widzenia programisty język wewnętrzny jest na najniższym poziomie. Alfabet tego języka składa się wyłącznie z cyfr dwójkowych: 0 i 1. Zdaniem w tym języku są rozkazy, a jego składnię i reguły znaczeniowe definiuje konstruktor komputera. Składnia określa precyzyjnie formaty poszczególnych rozkazów, a reguły znaczeniowe sposób działania rozkazów. Często, dla wygody programisty, w celu uproszczenia i skrócenia zapisu rozkazów stosuje się, oprócz właściwych symboli języka (cyfr 0 i 1), zapis ósemkowy lub szesnastkowy.

Język assemblerowy zajmujący następną, wyższą poziom różni się istotnie od języka wewnętrznego przede wszystkim jeśli chodzi o alfabet i składnię. Reguły znaczeniowe w odniesieniu do instrukcji, którym odpowiadają pojedyncze rozkazy komputera, są identyczne. Rozszerzenie dotyczy pewnych dodatkowych konstrukcji nazywanych dyrektywami (pseudorozkazami) oraz makroinstrukcjami.

Poziomy wyższe stanowią języki wysokiego poziomu (proceduralne), których składnia i reguły znaczeniowe są nieraz bardzo skomplikowane, a poszczególne instrukcje odpowiadają ciągom wielu rozkazów wewnętrznych komputera. Języki te dzielą się na: uniwersalne i specjalizowane. Choć w dowolnym języku programowania

można przedstawić właściwie każdy algorytm, to jednak pewne właściwości i konstrukcje języka mogą to zadanie ułatwiać lub utrudniać. W przeciwieństwie do języków wewnętrznych i assemblerowych, języki wysokiego poziomu są niezależne od typu danego komputera.

Przy okazji omawiania języków programowania należy zwrócić uwagę na pewne istotne kwestie terminologiczne, nie zawsze akcentowane w opracowaniach. Termin „rozkaż” (ang. *instruction*) należy stosować tylko wtedy, gdy mowa jest o języku wewnętrznym. Natomiast termin „instrukcja” (ang. *statement*) może być użyty tylko w odniesieniu do języków assemblerowych i języków wysokiego poziomu. Zdarza się często, że mówi się „assembler”, a z kontekstu wynika, że mowa jest o języku assemblerowym. Jest to błąd terminologiczny, assembler bowiem jest programem tłumaczącym (czyli tłumaczem) z języka assemblerowego na język wewnętrzny komputera. Języki assemblerowe nazywa się też językami symbolicznymi.

9.7. Dlaczego programowanie w języku wewnętrznym jest uciążliwe?

Programując w języku wewnętrznym trzeba wyrazić złożony problem przy użyciu ciągu wielu rozkazów wykonujących bardzo elementarne czynności w komputerze. Programy są bardzo długie, a przez to duże jest również prawdopodobieństwo pomyłki. Program w języku wewnętrznym jest nieczytelny i niezrozumiały dla osób „niewtajemniczonych”. Programy może pisać właściwie tylko specjalista, w równej mierze znający zarówno sprzęt, jak i metody programowania. Napisane programy są użyteczne tylko dla jednego typu komputera, również dokonywanie poprawek jest uciążliwe.

9.8. Czym się różni język assemblerowy od języka wewnętrznego?

Zasadnicza różnica polega na tym, iż w językach assemblerowych zarówno kody rozkazów, jak i ich argumenty zastępuje się nazwami symbolicznymi, natomiast w językach wewnętrznych stosuje się zapis dwójkowy. Zajęcie pamięci i długość programu są takie same. Jedna instrukcja symboliczna (z nielicznymi wyjątkami) jest zastępowana jednym rozkazem komputera.

Gdyby ktoś miał wątpliwości, czy warto w takim razie pisać program w języku assemblerowym i „dodatkowo” tłumaczyć go na język wewnętrzny, powinien uświadomić sobie, o ile łatwiej zapamiętać mnemotechniczne skróty rozkazów niż ich kody wewnętrzne.

Języki assemblerowe pozwalają ponadto nadawać symboliczne nazwy komórkom pamięci oraz różnym stałym i parametrom używanym w programie. Programista ma pełną swobodę w doborze tych nazw i zwykle stara się, by odzwierciedlały rolę, jaką

pełnią w programie identyfikowane przez nie obiekty. Czynności zlecane komputerowi przez instrukcje zostają w ten sposób umieszczone w czytelnym kontekście, co znacznie ułatwia zrozumienie programu.

Oprócz instrukcji stanowiących symboliczny równoważnik rozkazów w programach assemblerowych mogą wystąpić także tzw. dyrektywy lub pseudoinstrukcje. Niektóre języki assemblerowe pozwalają również na stosowanie tzw. makroinstrukcji.

Wszystkie te udogodnienia znacznie ułatwiają i przyspieszają proces pisania i uruchamiania programu. Poza tym należy wyraźnie podkreślić, że programista korzystający z języka assemblerowego ma pełny dostęp do tych wszystkich elementów komputera, do których istnieje dostęp na poziomie języka wewnętrznego.

9.9. Jaka jest postać instrukcji w języku assemblerowym?

Ponieważ język assemblerowy jest ściśle związany z danym komputerem, więc postać instrukcji, struktura programu i inne elementy nie są jednoznacznie określone. Zwykle jednak instrukcja składa się z następujących czterech elementów:

- pola etykiety
- pola operacji
- pola argumentów
- pola komentarza.

Etykieta jest symbolicznym adresem danej instrukcji, do której mogą odwoływać się inne instrukcje programu (np. instrukcje skoku). Etykietą może być dowolny ciąg znaków. Często jednak ma on jakieś znaczenie, np. START, KONIEC, WSKAŹNIK itp. Zwykle długość etykiety jest ograniczona. Czasami też wymaga się, by pierwszym znakiem była litera. Koniec etykiety może sygnalizować wybrany znak, np. dwukroppek, choć czasem wystarcza dowolny separator, np. odstęp lub tabulator.

Pole operacji zawiera symboliczny kod rozkazu, dobrany w ten sposób, by łatwo kojarzył się z funkcją danego rozkazu. Nazwy używane w tym miejscu są nazwami zastrzeżonymi, zdefiniowanymi przez autorów programu tłumaczącego. Programista nie może ich zmieniać. Najczęściej są to skróty angielskich nazw operacji, np.:

- MOV (ang. *MOVE* – prześlij)
- MUL (ang. *MULTIPLY* – mnoż)
- CLR (ang. *CLEAR* – zeruj)
- INC (*INCREMENT* – zwiększ)
- BPL (ang. *Branch if PLUS* – skocz, jeśli dodatnie) itp.

Pole argumentów zawiera identyfikatory rozkazu. Mogą to być liczby, nazwy zmiennych i stałych, identyfikatory rejestrów, bądź też wyrażenia symboliczne, których wartość jest wyliczana podczas tłumaczenia programu, np.:

```
INC R0           ;zwiększ o 1 zawartość rejestru R0
ADD # 2,DATA    ;Dodaj liczbę 2 do zawartości
                ;komórki DATA
JMP START       ;skocz do komórki o adresie
                ;START
```

W powyższym przykładzie skorzystano z symboli i składni stosowanych w językach assemblerowych. Przy każdej instrukcji umieszczono teksty (poprzedzone średnikiem) wyjaśniające istotę działania instrukcji. Takie teksty nazywamy komentarzami. Stanowią one ostatni z wymienionych powyżej elementów składowych instrukcji w języku assemblerowym.

Komentarze są przeznaczone dla innych programistów, którzy mogą być zmuszeni analizować cudzy program, lub też dla samego autora programu, który po pewnym czasie może zapomnieć (często to się zdarza), dlaczego napisał takie, a nie inne instrukcje. Komentarze stanowią ważny element dokumentacji programu, lecz w fazie tłumaczenia programu są całkowicie pomijane i nie mają żadnego wpływu na treść programu wynikowego.

Poszczególne pola instrukcji mogą mieć swoje ściśle ustalone położenie w stosunku do początku linii. Sztywne rozmieszczenie tych pól ułatwia zadanie autorom programu tłumaczącego. Stanowi jednak istotne ograniczenie i utrudnienie dla programisty. Dlatego najczęściej się go nie stosuje. Poszczególne pola mogą się zaczynać w dowolnym miejscu, byle tylko występowały w określonej kolejności. Oddzielają je zwykle separatory, którymi mogą być odstępki, tabulatory i znaki wyróżniające. Znakiem kończącym etykietę bywa dwukropek, argumenty operacji zwykle oddziela przecinek, a początek komentarza sygnalizuje średnik.

9.10. Jakie są wady języków assemblerowych?

Języki assemblerowe stanowią krok naprzód w stosunku do języków wewnętrznych. Uwalniają one programistę od troski o właściwe adresowanie. Poza tym program w języku assemblerowym, dzięki skrótom mnemotechnicznym określającym typy rozkazów, jest bardziej czytelny.

Programowanie w tym języku ma jednak pewne wady. Jest czasochłonne, co stało się czynnikiem bardzo istotnym z chwilą wprowadzenia szybkich komputerów, dla których programowanie jest przysłowiowym „wąskim gardłem” w całym procesie

obliczeniowym. Programy mogą być uruchamiane tylko na tym typie komputera, dla którego zostały napisane. Programowanie wymaga znajomości listy rozkazów komputera, jego organizacji logicznej, pamiętania wielu wiadomości szczegółowych. Utrudnia ono porozumiewanie się ludzi nie będących profesjonalistami w informatyce z informatykami i uniemożliwia tym samym poszerzenie kręgu odbiorców techniki komputerowej.

9.11. Dlaczego języki assemblerowe są nadal stosowane mimo ich wad?

Języki assemblerowe są podstawą tworzenia tzw. translatorów języków wysokiego poziomu. Translatory języków assemblerowych (czyli assembly) zajmują zdecydowanie mniej miejsca. Stosowanie języka assemblerowego umożliwia programiście świadome wykorzystywanie pewnych specyficznych właściwości konstrukcji komputera.

9.12. Dlaczego języki wysokiego poziomu stanowią duże ułatwienie w pracy programisty?

Języki wysokiego poziomu ułatwiają pracę programisty dzięki temu, że:

- pozwalają myśleć w języku zbliżonym do tego, do którego jest on przyzwyczajony
- pozwalają na pisanie zdecydowanie krótszych programów niż w języku assemblerowym i wewnętrznym
- zmniejsza się możliwość powstawania tzw. literówek
- wprowadzona diagnostyka poprawności programów umożliwia kontrolę występowania błędów składniowych i logicznych
- zwalniają programistę z trudu rozmieszczania programu i danych w pamięci, wyznaczania adresów, posługiwania się urządzeniami wejścia/wyjścia
- ze względu na uniezależnienie języka od typu komputera, możliwa jest wymiana typowych programów
- zwiększa się czytelność programu i jego zrozumiałość zarówno w przypadku poszukiwania błędów logicznych, jak i przy studiowaniu tego programu po dłuższej przerwie
- nauka programowania jest łatwiejsza ze względu na przejrzystość struktur językowych i fakt, że nie trzeba zajmować się dużą liczbą szczegółów
- można definiować różne struktury danych (wektory, macierze, struktury hierarchiczne itp.)

- jest możliwe tworzenie tzw. instrukcji złożonych pozwalających łatwo organizować pętle, wieloczęonowe rozgałęzienia itp.
- zapis programu jest na tyle przejrzysty, że sam tekst może stanowić jego dokumentację
- zwiększa się efektywność programowania.

Ponieważ jedna instrukcja języka wysokiego poziomu jest równoważna kilku, kilkunastu, a nawet kilkudziesięciu instrukcjom języka assemblerowego, więc efektywność pracy programistów programujących w językach wysokiego poziomu jest wyższa niż efektywność programistów posługujących się językami assemblerowymi.

Wprowadzenie języków wysokiego poziomu rozszerzyło krąg ludzi zainteresowanych techniką komputerową, ponieważ komunikacja z komputerem stała się prostsza.

9.13. Jakie są najważniejsze języki wysokiego poziomu?

Do najczęściej spotykanych języków wysokiego poziomu można zaliczyć:

- C++ – dla tego języka istnieją kompilatory (patrz pyt. 9.37) działające pod kontrolą praktycznie każdego systemu operacyjnego, jest elastyczny (umożliwia tworzenie zarówno prostych programów, jak i zaawansowanych aplikacji)
- Pascal – język przeznaczony do tworzenia aplikacji pracujących w systemie DOS, wraz z popularyzacją systemu Windows ustąpił miejsca językowi Delphi
- Delphi – jest przeznaczony dla środowiska MS Windows, jest obiektową odmianą języka Pascal
- Perl – język skryptowy przeznaczony do tworzenia programów w środowisku UNIX
- Java – umożliwia tworzenie apletów umieszczanych na stronach WWW. Aplety są to programy kompilowane przez twórcę, przesyłane z serwera do przeglądarki i tam uruchamiane. W odróżnieniu od aplikacji, aplety nie mają możliwości korzystania z zasobów dyskowych komputera
- Java script – jest rozszerzeniem HTML-u; jest to język skryptowy, program jest interpretowany przez przeglądarkę;
- Magic – program do tworzenia aplikacji korzystających z baz danych.

9.14. Co to jest programowanie strukturalne?

Programowanie strukturalne (ang. *structured programming*) zostało po raz pierwszy zaproponowane przez holenderskiego matematyka Dijkstrę jako programowanie hierarchiczne, metoda zstępującą (ang. *top-down*) z góry na dół.

Najkrócej rzecz ujmując, jest to sposób programowania wymuszający podział problemu na kolejne, coraz bardziej szczegółowe przybliżenia algorytmu. Systematyczny podział problemu na kolejne, coraz prostsze części składowe, których powiązania wzajemne są dobrze określone, ułatwia pracę, umożliwia podział pracy między zespoły oraz ułatwia kontrolę i wykrywanie błędów. Współpraca pomiędzy częściami odbywa się na zasadzie przekazywania parametrów. Ten sposób podejścia jest zalecany zarówno do formułowania algorytmu, jak i do zapisu go w postaci programu.

Wadą programowania strukturalnego jest kłopotliwa zmiana struktury danych (w przypadku, gdy operuje na niej wiele segmentów wymaga to zmiany we wszystkich tych segmentach).

9.15. Jaką postać w języku Pascal miałby przykładowy program znajdowania maksymalnego elementu z ciągu danych n liczb a_1, a_2, \dots, a_n ?

Program szukania maksymalnego elementu z ciągu danych n liczb może mieć w Pascalu następującą postać:

```
program FindMax (input, output);  
const n = 100;  
var licznik: 1..n;  
    A, M: real;  
begin  
    licznik := 1;  
    readln(A);  
    M := A;  
repeat  
    licznik := licznik + 1;  
    readln(A);  
    if A >= M then M := A  
until licznik = n;  
    writeln(M)  
end.
```

Algorytm powyższego programu został już przedstawiony w pyt. 9.4.

W programie tym, o nazwie *FindMax*, parametry *input* oraz *output* informują, że dane liczbowe będą wczytane ze standardowego dla danego komputera urządzenia wejściowego, a wyniki będą wyświetlone na standardowym urządzeniu wyjściowym. Po tym tzw. nagłówku programu pojawia się definicja stałej, ustalająca przykładowo wartość n równą 100 oraz deklaracje typów. Zmienna licznik może przyjmować wartości całkowite od 1 do n włącznie (tzw. typ okrojony), natomiast zmienne A i M mogą być rzeczywiste. Definicja i deklaracje opisują obiekty, na których ten program działa. Część operacyjna programu jest ujęta w nawiasy *begin* i *end*. Rozpoczyna się nadaniem wartości początkowych zmiennym *licznik* oraz M (po wczytaniu A instrukcją *read*). Następnie mamy instrukcję, która powoduje $(n-1)$ -krotne wykonanie instrukcji zawartych między symbolami *repeat* oraz *until* dla zmiennej *licznik* przyjmującej wartości od 2 do n . Instrukcja standardowa *write* powoduje drukowanie (wyświetlanie) wartości zmiennej M .

9.16. Co to jest programowanie obiektowe?

Programowanie obiektowe określane jest też mianem programowania zorientowanego obiektowo. W wykorzystujących programowanie obiektowe programach podstawowym pojęciem jest obiekt. Reprezentuje on procedury, jak i same dane. Działanie programu polega na wzajemnym komunikowaniu obiektów i modyfikacji ich cech. Nowe obiekty tworzone są na podstawie już istniejących. Możliwe jest dodawanie nowych obiektów i dziedziczenie wcześniejszych cech. Programowanie obiektowe wprowadza kilka terminów, które nie pojawiały się we wcześniejszych językach programowania. Klasa jest pewnym wzorcem rzeczy lub działania – zjawiska, którego dotyczy nasz program. Zawiera ona metody, czyli funkcje i procedury, które mogą być stosowane do danej klasy. Konkretnie elementy danej klasy to obiekty, które zawierają w sobie zarówno metody (programy) jak i klasy [14].

9.17. Jaka postać w języku C++ miałby przykładowy program znajdowania maksymalnego elementu ze stu losowych liczb?

Program szukania maksymalnego elementu ze stu liczb może mieć w C++ następującą postać:

```
#include <stdio.h>

int main(void)
{
    int i=1;
```

```
int a=0;
int max;
while (100>=i)
{
    a=random();
    if (max<a)
        max=a;
    i++;
    printf("%d\n",a);
}
printf("\n");
printf("Max: %d",max);
printf("\n");
return 0;
}
```

Algorytm powyższego programu był już prezentowany w pyt. 9.4 (tutaj dla odmiany liczby nie są czytane, a losowane).

Na początku programu znajduje się polecenie *include* dołączające bibliotekę *stdio.h* do programu. Kolejna linia rozpoczyna część główną programu *main*. Program jest umieszczony w nawiasach klamrowych `{ }`. Instrukcje: *int i=0*, *int a=0* oraz *int max* deklarują zmienne całkowite, na których będzie działał program. Instrukcja *while* rozpoczyna pętlę, która będzie wykonana sto razy. Wewnątrz pętli następuje przypisanie zmiennej *a* wartości losowej, porównanie jej wartości z wartością maksymalną, zwiększenie wartości licznika *i* o jeden oraz wydrukowanie wartości *a*. Jeśli porównywana wartość zmiennej *a* jest większa od dotychczasowej wartości maksymalnej, wówczas następuje przepisanie wartości zmiennej *a* do zmiennej *max*. Po osiągnięciu przez licznik *i* wartości 100 następuje zaprzestanie wykonywania pętli, wyświetlenie maksymalnej wartości spośród porównywanych liczb i zakończenie programu.

Komentarza może wymagać składnia instrukcji *printf*. Pierwszy argument określa format wyświetlania liczby, drugi jest argumentem przekazującym do funkcji wartość liczby.

9.18. Jakie są kryteria oceny języków programowania?

Należy podkreślić, że brak jest formalnych i precyzyjnych metod oceny języka programowania i często trzeba się niestety odwoływać do intuicji. Przy takiej właśnie intuicyjnej ocenie języka zwraca się uwagę na:

- prostotę
- łatwość nauczenia się i użytkowania
- czytelność i zwartość struktur programowych
- uniwersalność
- elastyczność
- modularność
- niezależność od komputera, lecz zarazem łatwość przystosowywania do różnych konfiguracji sprzętowych i programowych
- efektywność procesu tłumaczenia programu na ciąg rozkazów.

Niektóre z tych wymagań są wzajemnie sprzeczne, np. im język bardziej uniwersalny, tym trudniejszy do nauczenia.

9.19. Jakie są zasadnicze różnice między językami naturalnymi a językami programowania?

W językach programowania określa się formalnie zbiór reguł, za pomocą których można otrzymać wszystkie poprawne napisy w tym języku. Składnia (syntaktyka) tych języków jest więc sformalizowana.

W językach naturalnych nie mamy zbioru takich reguł, które pozwoliłyby uzyskać wszystkie możliwe zdania tego języka. Należy przy tym podkreślić, że same tylko reguły syntaktyczne nie zapewniają jeszcze sensowności zdań poprawnych syntaktycznie. Semantyka (reguły znaczeniowe) bowiem najtrudniej poddaje się formalizacji.

9.20. Co to są metody numeryczne?

Metody numeryczne są dużym działem matematyki, który dostarcza metod rozwiązywania złożonych zagadnień obliczeniowych, np. obliczanie całek oznaczonych, rozwiązywanie układów równań, szukanie pierwiastków równań przestępnych itp., za pomocą elementarnych operacji matematycznych. Przykładem najprostszym może być obliczanie całki oznaczonej jako sumy wartości funkcji podcałkowej obliczanych z pewnym wymaganym krokiem. Innym przykładem może być znana metoda stycznych, bądź metoda siecznych, rozwiązywania równań przestępnych. Metody nume-

ryczne zajmują się również określeniem dokładności otrzymywanych rozwiązań, gdyż na ogół algorytmy stosowane w metodach numerycznych są przybliżonymi rozwiązaniami problemów [5, 12, 18].

9.21. Czy komputer popełnia błędy arytmetyczne?

Może zdarzyć się taka sytuacja, w której komputer popełni błąd arytmetyczny. Jest to uzależnione od konstrukcji algorytmu i długości słowa, na którym wykonywane są operacje. Przykładowo, jeżeli programista nie przewidzi sytuacji, w której do bardzo małej liczby dodaje się, a następnie odejmuje tę samą dużą liczbę, to wskutek ograniczonej długości słowa procesora nastąpi zaokrąglenie i utrata liczby o małej wartości. Można tego uniknąć poprzez zmianę kolejności działań, tak by w operacjach występowały liczby o zbliżonych zakresach wartości [5, 12, 18].

9.22. Jakie wyróżniamy etapy opracowania programu?

Typowy proces opracowywania programu składa się z następujących kroków:

- analiza problemu i sformułowanie algorytmu
- pisanie (edytowanie) programu w języku programowania i wprowadzenie programu źródłowego do komputera
- weryfikacja programu podczas kompilacji (patrz pyt. 9.37)
- ewentualne poprawianie błędów formalnych (składniowych) wykrytych w kroku poprzednim i ponowne tłumaczenie
- uruchamianie programu i poprawianie błędów logicznych
- opracowanie dokumentacji.

Pierwszy krok jest najważniejszy. Od gruntownej analizy i wyboru właściwego algorytmu zależy pracochłonność dalszych etapów i jakość wyników. Bardzo ważnym czynnikiem jest wybór właściwego języka do zapisywania algorytmu. Wstępna wersja algorytmu powinna zawierać bloki opisujące ogólnie główne fazy przetwarzania, które w kolejnych krokach będą uściślane i szczegółowo rozrysowywane.

Krok drugi to zapisanie algorytmu za pomocą słów i zdań z wybranego języka programowania. Zaleca się tu staranność i systematyczność oraz stosowanie tylko takich konstrukcji, których znaczenie jest dla programisty w pełni zrozumiałe. Wprowadzanie i modyfikowanie tekstów programów źródłowych ułatwiają specjalne edytory. Oprócz bezpośredniego wprowadzania pozwalają one łatwo zlokalizować wybrany fragment tekstu, wstawić lub usunąć określony ciąg znaków.

Następnym krokiem jest próba weryfikacji nowo wprowadzonego programu. Z reguły wynikiem takiej operacji jest informacja o wykrytych błędach. Jest to nieunik-

nione, bo nawet jeżeli z punktu widzenia języka program jest poprawny, to zwykle pomyłki następują przy jego wprowadzaniu.

Poprawiamy wykryte błędy i ponownie weryfikujemy nasz program, a kroki te powtarzamy dopóty, dopóki nie uzyskamy poprawnego wyniku weryfikacji.

Nie znaczy to jeszcze, że program jest poprawny. Oznacza tylko, że z formalnego punktu widzenia użyliśmy prawidłowych konstrukcji języka. O tym, czy program jest poprawny, przekonamy się dopiero po jego uruchomieniu i sprawdzeniu, czy działa zgodnie z naszymi oczekiwaniami. Jeżeli będziemy mieli szczęście, to uruchomiony program da nam od razu prawidłowe wyniki i pozostanie nam tylko opracowanie dokumentacji.

Najczęściej jednak program działa w sposób nieoczekiwany i zacznie się teraz najtrudniejsza faza całego procesu, a mianowicie poszukiwanie błędów logicznych. Polega ona na dokładnym testowaniu wybranych fragmentów programu i sprawdzeniu, czy działają one zgodnie z przewidywaniem; a jeżeli nie działają zgodnie z oczekiwaniami, poszukuje się przyczyn. Wiele systemów dysponuje w tym celu specjalnymi programami pomocniczymi, które ułatwiają testowanie zwłaszcza programów assemblerowych. Pozwalają one np. wykonywać program krok po kroku, instrukcja po instrukcji, z równoczesnym śledzeniem zawartości wybranych komórek. Okazuje się, że faza uruchamiania pochłania zwykle od 2/3 do 3/4 całego czasu potrzebnego na opracowanie programu.

Ostatnim krokiem jest opracowywanie dokumentacji. Nie jest to potrzebne w przypadku małych programów jednostkowych pisanych na własne potrzeby. Wszystkie większe prace, zwłaszcza takie, które mogą być w przyszłości kontynuowane przez kogoś innego, powinny być bardzo starannie udokumentowane. Obowiązkowym elementem stanowiącym minimum dokumentacji powinny być wyczerpujące komentarze w tekście programu źródłowego.

9.23. Jakie cechy powinien mieć dobry program?

Pojęcie dobry program nie jest ściśle określone. Można określić kilka składników wpływających na ocenę programu, są to:

- poprawność
- uniwersalność
- odporność
- stabilność
- optymalność.

Od programu wymaga się poprawności, jednak nie jest to łatwe do spełnienia. Przede wszystkim musi być podane ściśle określenie terminu „dobry program”. Te-

stowanie może wykazać obecność pomyłki, ale nigdy nie wykaze nieobecności błędu. Metody sprawdzania poprawności powinny być podobne do technik dowodów matematycznych, ponieważ stwierdzenia, które chcemy udowodnić, wyraża się w matematyce. Dowód poprawności programu powinien być ustalony jednocześnie z konstrukcją programu, a nie pozostawiany na dalsze etapy.

Na uniwersalność programu w stosunku do zmienionych założeń ma wpływ jego modularna struktura. Chodzi o to, że gdy zmienia się nieznacznie założenia dla programu, bądź rodzaj danych wejściowych, wówczas, bez istotnych przeróbek, wystarczy powinna wymiana jednego z modułów programu.

Zaleca się sprawdzanie, czy działanie programu jest poprawne w przypadku danych wejściowych w ogóle nie przewidywanych. Odporny program po prostu powinien użytkownika powiadomić, że na wejściu pojawiły się dane, których nie przewidywaliśmy.

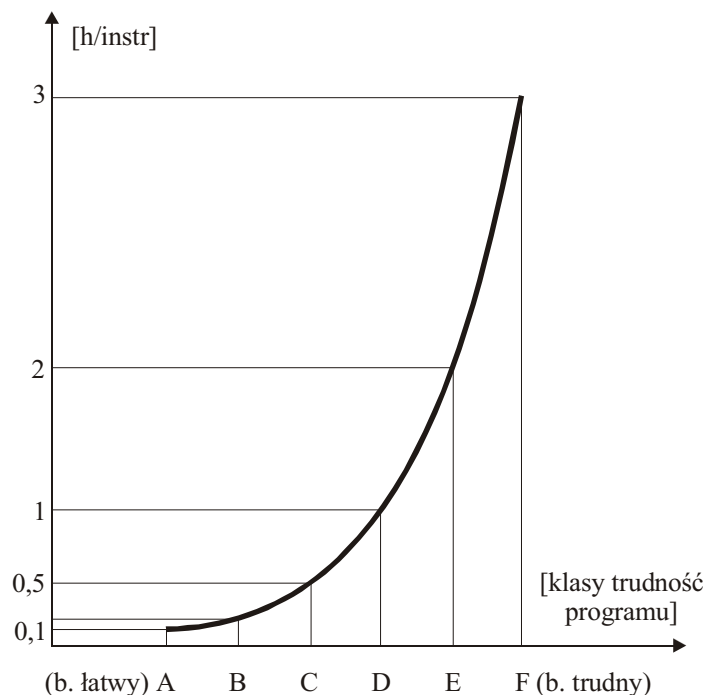
Stabilność jest pojęciem podobnym do odporności i uniwersalności. Program ma tę właściwość, gdy radzi sobie z dużymi klasami danych podlegających zakłóceniom i pochodzących z wielu źródeł. Stabilny program nie powinien przestać funkcjonować w przypadku prostego błędu danych wejściowych.

W programowaniu mamy do czynienia z kompromisem pomiędzy oszczędnością czasu wykorzystywania procesora, oszczędnością pamięci, krótkim czasem reakcji systemu i krótkim czasem oczekiwania na obsługę. Najczęściej wymagania te są wzajemnie sprzeczne. Każdy podręcznik programowania podaje wiele wskazówek pozwalających optymalizować program, np. ze względu na oszczędność pamięci lub ze względu na czas wykonywania tego programu.

9.24. Jaka jest wydajność programistów?

Biorąc pod uwagę przestrożę mówiącą, że „traktowanie osobomiesięcy jako jednostki miary wielkości zadania jest niebezpieczne i złudne”, można spróbować odpowiedzieć na powyższe pytanie. Należy wziąć przy tym pod uwagę fakt, że według niektórych specjalistów czas potrzebny na wykonanie ostatnich 20% prac nad programem jest większy niż na uprzednie 80%.

Poniżej przedstawiona jest (wg Metzelaara) zależność czasu opracowania jednej instrukcji wynikowej (w h/instr.) od stopnia trudności programu (wraz z analizą, projektem, edytowaniem, weryfikowaniem, tworzeniem dokumentacji i uruchamianiem).



Rysunek 9.1. Wykres zależności czasu opracowania jednej instrukcji wynikowej od stopnia trudności programu (A – bardzo łatwy, F – bardzo trudny)

9.25. Jakie znamy sposoby testowania programów?

W praktyce testowanie programu na wszystkich możliwych teoretycznie zbiorach danych jest niemożliwe. Może to uzmysłowić prosty przykład. Załóżmy, że program działa na sześciu danych wejściowych. Każda z tych danych jest reprezentowana przez słowo 16-bitowe. Pełne przetestowanie programu na wszystkich możliwych układach danych ($2^{16 \cdot 6} = 2^{96} \approx 10^{38}$) wymagałoby 10^{38} s przy założeniu, że program wykonuje się jedną sekundę. Na podstawie nawet 100 testowań (wobec 10^{38}) nie można przecież jednoznacznie rozstrzygnąć o poprawności programu. W związku z powyższym na ogół stosuje się trzy sposoby testowania:

- test funkcjonalny
- test według oceny użytkownika
- test według struktury programu.

Test funkcjonalny polega na praktycznym zbudowaniu macierzy zawierającej różne zestawy danych, na których testujemy program, i odpowiadające im wyniki, osobno dla poszczególnych funkcji (fragmentów programu) realizowanych przez program. Jeśli dla danego zestawu danych uzyskujemy dobre wyniki dla danej funkcji, to

stawiamy np. znak plus, w przeciwnym razie znak minus. Z obserwacji takiej macierzy można doszukać się przyczyn nieprawidłowej pracy programu.

Użytkownik, który jako jedyny ma najlepsze wyobrażenie o spodziewanych wynikach programu, porównuje otrzymane wyniki z tymi wyobrażeniami. Proponując informacje wejściowe (próbne) do testowania programu, użytkownik bierze pod uwagę różnorodność tych danych i jej wpływ na wyniki.

Wybór danych wejściowych do testowania może zależeć np. od tego, jaką część programu chcemy sprawdzić. Należy dbać o to, aby poszczególne moduły można było testować etapami, niezależnie od siebie. Jeśli moduły opracowywali różni programiści, wskazane jest przetestowanie modułów oddzielnie przed połączeniem.

9.26. Co to jest oprogramowanie systemu komputerowego i jak je klasyfikujemy?

Oprogramowaniem systemu komputerowego nazywamy zbiór wszystkich programów możliwych do wykonania w tym systemie, czyli zbiór algorytmów zapisanych w języku i w postaci rozumianej przez dany komputer.

Zwykle wyróżnia się dwie podstawowe kategorie oprogramowania:

- oprogramowanie podstawowe,
- oprogramowanie użytkowe (aplikacyjne).

W skład oprogramowania podstawowego wchodzi przede wszystkim system operacyjny zarządzający pracą całego systemu komputerowego, translatory (kompilatory języków programowania, standardowe programy narzędziowe ułatwiające pracę programiście podczas tworzenia i uruchamiania programów, a także programy diagnostyczne ułatwiające lokalizację uszkodzeń.

Na oprogramowanie użytkowe (aplikacyjne) składają się wszystkie programy tworzone we własnym zakresie przez użytkownika komputera (lub dostarczane przez producenta bądź też wyspecjalizowane firmy). Są one przeznaczone do wykonywania określonych zadań wynikających z indywidualnych potrzeb użytkownika. Mogą to być programy rozwiązujące układy równań liniowych, wykonujące obliczenia statystyczne, sterujące linią produkcyjną w zakładzie przemysłowym, sporządzające miesięczne listy płac, obsługujące kasy rejestracyjne w domu towarowym, symulujące lądowanie pojazdu kosmicznego na księżycu i wiele, wiele innych.

Sprzęt wraz z oprogramowaniem podstawowym stanowią konieczne narzędzia umożliwiające zarówno tworzenie, jak i uruchamianie programów użytkowych (aplikacyjnych).

Nie zawsze było to oczywiste, lecz obecnie nikt już nie ma wątpliwości, że oprogramowanie (zwłaszcza podstawowe) stanowi jeden z najistotniejszych zasobów sys-

temu komputerowego traktowanego jako całość. W początkach rozwoju informatyki, gdy możliwości sprzętu były ograniczone, jego koszt wysoki, a praca programisty względnie tania, na pierwsze miejsce wysuwał się sprzęt, jako najdroższy element systemu. Obecnie sytuacja całkowicie się odwróciła i już od kilkunastu lat wartość oprogramowania istotnie przewyższa wartość sprzętu. Jakość oprogramowania decyduje o możliwościach i wygodzie korzystania z komputera, a żaden producent nie jest w stanie utrzymać się na rynku, jeżeli wraz z komputerem nie dostarcza dostatecznie bogatego oprogramowania.

9.27. Co to jest system operacyjny i do czego służy?

System operacyjny jest zbiorem specjalnych programów zarządzających pracą systemu komputerowego i ułatwiających zarówno tworzenie, jak i wykonywanie programów użytkowych. Jest on pośrednikiem między programistą (i jego programem) a rzeczywistym komputerem. Ukrywa on przed użytkownikiem wszystkie zawiłości wewnętrznej architektury komputera i udostępnia mu inny, wygodniejszy i prostszy w użyciu komputer (patrz pyt. 1.8, rys. 1.2).

Do głównych zadań systemu operacyjnego należą:

- optymalizacja wykorzystania zasobów komputera (procesora, pamięci, urządzeń zewnętrznych)
- ochrona programów i danych przed przypadkowym lub celowym uszkodzeniem
- maksymalne uniezależnienie programów użytkowych od ewentualnych zmian w konfiguracji komputera (np. spowodowanych uszkodzeniem któregoś z urządzeń);
- ułatwienie pracy z komputerem poprzez udostępnienie dodatkowych funkcji i zasobów nie istniejących w komputerze bez oprogramowania.

W następnym punkcie prześledzimy pokrótce ewolucję systemów operacyjnych. Pozwoli to łatwiej zrozumieć znaczenie i rolę poszczególnych funkcji systemu.

9.28. Jak doszło do powstania systemów operacyjnych?

Początkowo korzystanie z komputera wymagało nie tylko umiejętności programowania, ale także gruntownej znajomości budowy komputera i sposobu jego obsługi. Poza właściwymi programami użytkowymi i nielicznymi translatorami nie istniało inne oprogramowanie wspomagające. Użytkownik komputera musiał przede wszystkim zarezerwować sobie czas, w którym komputer był tylko do jego wyłącznej dyspozycji. Kolejny programista musiał oczywiście czekać w kolejce. Następnie należało wczytać program źródłowy i rozpocząć tłumaczenie. Po jego zakończeniu programista dostawał zwykle listę błędów, które musiał zlokalizować i poprawić w programie źró-

dłowym. Potem czynności należało powtórzyć od początku. Najczęściej jednak napisany program zatrzymywał się w nieoczekiwanym miejscu lub... nie chciał się zakończyć. Niewiele wówczas można było zrobić. Zwykle brakowało już czasu (inni programiści czekali w kolejce) i najczęściej robiono wtedy wydruk zawartości pamięci, który później należało żmudnie analizować.

Powyższy sposób korzystania z komputera był podwójnie nieefektywny. Kosztowny wówczas komputer był do wyłącznej dyspozycji tylko jednego użytkownika, który tracił znacznie więcej czasu na obsługę urządzeń niż na wykonanie swojego programu. Stopień wykorzystania sprzętu przy takim trybie pracy był minimalny. Pierwszym krokiem zmierzającym do poprawy tej sytuacji było więc zautomatyzowanie czynności przygotowawczych: wprowadzenia programów i danych, drukowania wyników, magazynowania i odszukiwania informacji na nośnikach magnetycznych. Coraz więcej czynności wymagających uprzednio ręcznych manipulacji można było zainicjować pisząc tylko odpowiednie polecenia na tzw. kartach sterujących (papierowe karty dziurkowane) wprowadzanych wraz z programem źródłowym lub na specjalnej konsoli operatorskiej. Procedura obsługi komputera uprościła się na tyle, że programistów zastąpili operatorzy. Zaczęto także stosować biblioteki gotowych programów, w szczególności tych, które realizowały operacje często powtarzające się w różnych programach. Wszystkie te udogodnienia pojawiły się pod koniec lat pięćdziesiątych w wyniku wprowadzenia stosunkowo prostych programów sterujących, zwanych monitorami.

Na początku lat sześćdziesiątych, w związku z istotnymi zmianami konstrukcyjnymi polegającymi na wprowadzeniu przerwań i kanałów wejścia/wyjścia rozbudowano monitory tak, by wykorzystać w pełni nowe możliwości sprzętu. Chodziło głównie o poprawę efektywności wykorzystania procesora poprzez umożliwienie mu pracy podczas trwania operacji wejścia/wyjścia. Te nowe programy sterujące pracą komputera i wykonywaniem programów użytkowych nazwano programami zarządzającymi.

Dalsza ewolucja doprowadziła do powstania programów nadzorujących. Wykonywały one coraz więcej złożonych operacji. Zaczęły rezydować na stałe w pamięci. Pojawiły się także pierwsze mechanizmy zabezpieczające programy i informacje przed niepożądanym, wzajemnym oddziaływaniem. Było to konieczne, gdyż korzystanie z komputera stało się na tyle proste, że pojawili się użytkownicy stosunkowo mało doświadczeni, którzy nieświadomie mogli spowodować znaczne szkody.

Miano właściwego systemu operacyjnego przypisuje się dopiero systemowi OS/360 opracowanemu dla rodziny komputerów IBM 360 w 1965 r. Natomiast w Europie za znaczące uznaje się systemy GEORGE opracowane dla komputerów ICL 1900 (odpowiednik w Polsce – Odra seria 1300) i system THE opracowany przez Dijkstrę na uniwersytecie w Eindhoven. Ten ostatni uważa się za klasyczny przykład systemu hierarchicznego.

9.29. Jaka jest różnica między procesem a programem?

Program w postaci pliku oczekującego na uruchomienie lub zapisanego w pamięci komputera nie jest procesem. Stanie się nim wówczas, gdy zacznie się wykonywać. Może to nastąpić, gdy zostanie uaktywniony przez użytkownika lub inny proces i zacznie wykorzystywać zasoby komputera.

9.30. Na czym polega działanie systemu operacyjnego?

System komputerowy składa się z pewnej liczby zasobów. Mianem tym określa się nie tylko elementy składowe komputera, takie jak procesor, pamięć, urządzenia wejścia/wyjścia, lecz także przechowywane w nim programy oraz dane, a mówiąc ogólnie, informacje.

O dostęp do tych zasobów może ubiegać się równocześnie wiele procesów. Podczas pracy komputera trwa nieustanna rywalizacja między procesami o poszczególne zasoby. Typowym zasobem, który jest rozdzielany pomiędzy różne procesy, jest procesor, a dokładniej jego czas.

Działanie systemu operacyjnego sprowadza się więc do umiejętnego zarządzania zasobami komputera. Musi on efektywnie wykorzystywać poszczególne zasoby, jak i rozwiązywać wszystkie możliwe konflikty wynikające z ich współdzielenia między procesami. W tym celu musi śledzić stan każdego zasobu, decydować o tym, który proces, kiedy i na jak długo może otrzymać dany zasób, a także pilnować, żeby we właściwym czasie uprzednio przydzielone zasoby zostały zwrócone.

Wszystkie zasoby komputera dzieli się na cztery charakterystyczne grupy związane z:

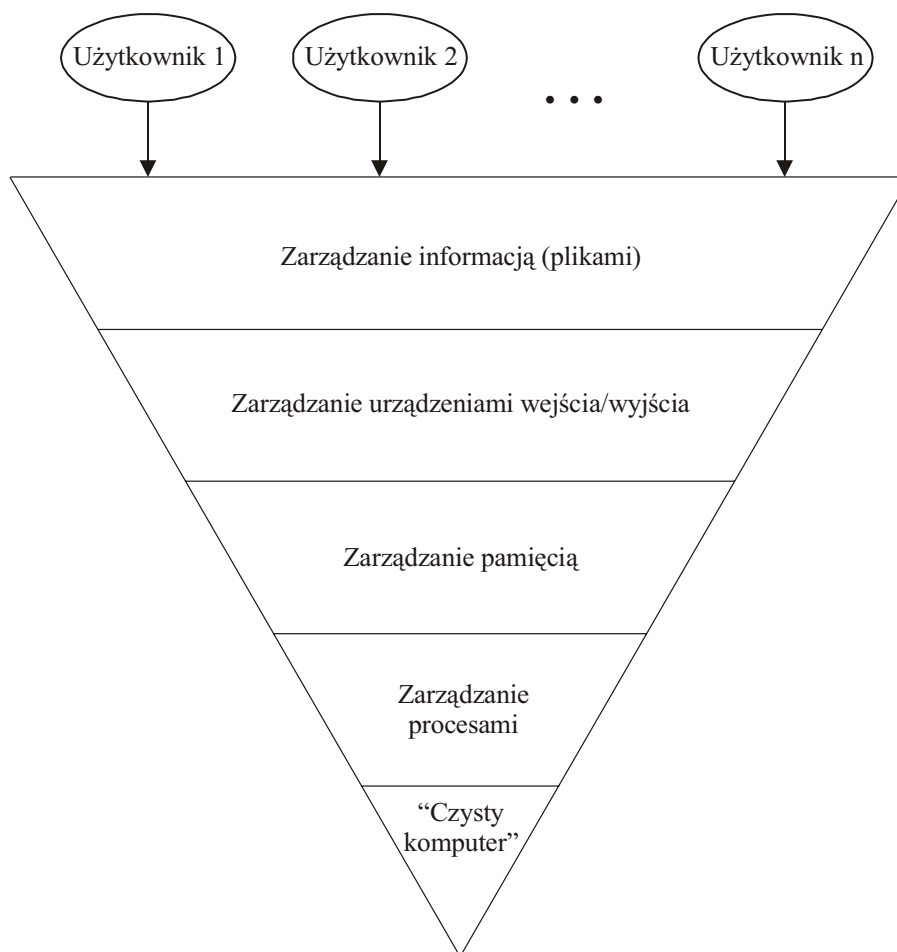
- procesorem
- pamięcią
- urządzeniami
- informacją.

W ten sam sposób grupuje się też zagadnienia związane z zarządzaniem tymi zasobami.

9.31. Jaka jest typowa struktura systemu operacyjnego?

Jednym z częściej stosowanych modeli przy opisie i analizie systemów operacyjnych jest wielopoziomowy model hierarchiczny zwany też warstwowym. Jego zaletą jest przejrzystość, gdyż jest on w sposób naturalny uporządkowany. Budowę takiego modelu rozpoczyna się od „czystego” komputera. Jest to wyłącznie sam sprzęt, bez

żadnego oprogramowania. Zajmuje on najniższy z możliwych poziomów modelu z rys. 9.2. Na tym poziomie pracowano przed wprowadzeniem systemów operacyjnych.



Rysunek 9.2 Wielopoziomowy, hierarchiczny model systemu operacyjnego

Następny poziom tworzą procedury lub programy bezpośrednio związane ze sprzętem i realizujące podstawowe funkcje, wykorzystywane następnie przez pozostałe moduły zarządzające zasobami na poziomach wyższych. Te podstawowe funkcje to przede wszystkim obsługa przerw, w tym m.in. zegara systemowego oraz operacje synchronizujące, niezbędne przy nadzorowaniu przydziału zasobów. Zagadnienia synchronizacji i wzajemnej koordynacji procesów rywalizujących o pewien zasób należą do najważniejszych problemów zarówno teoretycznych, jak i praktycznych. Chodzi bowiem o znalezienie prostych i efektywnych mechanizmów, które wykluczają możliwość równoczesnego zajęcia jakiegoś zasobu przez dwa różne procesy. Jest to sytuacja zwana zakleszczeniem (ang. *deathlock*), w której np. proces A zatrzymuje

się, gdyż oczekuje na drukarkę, aktualnie używaną przez proces B. Proces B jednak przed zwolnieniem drukarki chce skorzystać ze stacji dysków, używanej przez proces A. Nie może z niej skorzystać i zatrzymuje się, czekając na jej zwolnienie, co nigdy nie nastąpi, gdyż proces A właśnie czeka na drukarkę i nie jest w stanie zwolnić stacji dysków.

Kolejny poziom tworzą moduły zarządzania pamięcią. Odpowiadają one za przydzielanie i odbieranie pamięci poszczególnym procesom, a także za ochronę przydzielonych obszarów pamięci przed próbą dostępu z innego procesu.

Na poziomie czwartym odbywa się zarządzanie operacjami wejścia/wyjścia, polegające na śledzeniu stanu urządzeń zewnętrznych i inicjowaniu poszczególnych procesów.

Piąty poziom tworzą moduły zarządzania informacją definiujące system plików. Odpowiada on za gromadzenie i udostępnianie informacji w pamięciach zewnętrznych. Realizuje operacje, takie jak: tworzenie, usuwanie, otwieranie, zamykanie, odczytywanie i zapisywanie plików.

Opisane poziomy systemu operacyjnego definiują jakby nowy, lepszy i wygodniejszy komputer.

9.32. W jaki sposób system operacyjny zarządza czasem procesora?

Konieczność zarządzania czasem procesora pojawiła się z chwilą wprowadzenia wieloprogramowości. Ten tryb korzystania z komputera z definicji wymaga dzielenia czasu procesora między programy wykonywane współbieżnie. Możliwość pseudorównoległej pracy realizuje się przydzielając kolejno procesor według ustalonego kryterium wszystkim programom współistniejącym w systemie. Zagadnienia związane z zarządzaniem czasem procesora są dość złożone, poniżej przedstawione zostaną niektóre kryteria przełączania procesora między programami.

Procesy już istniejące w systemie, czyli takie, których programy już jakoś zostały wprowadzone do komputera i skierowane do wykonania, mogą znajdować się w trzech możliwych stanach:

- GOTÓW (ang. *ready*)
- WYKONYWANY (ang. *running*)
- CZEKAJĄCY (ang. *waiting*).



Rysunek 9.3. Stany procesu

Na rysunku 9.3 pokazano możliwe przejścia między tymi stanami i przyczyny, które je powodują. Proces jest w stanie GOTÓW, gdy jedynym zasobem potrzebnym do tego, by zaczął się wykonywać jest procesor. Z chwilą przydzielenia mu procesora przechodzi do stanu WYKONYWANY. Pozostaje w tym stanie tak długo, jak długo dysponuje wszystkimi potrzebnymi mu zasobami. Gdy tylko zabraknie mu jakiegoś zasobu lub gdy będzie musiał poczekać na zakończenie operacji wejścia/wyjścia, przechodzi do stanu CZEKAJĄCY, w którym pozostaje, dopóki jego potrzeby nie zostaną zaspokojone. Po ich zaspokojeniu przechodzi do stanu GOTÓW i zaczyna rywalizować z innymi procesami o dostęp do procesora.

Do zadań systemu operacyjnego należy nadzorowanie wszystkich tych przejść między poszczególnymi stanami procesu. Każdy z tych stanów charakteryzuje wektor stanu zawierający (mówiąc w skrócie) wszystkie informacje niezbędne do ewentualnego zawieszenia, a następnie wznowienia danego procesu. Są to między innymi zawartości wszystkich rejestrów, informacje o poszczególnych zasobach itd.

System operacyjny tworzy kolejki procesów znajdujących się w stanach GOTÓW i CZEKAJĄCY. W najprostszym przypadku są to kolejki obsługiwane w kolejności napływu klientów (typu FIFO: ang. *First In-First Out* – pierwszy przybył, pierwszy obsłużony). Są jednak możliwe również bardziej wyszukane strategie porządkowania procesów znajdujących się w kolejkach. W szczególności procesom można przydzielać priorytety, które będą decydowały o miejscu w kolejce. Oczywiście proces o najwyższym priorytecie będzie zajmował pierwsze miejsce w kolejce.

Organizacja, nadzorowanie i obsługiwanie kolejki procesów gotowych jest właśnie zadaniem modułów zarządzających czasem procesora. Obsługę kolejek procesów czekających na zasób lub wystąpienie jakiegoś zdarzenia realizują inne moduły, np. zarządzające pamięcią lub urządzeniami wejścia/wyjścia.

9.33. W jaki sposób system operacyjny zarządza pamięcią operacyjną?

Drugim obok procesora bardzo ważnym zasobem komputera, o który ubiegają się procesy, jest pamięć operacyjna. W przeciwieństwie do zarządzania procesorem można zaspokoić żądania kilku procesów równocześnie, przydzielając im różne obszary pamięci. Zarządzanie pamięcią sprowadza się więc do umiejętnego gospodarowania pulą dostępnych obszarów pamięci. Wymaga to m.in.:

- śledzenia stanu zajętości pamięci
- określania strategii przydziału pamięci
- określania strategii zwalniania pamięci
- wzajemnej ochrony przydzielonych obszarów pamięci.

Elastyczne gospodarowanie pamięcią stało się możliwe z chwilą rozdzielenia obszarów adresowych procesu i fizycznych obszarów adresowych, czyli rzeczywistych adresów pamięci. We współczesnych systemach wieloprogramowych proces może być bowiem umieszczony w dowolnym miejscu pamięci. Nie może więc odwoływać się do bezwzględnych adresów fizycznych, gdyż zmieniają się one ze zmianą położenia procesu w pamięci.

9.34. W jaki sposób system operacyjny zarządza informacją?

Zarządzanie informacją w systemie komputerowym przypomina działanie zwykłej biblioteki. Celem głównym jest efektywne przechowywanie, wyszukiwanie i udostępnianie informacji, na którą składają się zarówno programy, jak i dane. Moduły zarządzania informacją muszą więc wykonywać następujące funkcje:

- rejestrowanie i aktualizowanie informacji określających miejsce przechowywania oraz stan poszczególnych zasobów informacyjnych zgromadzonych w systemie
- określanie strategii dotyczących sposobu przechowywania, metod udostępniania i ochrony informacji
- przydzielanie i zwalnianie poszczególnych zasobów informacyjnych.

Przed bliższym omówieniem szczegółów należy zdefiniować pewne pojęcia charakteryzujące zasoby informacyjne: rekord i plik.

Rekordem logicznym nazywamy jednostkę informacji mającą pewne znaczenie dla programisty. Może to być np. pojedyncza liczba, ciąg znaków stanowiący jedną instrukcję programu źródłowego, ciąg 3 liczb definiujących wektor w przestrzeni trójwymiarowej lub dowolna mniej lub bardziej złożona struktura danych mająca pewne znaczenie logiczne.

Plikiem (ang. *file*) nazywamy pewien ciąg rekordów logicznych. Zawiera on zbiór informacji jednorodnych, np. program źródłowy, który jest ciągiem instrukcji, lub zestaw dokumentów opisujących transakcje magazynowe w pewnym przedsiębiorstwie w ciągu miesiąca.

Rekordem fizycznym nazywamy fizyczną jednostkę informacji, którą można przesyłać do i z pamięci zewnętrznej lub urządzenia zewnętrznego. W przypadku pamięci dyskowych będzie to ciąg bitów mieszczący się np. w jednym sektorze lub na jednej ścieżce.

W zależności od właściwości fizycznych urządzenia pamiętającego rekordy logiczne mogą zajmować część, dokładnie jeden, lub też kilka rekordów fizycznych.

Wszystkie informacje przechowywane w systemie komputerowym muszą mieć postać plików. Zagadnienia zarządzania informacją można więc sprowadzić do zagadnień gospodarowania systemem plików (ang. *file system*). System plików musi przede wszystkim umożliwiać identyfikowanie i lokalizację potrzebnych informacji. Pojedynczy plik jest identyfikowany przez swą nazwę, ściślej przez identyfikator zawierający właściwą nazwę, np. SORT, DANE, WYNIKI oraz dodatkowy ciąg znaków określający typ informacji przechowywanej w danym pliku. Na przykład skróty: GIF, JPG, BMP mogą oznaczać pliki zapisane w różnych formatach graficznych. Części nazwy oddziela kropka.

Zbiór plików przechowywanych w jednym miejscu jest opisany za pomocą katalogu i zawiera:

- identyfikator pliku
- atrybuty pliku określające organizację i definiujące prawa dostępu do pliku
- informacje lokalizujące fizyczne rekordy pliku
- informacje pomocnicze, takie jak data i czas utworzenia pliku, data i czas ostatniej modyfikacji.

Przez organizację pliku rozumie się dwa pojęcia: organizację fizyczną określającą sposób rozmieszczenia pliku w pamięci fizycznej i organizację logiczną rozpatrywaną z punktu widzenia metody dostępu do danych.

9.35. W jaki sposób system operacyjny zarządza urządzeniami i operacjami wejścia/wyjścia?

Do podstawowych zadań modułów zarządzających urządzeniami wejścia/wyjścia należy:

- śledzenie stanu urządzeń i jednostek sterujących
- przydzielanie i zwalnianie tych zasobów zgodnie ze strategią właściwą dla danego typu zasobu
- bezpośrednia obsługa urządzeń za pomocą specjalnych programów standardowych
- udostępnianie użytkownikom komputera urządzeń wirtualnych.

Pierwsze dwie grupy zadań nie odbiegają istotnie od podobnych zadań realizowanych w innych modułach zarządzających zasobami. Nie oznacza to, że są one proste i łatwe. Istotne cechy charakteryzujące zarządzanie operacjami wejścia/wyjścia najlepiej jednak uwidaczniają się w dwóch ostatnich grupach zadań.

Standaryzacja obsługi urządzeń zewnętrznych stanowiła istotny krok naprzód w zakresie ułatwienia korzystania z komputera. Standardowe programy obsługi urządzeń bezpośrednio sterują ich pracą, obsługują przerwania i reagują na błędy. Z punktu widzenia systemu operacyjnego programy obsługi urządzeń są procesami działającymi w sposób ciągły.

Definiują one urządzenia logiczne, identyfikowane nazwami. Ten fakt ma olbrzymie znaczenie z punktu widzenia użytkownika. Gdyby program, który czyta sekwencyjnie rekordy, przetwarza je i sekwencyjnie wyświetla wyniki, miałby być wykonywany w komputerze, który nie ma standardowych programów obsługi urządzeń, znaczna część programu musiałaby realizować tę obsługę we własnym zakresie. Co więcej, program ten mógłby działać tylko na określonym zestawie sprzętu. Tymczasem w systemie ze standardową obsługą urządzeń można program napisać tak, że będzie działał z różnymi urządzeniami. W zależności od konkretnych rozwiązań wystarczy tylko zmienić nazwy urządzeń logicznych.

Drugim istotnym udogodnieniem jest tworzenie nowych, wirtualnych urządzeń wejścia/wyjścia. Odnosi się to zwłaszcza do urządzeń wykorzystywanych wspólnie, takich jak np. drukarka. W przypadku drukarki nie można oczywiście dopuścić, żeby wydruki produkowane przez różne programy przeplatały się co kilka wierszy. Przydzielenie drukarki na wyłączność jednemu programowi i przekazywanie jej następnemu dopiero po zakończeniu pracy przez poprzednika też jest nie do przyjęcia, gdyż stopień wykorzystania urządzenia byłby niewielki. Właściwym rozwiązaniem tego problemu jest zdefiniowanie urządzeń wirtualnych. Można zdefiniować tyle drukarek wirtualnych, ile jest programów, które ich potrzebują. Wydruki ostatecznie ukazują się na tej samej rzeczywistej drukarce, lecz każdy z programów działa tak, jakby miał

własną. Za realizację urządzeń wirtualnych odpowiada moduł nazywany spoolerem (ang. *Simultaneous Peripheral Operation On-Line*). Powoduje on buforowanie na dysku wydruków pochodzących od poszczególnych programów i ustawia je w kolejce do rzeczywistej drukarki.

9.36. W jaki sposób przebiega komunikacja operatora komputera z systemem operacyjnym?

Użytkownik komputera komunikuje się z systemem operacyjnym zazwyczaj przy użyciu klawiatury i myszki, a wyniki może obserwować na ekranie monitora lub na drukarce.

System operacyjny posiada graficzny lub tekstowy interfejs, który ułatwia użytkownikowi komunikację z komputerem. Obecnie najczęściej spotyka się interfejsy graficzne, korzystające z intuicyjnych metod komunikacji (MS Windows, X-windows). Interfejsy tekstowe są popularne wśród systemów UNIX. Choć nie wyglądają tak przystępnie jak ich graficzni konkurenci, to są często wykorzystywane przez zaawansowanych użytkowników.

Za pośrednictwem systemu operacyjnego użytkownik komputera może:

- uruchamiać i zatrzymywać programy
- instalować jakiegokolwiek oprogramowanie w systemie
- zarządzać urządzeniami zewnętrznymi i przydzielać je dla poszczególnych programów
- przydzielać priorytety poszczególnym programom
- wykonywać operacje na plikach.

System operacyjny wyświetla na monitorze komunikaty przeznaczone dla użytkownika. Komunikaty te z reguły stanowią odpowiedź na zlecenie bądź też są informacją o pewnych zdarzeniach.

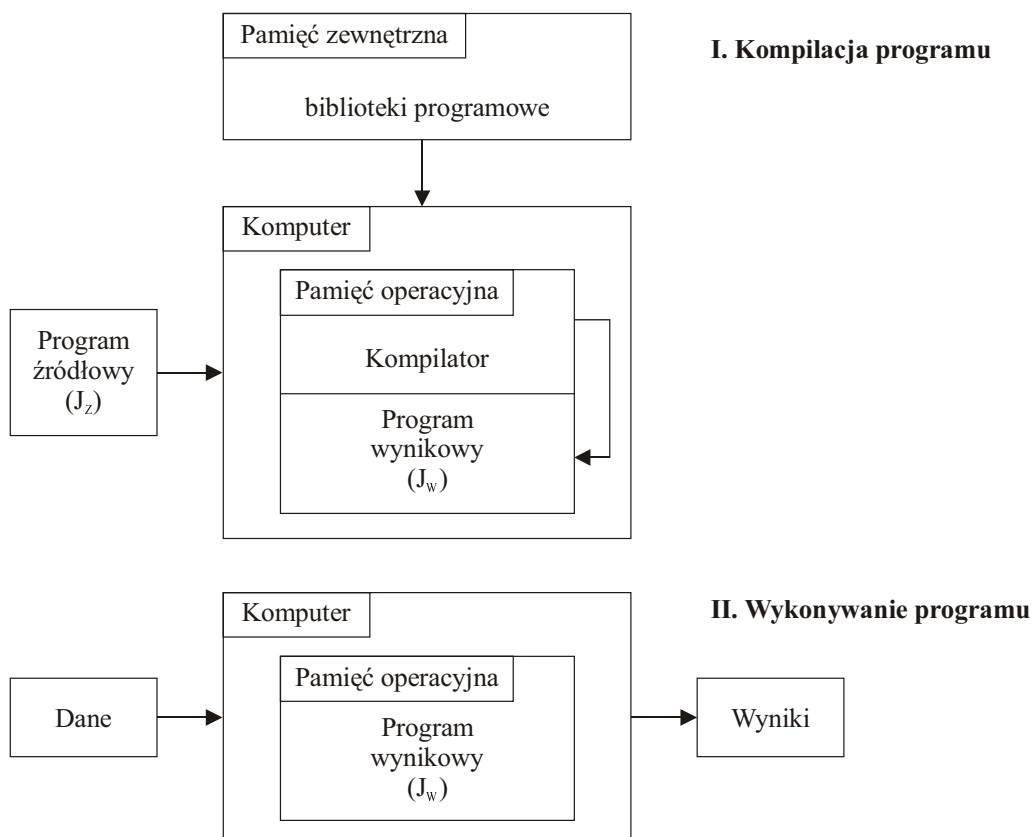
9.37. Co to są translatory i jaka jest różnica między kompilatorami a interpreterami?

Translator to „tłumacz” programu napisanego w języku wysokiego poziomu lub assemblerowym na program w języku wewnętrznym. Wyróżnia się kompilatory i interpretery.

Przy kompilacji uzyskuje się w pamięci komputera przekład całego programu źródłowego, zwany programem wynikowym. Dopiero po zakończeniu procesu kompilacji można przystąpić do wykonania tego programu. Translacja interpretacyjna po-

lega na tym, że tłumaczenie poszczególnych jednostek składniowych programu źródłowego następuje wtedy, gdy przychodzi kolej ich wykonania, czyli można wykonywać program, nie czekając na zakończenie translacji całego programu źródłowego. Mamy tu więc do czynienia z takim trybem translacji, który można streścić w sformułowaniu „czytaj fragment programu, tłumacz go i wykonaj”. W technice interpretacji nie mamy więc do czynienia w ogóle z pojęciem programu wynikowego, które jest charakterystyczne jedynie dla techniki kompilacyjnej. Technika kompilacyjna jest szybsza, a technika interpretacji prostsza. Różnica szybkości w obu technikach jest widoczna w programach zawierających pętlę, gdyż w technice interpretacyjnej tłumaczenie jest wykonywane tyle razy, ile istnieje obiegów pętli.

Technikę kompilacyjną symbolicznie ilustruje rysunek 9.4.



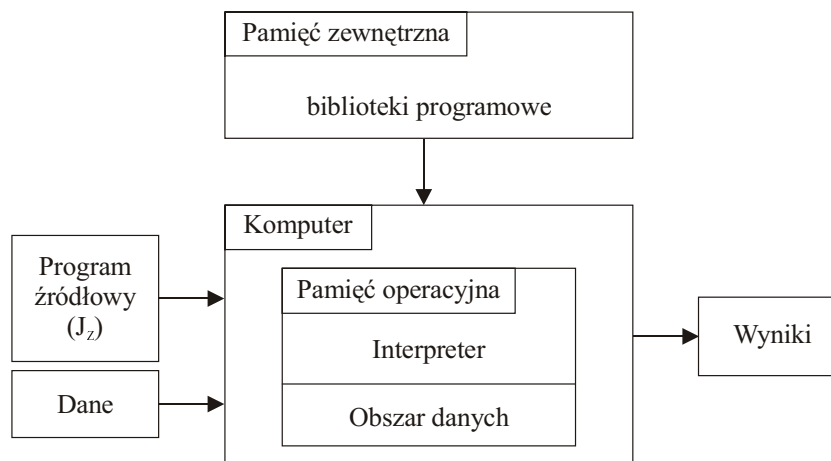
Rysunek 9.4 Kompilacja

Z rysunku 9.4 wynika, że program zwany kompilatorem traktuje napisany przez użytkownika program w języku źródłowym (J_z) tak jak dane do programu. Wynikiem działania kompilatora jest program zapisany w języku wewnętrznym (J_w). W procesie kompilacji są także wstawiane w odpowiednie miejsca programu wynikowego typowe

podprogramy biblioteczne, jeśli ich użycia żąda program źródłowy. Podprogramy biblioteczne mogą być napisane w języku wewnętrznym, asemblerowym lub wysokiego poziomu. W dwóch ostatnich przypadkach konieczne jest także ich tłumaczenie.

Interpreter, jak to przedstawiono na rysunku 9.5, czyta zarówno program źródłowy, jak i dane do tego programu i dokonuje tłumaczenia, po czym wykonuje fragment programu. Kroki te kolejno powtarza, aż do zakończenia programu.

W wielu istniejących translatorach z powodzeniem łączy się zasady kompilacji i interpretacji. W praktyce rzadko dokonuje się bezpośredniej translacji programów z języka źródłowego (wysokiego poziomu) na język wewnętrzny. Bardzo często ogniwem pośrednim jest język asemblerowy. Oznacza to, że translatory najpierw dokonują tłumaczenia z języka wysokiego poziomu na język asemblerowy, a następnie z języka asemblerowego na język wewnętrzny. W ten właśnie sposób, przy wielu różnorodnych językach wysokiego poziomu, uzyskuje się jednolitą formę przejściową dla wszystkich programów i wówczas drugi etap translacji przebiega identycznie dla wszystkich programów, niezależnie od ich postaci źródłowej. Zastosowanie tego typu translacji umożliwia pisanie jednego programu źródłowego w ten sposób, że jego poszczególne fragmenty są pisane w różnych językach wysokiego poziomu, najbardziej odpowiednich do specyfiki obliczeń stosowanych w danym fragmencie.



Rysunek 9.5. Interpretacja

9.38. Co to jest odwrotna notacja polska (ONP)?

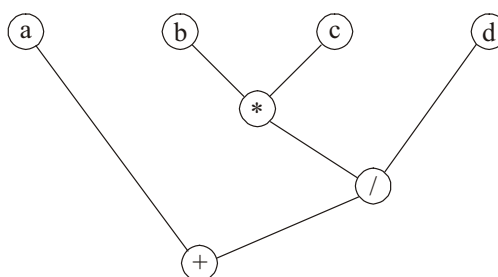
Odwrotna notacja polska (ONP) jest jednym z wariantów beznawiasowego zapisu wyrażeń formalnych, głównie arytmetycznych. Autorem metody jest polski logik J. Łukasiewicz (1878-1956).

Wiadomo, że zapisane działania arytmetyczne nie muszą być wykonywane po kolei, od lewej ku prawej stronie, gdyż o kolejności ich wykonywania decydują także nawiasy. Na przykład zapis działania $a+b/c$, wcale nie oznacza, że najpierw trzeba wykonać dodawanie (pierwszy operator od lewej). Beznawiasowa konwencja zapisu wyrażeń arytmetycznych polega na tym, że zamiast pisać znaki działań (operatory) między argumentami tych działań (operandami), piszemy je za argumentami. Poniżej, dla ilustracji, podajemy kilka przykładów (tab. 9.1).

Tabela 9.1. Beznawiasowa konwencja zapisu działań przy użyciu ONP

Zapis tradycyjny	ONP
$a+b$	$ab+$
$(a-b)*c$	$ab-c*$
$a+b * c/d$	$abc * d/+$

Przeglądanie wyrażenia zapisanego w ONP od strony lewej do prawej pozwala wykonać działanie stojące za dwoma argumentami, lecz wynik tego działania musi być dalej traktowany jako argument kolejnego działania. Łatwo zauważyć, że ONP nie zmienia kolejności występowania argumentów (operandów), lecz może zmieniać w zapisie kolejność znaków działań (operatorów) w stosunku do pierwotnego wyrażenia. W zapisie $abc * d/+$ kolejność działań będzie taka jak na rysunku 9.6.



Rys. 9.6. Kolejność działań w zapisie $abc * d/+$

9.39. Dlaczego stosujemy narzędzia wspomagające tworzenie programów?

Narzędzia wspomagające tworzenie programów stosowane są po to, by ułatwić pracę programistom przez zautomatyzowanie niektórych funkcji, np. deklarację zmiennych. Coraz częściej, ze względu na stopień złożoności, bez narzędzi wspomagających nie byłoby w ogóle możliwe powstanie programu. Następną zaletą, wynikającą ze stosowania narzędzi wspomagających, jest zapewnienie kompatybilności między różnymi platformami programowymi i sprzętowymi, np. język C. Wspomaganie ułatwia proces wyszukiwania i usuwania błędów. Dzięki niemu programy mogą reprezentować wyższą jakość. Programy wspomagające stosowane są również przy powstawaniu dużych projektów, nad którymi pracuje wiele osób.

9.40. Co rozumiemy pod pojęciem CASE?

CASE (ang. *Computer Aided Software Engineering*) jest to komputerowo wspomagana inżynieria oprogramowania. Zapewnia ona środowisko i narzędzia, które są niezbędne do tworzenia programu.

Systemy CASE najczęściej są specjalizowane i ukierunkowane na wspomaganie tworzenia baz danych bądź systemów czasu rzeczywistego.

Narzędzia CASE działają na dwóch poziomach: niższym i wyższym. Do niższego poziomu zaliczają się generatory aplikacji przyspieszające pisanie oprogramowania. Narzędzia CASE wyższego poziomu ułatwiają projektowanie, tworzenie, pomagają także w testowaniu i bieżącej konserwacji oprogramowania.

Oczekiwania wobec narzędzi CASE są następujące:

- wspomaganie w ciągu pełnego cyklu życia systemu
- uniwersalność stosowanych metodologii, tak by było możliwe samodzielne ich definiowanie lub modyfikowanie
- wspieranie nowoczesnych cykli i metodologii
- zapewnianie wysokiej jakości na każdym etapie prac
- integracja ze środowiskiem wytwarzania
- łatwa asymilacja w przedsiębiorstwie
- bezawaryjna praca.

Wszystko to ma istotny wpływ na klasy narzędzi, z którymi mamy dziś styczność [31].

10. Ewolucja komputerów i ich zastosowania

10.1. W jakich dziedzinach stosujemy komputery?

Obszar zastosowań komputerów zmieniał się wraz z rozwojem techniki. Początkowo komputery były stosowane do wykonywania złożonych i żmudnych obliczeń. Wtedy urządzenia te były bardzo kosztowne, miały duże wymiary, były dość zawodne i mało popularne.

Spadek ceny, miniaturyzacja elementów oraz wzrost niezawodności spowodowały upowszechnienie komputerów. Obecnie trudno znaleźć dziedzinę, w której komputery nie znalazły jeszcze zastosowania.

Jako przykłady zastosowań można wymienić kilka dziedzin:

- komunikacja (poczta elektroniczna, wideokonferencje)
- obliczenia naukowe (matematyka, fizyka, astronomia, biologia, chemia)
- projektowanie wspomagane komputerowo – CAD (ang. *Computer Aided Design*)
- wytwarzanie wspomagane komputerowo – CAM (ang. *Computer Aided Manufacturing*)
- symulacje komputerowe
- gromadzenie i wyszukiwanie informacji (bazy danych)
- systemy nauczania komputerowego
- zdalne nauczanie z wykorzystaniem sieci komputerowych (ang. *Distance Learning*)
- wspomaganie pracy biurowej (edytory tekstu, arkusze kalkulacyjne)
- diagnostyka medyczna
- sterowanie procesami przemysłowymi
- telekomunikacja (telefony, centrale, nadzór, projektowanie).

Oczywiście powyższa lista nie jest kompletna i przykładów można znaleźć więcej, gdyż ludzie ciągle wyszukują nowe obszary, w których można zastosować komputery.

Istotne jest to, że zdezaktualizował się niezwykle szybko klasyczny w latach osiemdziesiątych podział zastosowań komputerów (obliczenia naukowo-techniczne, przetwarzanie danych ekonomicznych, symulacje komputerowe i sterowanie procesami technologicznymi). Dziś taka klasyfikacja jest już anachroniczna, a to przede wszystkim za sprawą wejścia komputerów osobistych w nasze życie codzienne.

10.2. Co kryje się pod pojęciem multimedia?

Pojęcie to określa techniki komputerowe umożliwiające prezentowanie rzeczy i zjawisk poprzez: dźwięk, obraz, animację, muzykę, tekst czy słowo mówione. Multimedia to połączenie różnych technik prezentacji.

Techniki multimedialne (łącznie dźwięk i wideo) wymagają dużych pojemności nośników oraz szybkich kanałów transmisyjnych. Prezentacje multimedialne często przechowywane są na dyskach CD-ROM i DVD. Do prowadzenia wideokonferencji wysokiej jakości wymagane są kanały o prędkościach transmisji kilkudziesięciu Mb/s.

Techniki multimedialne znajdują zastosowanie m.in. w:

- systemach konferencyjnych (wideotelefony, wideokonferencje)
- przesyłaniu obrazów wysokiej jakości (np. medycznych)
- wspomaganiu nauczania (encyklopedie, słowniki)
- telewizji (cyfrowej lub na żądanie – ang. *TV, Video on Demand*).

10.3. Jakie są tendencje rozwoju technik multimedialnych?

Przyszłości multimedii należy poszukiwać w rozwoju szybkich sieci oraz w coraz większym stopniu kompresji obrazów i dźwięków. Wynika to ze zwiększonego zapotrzebowania na transmisję wysokiej jakości. Zwiększenie stopnia kompresji umożliwi przesyłanie wielu sygnałów wideo wysokiej jakości tym samym medium. Stanie się to niezbędne po wprowadzeniu usługi wideo na żądanie.

Wystąpi również konieczność doprowadzenia zakończenia szybkiej sieci transmisji danych bezpośrednio do użytkownika. Może to być zrealizowane poprzez wykorzystanie sieci telewizji kablowej, istniejącej instalacji telefonicznej – ADSL (ang. *Asymmetric Digital Subscriber Line*) lub doprowadzenie światłowodu bezpośrednio do mieszkania.

10.4. Co to jest sztuczna inteligencja?

Sztuczna inteligencja (ang. *artificial intelligence*) zajmuje się konstruowaniem algorytmów mających znamiona inteligencji człowieka. Inteligencja jest zdolnością interpretowania zachodzących zjawisk, umiejętnością korzystania z nabytych doświadczeń, zdolnością określania celów oraz sposobów ich osiągnięcia.

Algorytmy sztucznej inteligencji mają zdolność uczenia się, podejmowania skomplikowanych decyzji w zmiennych warunkach, rozumowania abstrakcyjnego oraz zdolność analizy i syntezy.

Twórcy sztucznej inteligencji prezentują wiele nurtów. Najbardziej radykalne podejście zakłada powstanie komputera równoważnego ludzkiemu mózgowi lub go przewyższającego.

Sztuczna inteligencja znajduje zastosowanie w:

- rozpoznawaniu obrazów i dźwięków
- przeszukiwaniu baz danych
- systemach eksperckich.

10.5. Co to są systemy eksperckie?

System ekspercki jest programem wykorzystującym algorytmy sztucznej inteligencji oraz obszerną wiedzę, wprowadzoną przez ludzi (najlepszych ekspertów). Wiedza, jaką ma do dyspozycji program, jest uważana za model ekspertyzy. Wiadomościami takimi dysponuje jedynie wąskie grono wysoko opłacanych specjalistów.

Przyczyną powstania programów eksperckich jest chęć obniżenia kosztów i przyspieszenia pracy. Programy takie działają znacznie szybciej niż ludzie, mogą działać bez przerwy (nie muszą odpoczywać) i są obiektywne.

Przed przystąpieniem do tworzenia systemu eksperckiego należy przeprowadzić analizę, czy jego budowa jest opłacalna. Może się okazać, że gromadzenie wiedzy ekspertów, tworzenie bazy danych i konstruowanie systemu pochłonie większe koszty niż zatrudnienie grona specjalistów.

10.6. Czy współczesne komputery mogą rozpoznawać mowę?

Istnieje wiele systemów potrafiących rozpoznawać mowę. Komputer w inny sposób niż człowiek analizuje dźwięki. Człowiek potrafi dopowiedzieć sobie niedosłyszane fragmenty, może też domyślać się kontekstu wypowiedzi.

Najprostsza metoda rozpoznawania sygnałów polega na porównywaniu próbek głosu z określonymi wzorcami. Ponieważ sygnały wzorcowy i badany różnią się, podczas porównywania komputer ocenia ich podobieństwo, korzystając np. z sieci neuronowych. Badana próbka głosu ma zmienianą amplitudę, jest rozciągana i ściskana w dziedzinie czasu. Operacje te są niezbędne ze względu na zmienną głośność i szybkość mówienia. Program tego typu należy nauczyć głosu mówcy, który ma być rozpoznawany (trenowanie programu). Trening polega na wypowiedaniu pojedynczych fraz, które program ma rozpoznawać, lub czytaniu zadanego fragmentu tekstu.

Rozpoznawanie mowy jest obecnie wprowadzane w infoliniach i służy wybieraniu określonych pozycji w ofercie firmy. Systemy tego typu stosowane są również w niektórych telefonach komórkowych, umożliwiając wydawanie poleceń głosem.

Systemy te są dość proste i reagują na pojedyncze frazy. Znacznie większe problemy sprawia rozpoznawanie spontanicznej wypowiedzi, w której przerwy pomiędzy frazami nie występują. Identyfikacja takiego tekstu musi być oparta na znacznej ilości informacji kontekstowych (wynikach fragmentu wypowiedzi już przetworzonej).

10.7. Jak działają urządzenia analizujące obrazy? Jakie mają zastosowania?

Analiza obrazu polega na pozyskiwaniu informacji przedstawionych w formie obrazu. Obraz w formie cyfrowej jest reprezentowany jako tablica punktów o określonych kolorach. Rozpoznawanie obrazu polega na szukaniu związków pomiędzy punktami tablicy, usuwaniu lub uwypuklaniu niektórych cech szczególnych.

Przetwarzanie obrazu można prowadzić na wiele sposobów. Najprostszą metodą jest podstawianie kolorów. Prosta zmiana wartości elementów tablicy reprezentujących obraz może powodować uwypuklenie poszukiwanej informacji.

Kolejną grupę metod analizy obrazów stanowią operacje arytmetyczne i logiczne na obrazach. Takiej obróbce można poddawać pojedynczy obraz lub grupę obrazów. Operacje na jednym obrazie są wykonywane w celu zmiany jasności obrazu, zmiany rozdzielczości lub barw opisujących pojedynczy punkt.

Obraz może być też poddawany korelacji. Polega ona na przetworzeniu obrazu filtrem, którego parametry są opisywane przez macierz kwadratową. Wartość przetwarzanego punktu jest sumą wartości punktu bieżącego oraz punktów sąsiednich, pomnożonych parametrami macierzy filtru. Poprzez zastosowanie odpowiedniej macierzy można:

- usunąć szumy z obrazu (filtr dolnoprzepustowy)
- poprawić ostrość obrazu (filtr górnoprzepustowy)
- wykryć krawędzie obiektów.

Analizę obrazu można przeprowadzić w dziedzinie częstotliwości. Opisu tego dokonuje się przy użyciu transformaty Fouriera. Mówi ona o rozkładzie elementów o różnych częstotliwościach w obrazie.

Rozpoznawanie obrazów może mieć zastosowanie np. w:

- medycynie (wspomaganie diagnostyki – analiza zdjęć)
- wojsku (rozpoznawanie kształtu czołgów, samolotów – swój czy obcy)
- policji (wyszukiwanie określonej twarzy w tłumie)

- meteorologii (analiza zachmurzenia na podstawie zdjęć satelitarnych)
- rozpoznawaniu pisma (programy OCR).

10.8. Co to są sieci neuronowe?

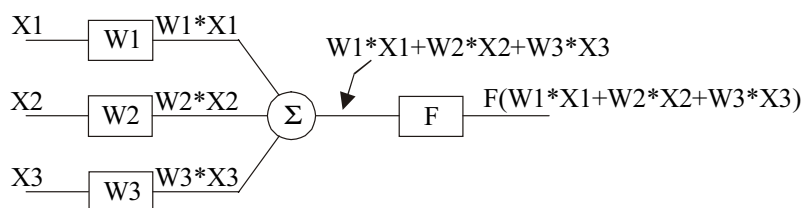
Pierwowzorem sztucznej sieci neuronowej jest układ nerwowy, którego głównymi elementami są komórki nerwowe i połączenia między nimi. Mózg człowieka składa się z ok. 10^{10} komórek nerwowych i 10^{15} połączeń między nimi.

Prace nad biologicznym układem nerwowym, zachowaniami pojedynczej komórki i grupy komórek doprowadziły do powstania sztucznych sieci neuronowych.

Do najważniejszych właściwości tych sieci należą:

- umiejętność uczenia się
- zdolność przewidywania, prognozowania i uogólniania
- umiejętność rozpoznawania klas informacji według wyuczonego wzorca
- wyszukiwanie związków zachodzących między informacjami wejściowymi
- zdolność do wyszukiwania rozwiązań optymalnych
- odporność na uszkodzenia (uszkodzenie pojedynczych neuronów ma niewielki wpływ na zniekształcenie wyniku).

Pojedynczy sztuczny neuron składa się z sumatora, do którego dostarczane są sygnały wejściowe przemnożone przez wagi. Na wyjściu sumatora znajduje się blok aktywacji (element kształtujący odpowiedź neuronu, zależnie od sygnału z wyjścia sumatora).



Rysunek 10.1. Model sztucznego neuronu

Aby sieć neuronowa mogła funkcjonować, musi zostać poddana nauczaniu. Podczas tego procesu dobierane są współczynniki (wagi) sumatorów. Nauka sieci może być prowadzona samodzielnie lub „z nauczycielem”. Nauka samodzielna polega na prezentowaniu sieci różnych informacji, na które reakcja sieci jest wzmacniana. Po odpowiednio długim okresie nauki reakcja sieci na różne klasy informacji różni się

znacznie. Nauka z nauczycielem polega na dostarczaniu sieci informacji wejściowych wraz z pożądanymi wynikami.

Po zakończeniu nauki sieć odpowiada prawidłowo, z określonym prawdopodobieństwem, na nowe informacje.

Sieci neuronowe znajdują zastosowanie przede wszystkim w rozpoznawaniu obrazów (analiza pisma, prognozowanie pogody) i systemach eksperckich (wspomaganie podejmowania decyzji, pomoc w projektowaniu).

10.9. Co to są komputery optyczne?

Zapewne wyprodukowanie komputera w pełni optycznego jest jeszcze odległe. Trudno sobie wyobrazić rejestr przechowujący informacje w postaci strumienia światła, czy bramkę logiczną sterowaną światłem. Możliwe jest stosowanie połączeń hybrydowych (elektronika w połączeniu z optyką). Elementy optyczne mogą być wykorzystywane np. jako szyny danych o bardzo dużych prędkościach transmisji. W taki sposób są już łączone układy scalone superkomputerów równoległych. Pojedyncze włókno tak wykonanej szyny ma prędkość transmisji kilku Gb/s przy odległości do kilkuset metrów.

Przetwarzanie optyczne jest już obecnie powszechnie używane do transmisji danych w sieciach komputerowych. Dzięki zastosowaniu światłowodów możliwa jest prędkość transmisji 2,5 Gb/s na odległości wielu kilometrów. Można spodziewać się dalszego wzrostu szerokości pasma transmisyjnego światłowodów (teoretyczna granica wynosi 25 Tb/s).

10.10. Jaki wpływ mają komputery na społeczeństwo?

Koniec wieku XX przyniósł wiele zmian w życiu ludzi. Jedną ze wspaniałych zdobyczy ludzkości okazał się komputer i komputeryzacja. Nowoczesność urządzeń oraz ich wielofunkcyjność sprawiła, że komputery obecnie spotyka się na każdym kroku. Oczywiście komputer jak każdy wynalazek ma swoje dobre i złe strony. Niektóre pozytywne aspekty wprowadzenia komputerów to:

- pomoc w pracy wielu firm i instytucji
- łatwość pozyskiwania informacji
- ułatwienie komunikacji międzyludzkiej.

Bez komputerów nie udałoby się człowiekowi XX wieku polecieć na Księżyc, przeszczepić serca czy korzystać z telefonii komórkowej. Zapewne bez komputerów ludzie byłiby ubożsi o wiele wrażeń.

Z drugiej strony komputer jest tylko urządzeniem, przy którym wiele milionów ludzi zasiada codziennie do pracy i zabawy na długie godziny. Niektórzy robią to bez opamiętania stając się maniakami komputerowymi. Oczywiście stwarza to niespotykane dla ludzi zagrożenia. Wymienianie dolegliwości, które wiążą się z pracą przy komputerze, zapewne byłoby długie. Najważniejszymi skutkami są:

- zmiany zwyrodnieniowe układu kostno-stawowego
- uszkodzenia i zespoły bólowe mięśni
- upośledzenie ostrości wzroku i zaburzenia widzenia
- zaburzenia psychiczne i uzależnienie
- impotencja i bezpłodność.

Wiele zaburzeń, które pojawiają się przy długotrwałej pracy z komputerem, nie zostało do tej pory zakwalifikowanych do chorób zawodowych. Siedząc przed komputerem należy pamiętać o odpowiednim przygotowaniu miejsca pracy, ograniczeniu długości pracy oraz dokonywaniu systematycznych przerw.

Niezbędna zwłaszcza jest ochrona kobiet ciężarnych, u których stwierdzono częstsze poronienia i uszkodzenia płodu. Dlatego lekarze apelują o ostrożną pracę przy komputerze, aby ochronić swoje zdrowie fizyczne i psychiczne.

10.11. Na czym polega wizja społeczeństwa informacyjnego?

Postęp techniki, telekomunikacji, informatyki oraz multimedii przyczynił się do powstania wizji społeczeństwa informacyjnego.

Termin ten określa społeczeństwo, w którym informacja jest traktowana jako dobro niematerialne i staje się ważniejsza od dóbr materialnych. W taki sposób informacja staje się towarem.

W wyniku działań podjętych w Unii Europejskiej, opracowano szereg projektów rozwiązań w wybranych dziedzinach, które powinny charakteryzować społeczeństwo informacyjne [32]. Są to m.in.:

- praca i nauczanie na odległość (ang. *homeworking, teleworking, distance learning*)
- zdalna opieka medyczna
- handel elektroniczny (ang. *electronic commerce, e-commerce*)
- bank wirtualny, internetowy (ang. *home banking*)
- organizacja działalności gospodarczej i administracji państwowej.

Praca na odległość jest nowym zjawiskiem i polega na wykonywaniu pracy zawodowej bez opuszczania mieszkania. Podejście takie powoduje zmniejszenie kosztów związanych z dojazdem do pracy i pozwala zaoszczędzić czas tracony na dojazd. Aby

było to możliwe, należy mieć dostęp do infrastruktury telekomunikacyjnej i oczywiście komputer. Taki styl pracy nie będzie możliwy oczywiście na wszystkich stanowiskach. Zjawisko to jest spotykane już obecnie. Przewagą nauczania na odległość nad tradycyjną formą nauki, będzie podwyższenie poziomu nauczania przy jednoczesnym obniżeniu kosztów edukacji.

Zdalna opieka medyczna przyczyni się również do zmniejszenia kosztów, a także umożliwi odbywanie specjalistycznych konsultacji bez potrzeby przemieszczania pacjentów. Będzie to szczególnie istotne poza największymi aglomeracjami, gdzie dostęp do wysokiej klasy specjalistów jest utrudniony.

Handel elektroniczny spowoduje obniżenie kosztów transportu, umożliwi klientom dokonywanie zakupów bez potrzeby opuszczania mieszkania. Klienci, dzięki dostępowi do wielu sklepów jednocześnie, będą mieli większą możliwość wyboru towaru.

Organizacja działalności gospodarczej, usprawni przepływ informacji między odległymi placówkami firm (obniżenie kosztów). Zmiana w organizacji administracji państwowej polega np. na udostępnieniu aktów prawnych w sieci. Ważnym elementem wspomagającym działanie odległych oddziałów firm jest obecnie już stosowana elektroniczna wymiana korespondencji.

10.12. Jakie cechy mają komputery piątej generacji?

Komputery kolejnej generacji cechują się przede wszystkim sztuczną inteligencją. Komputery te są zdolne wykonywać do miliarda decyzji na sekundę.

Komputery tej generacji będą działały przy użyciu sieci neuronowych. Przewiduje się, że komputery te będą miały inną budowę. Zamiast struktur półprzewodnikowych będą stosowane struktury biologiczne i przełączanie światła.

11. Spis literatury

- [1] Amoroso E.: *Wykrywanie intruzów*, Wydawnictwo ReadMe, 1999;
- [2] Aspinwall J.: *IRQ, DMA i I/O*, Wydawnictwo ReadMe, 1999;
- [3] Ball B.: *Linux*, Wydawnictwo Helion, 1998;
- [4] Banachowski L., Diks K., Rytter W.: *Algorytmy i struktury danych*, Wydawnictwo WNT, Warszawa, 1999;
- [5] Björck A., Dahlquist G.: *Metody numeryczne*, PWN, 1987;
- [6] Buchanan W.: *Sieci komputerowe*, Wydawnictwa Komunikacji i Łączności, 1997;
- [7] Chalk B.S.: *Organizacja i architektura komputerów*, Wydawnictwa Naukowo-Techniczne, Warszawa, 1998;
- [8] Dec Z., Konieczny R.: *ABC komputera 99*, Wydawnictwo Edition, 2000;
- [9] Dembowski K.: *Warsztat komputerowy*, Wydawnictwo Robomatic; 2000;
- [10] Derfler F.: *Poznaj sieci*, Wydawnictwo Mikom, Warszawa, 1999;
- [11] Drozdek A.: *Wprowadzenie do kompresji danych*, Wydawnictwa Naukowo-Techniczne, Warszawa, 1999;
- [12] Dryja M., Jankowscy J. i M.: *Przegląd metod i algorytmów numerycznych, cz.2*, WNT, 1988;
- [13] Duch W.: *Fascynujący świat komputerów*, Wydawnictwo Nakom, Poznań, 1997;
- [14] Duch W.: *Fascynujący świat programów komputerowych*, Wydawnictwo Nakom, Poznań, 1997;
- [15] Frisch A.: *UNIX Administracja systemu – drugie wydanie*, Wydawnictwo ReadMe, 1997;
- [16] Garfinkel S., Spafford G.: *Bezpieczeństwo w UNIX-ie i Internecie*, Wydawnictwo ReadMe, 1999;
- [17] Gutowski Z., Molski M.: *Komputer w pytaniach i odpowiedziach*, Wydawnictwa Naukowo-Techniczne, Warszawa, 1988;
- [18] Jankowscy J. i M.: *Przegląd metod i algorytmów numerycznych, cz.1*, WNT, 1988;

- [19] Jarczewski P., Jeliński A.: *Koncepcja edukacyjnych fundamentów informatyki*, Praca magisterska pod kierunkiem dra Mariana Molskiego, ATR, Bydgoszcz 2000;
- [20] Kolan Z.: *Urządzenia techniki komputerowej*, Wydawnictwo CWK Screen, 1999;
- [21] Kutylowski M.: *Kryptografia. Teoria i praktyka zabezpieczania systemów komputerowych*, Wydawnictwo ReadMe, 1999;
- [22] Metzger P., Jełowicki A.: *Anatomia PC*, Wydawnictwo Helion, 1999;
- [23] Mielczarek W.: *Szeregowe interfejsy cyfrowe*, Wydawnictwo Helion; 1994;
- [24] Miller M.A.: *TCP/IP – wykrywanie i usuwanie problemów*, Wydawnictwo ReadMe, 1999;
- [25] Molski M.: *Wstęp do techniki cyfrowej*, WKŁ, Warszawa 1989;
- [26] Molski M.: *Podstawy bezpieczeństwa systemów informatycznych*, MSG Media, Bydgoszcz 1998;
- [27] Mroczek J., Kosowski K.: *Alchemia komputerów*, Wydawnictwo Grom;
- [28] Nunemacher G.: *Przewodnik po sieciach lokalnych (LAN)*, Wydawnictwo Mikom, Warszawa, 1996;
- [29] Perry G.: *Poznaj komputer PC w 24 godziny*, Wydawnictwo Intersoftland, 1999;
- [30] Perry G.: *Poznaj podstawy programowania w 24 godziny*, Wydawnictwo Intersoftland, 1999;
- [31] Piechówka M., Szejko S.: *CASE, jaki jest, każdy widzi*, Miesięcznik Informatyka, kwiecień 1998,
http://www.pol.pl./informatyka/1998_04/f6.htm
- [32] Piotrowski A.J.: *W dużym skrócie o Społeczeństwie Informacyjnym*,
<http://www.kgn.gov.pl/gsi/artukul.html>;
- [33] Rosch W.L.: *Multimedia od A do Z*, Wydawnictwo Intersoftland, 1997;
- [34] Schmidt F.: *SCSI. Protokoły, zastosowania i programowanie*, Wydawnictwo Mikom;
- [35] Siołek O., Bibok O.: *Przy komputerze bez stresu, czyli jak obłaskawić maszynę*, Edition, 2000;
- [36] Skarbek W.: *Multimedia – Sprzęt i oprogramowanie*, Akademicka Oficyna Wydawnicza PLJ, Warszawa, 1999;

-
- [37] Sławik M., Bremer A.: *Elementy informatyki dla szkół średnich*, Wydawnictwo Videograf II, Katowice 2000;
 - [38] Sportack M.: *Sieci komputerowe – księga eksperta*, Wydawnictwo Helion, 1999;
 - [39] Sysło M.: *Algorytmy*, Wydawnictwa Szkolne i Pedagogiczne, Warszawa, 1997;
 - [40] Tanenbaum A.S.: *Rozproszone systemy operacyjne*, PWN, 1997;
 - [41] Tokarski J.: *Słownik wyrazów obcych*, PWN Warszawa 1977;
 - [42] Woodcock J.: *Komputer krok po kroku*, Wydawnictwo ReadMe, 1999;
 - [43] Wróblewski P.: *ABC Komputera*, Wydawnictwo Helion, 1999;
 - [44] *Encyklopedia powszechna PWN*, PWN Warszawa;
 - [45] *Wademecum teleinformatyka*, IDG Poland, 1998.

12. Skorowidz

A

abacysta · 143
ABC · 14
adres IP · 130
adres pamięci · 28
adresowanie
 bezpośrednie · 35
 natychmiastowe · 35
 pośrednie · 35
 względem licznika rozkazów · 35
 względem rejestru bazowego · 35
ADSL · 176
AGP · 102
Aiken · 13
akcelerator graficzny · 48
akumulator · 34
Al Chopezmi · 143
algebra Boole'a · 24
algorytm · 15; 143
algorytmista · 143
ALU · 26
Apple · 52
architektura systemu komputerowego · 103
archiwizacja
 częściowa · 73
 pełna · 73
ARPA · 129
ARPANET · 129
artificial intelligence · 176

arytmometr · 34

ASCII · 77

Atanasoff · 14

B

backup · 72

bajt · 22

bezprzerwowy system zasilania · 31

BIOS · 79

bit · 22

BPP · 80

brama w sieci · 126

bramka logiczna · 24

bridge · 126

C

C++ · 151

CA · 142

cache · 27

cache L1 · 63

cache L2 · 63

CAD · 175

CAM · 175

CASE · 174

CD-R · 57; 58

CD-RW · 58

Centronix · 78

Chat · 139
 chipset · 100
 CISC · 51
 CMOS · 79
 control flow computer · 15
 cpi · 85
 CRT · 97
 CSK · 77
 Cyformat · 77
 czas dostępu · 28
 częstotliwość
 odświeżania · 84
 taktowania procesora · 49
 czytnik kodu kreskowego · 75

D

DAT · 72
 data flow computer · 15; 110
 DDP · 122
 Delphi · 151
 demodulacja · 92
 digitizer · 75; 88
 Dijkstra · 151
 DMA · 81
 DNS · 133; 136
 domena · 130
 dpi · 87
 DRAM · 56
 driver · 102
 drukarka · 75
 atramentowa · 85; 86
 igłowa · 85; 86

 laserowa · 85; 86
 termiczna · 85; 87
 duplex · 91; 115
 DVD · 71
 dyrektywa · 148
 dysk
 elastyczny · 67
 twardy · 67

E

Eckert · 14
 e-commerce · 181
 ECP · 80
 EDI · 127
 EDVAC · 14
 EEPROM · 57
 EIDE · 100
 elektroniczna wymiana dokumentów · 127
 e-mail · 138
 ENIAC · 14
 EPP · 80
 EPROM · 57
 Ethernet · 121

F

Fast Ethernet · 121
 FIFO · 80; 166
 flaga · 39
 Flash · 57
 FM · 66
 format rozkazu · 36

FTP · 135

G

generacja komputerów · 18

GEORGE · 162

giga · 28

Gigabit Ethernet · 121

głośnik · 75

H

half duplex · 91

handel elektroniczny · 142

handshake · 80

handy skaner · 87

hardware · 103

home banking · 181

homeworking · 181

HP-UX · 110

I

IBM (CP852) · 77

IBM PC · 17

ICAN · 130

IDE · 100

indeksowanie · 35

instrukcja · 146; 147

integer · 47

Intel · 50

inteligentny budynek · 124

interfejs · 79

graficzny · 170

Internet · 129

interpreter · 170; 172

postscriptu · 85

IPv6 · 134

IPX · 122

IRC · 131; 139

Irix · 110

ISA · 99

ISO-8859-2 · 77

J

Java · 151

Java script · 151

język

assemblerowy · 146

programowania · 145

wewnętrzny · 146

wysokiego poziomu · 146

K

kalkulator · 143

kamera · 75

cyfrowa · 88

kanał IRC · 139

karta dźwiękowa · 89

karta rozszerzeń · 99

kategoria okablowania · 123

kilo · 28

klawiatura · 75

klient – serwer · 127

kod · 23

BCD · 44
 kreskowy · 98
 operacji · 36

kodowanie · 22
 cyfrowe · 77

kolejka procesów · 166

komentarz w programie · 149

kompilator · 170

kompresja · 74
 LZW · 95

komputer · 13
 osobisty · 17
 piątej generacji · 182
 przemysłowy · 17
 przenośny · 17
 sterowany danymi · 110
 wirtualny · 18

koncentrator · 117; 126

koprocesor · 47
 arytmetyczny · 47
 graficzny · 47

L

LAN · 116

land · 70

laptop · 104

LCD · 97

licznik rozkazów · 34

Linux · 110

lista rozkazów · 15
 procesora · 36

login · 137

logowanie · 107

lpi · 85

LSI · 16

L

łącze
 dedykowane · 131
 dzierżawione · 131
 stałe · 131

Łukasiewicz J. · 172

M

Magic · 151

magistrala · 26
 adresowa · 29
 danych · 29
 sterująca · 29

makroinstrukcja · 148

MAN · 116

Mark I · 13

Mauchly · 14

Mazovia · 77

mega · 28

metoda zstępująca · 151

metody numeryczne · 155

MFM · 66

mikrofon · 75

mikroprocesor · 50

MINI-DIN · 96

mirroring · 73

model
 hierarchiczny · 163
 warstwowy · 163
 warstwowy ISO-OSI · 113

modem · 75; 92
 GSM · 141

modulacja · 92

amplitudy AM · 93
częstotliwości FM · 93

monitor · 162

graficzny · 83
LCD · 97

MOS · 56

most · 126

Motorola · 52

MSI · 16

multimedia · 176

myszka · 75; 96

N

nagrywarka · 69

NAT · 134

NetBEUI · 122

netykieta · 131

nick · 139

notebook · 17; 104

O

obiekt · 153

obsługa przerwania · 42

OCR · 90; 179

odwrotna notacja polska · 172

okablowanie strukturalne · 124

ONP · 172

oprogramowanie · 160

podstawowe · 160
użytkowe · 160

OS/360 · 162

P

pakiet · 113

palmtop · 17; 104

pamięć · 55

magnetoptyczna · 69
podręczna · 27
podręczna drugiego poziomu · 49
podręczna pierwszego poziomu · 49
ulotna · 55

panel LCD · 75

Pascal · 151

password · 137

PCI · 99

PCMCIA · 99

Perl · 151

pętla · 38

ping · 135

pit · 70

plik · 168

ploter · 75; 87

płyta główna · 101

PN-93 T-42118 · 77

poczta elektroniczna · 138

podprogram · 39

pojemność pamięci · 28

pole

argumentów · 149
operacji · 148

polecenia AT · 92

połączenie

komutowane · 131
wideokonferencyjne · 140

port AGP · 102

- postscript · 85
- potokowy · 37
- PowerPC · 52
- pozycyjny system dwójkowy · 22
- proces · 163
- procesor · 26
- program · 145
 - nadzorujący · 162
 - zarządzający · 162
 - źródłowy · 145
- programowanie
 - obiektywne · 153
 - strukturalne · 151
- projektor · 75
- PROM · 58
- protokół
 - HTTP · 137
 - ICMP · 132
 - IP · 122; 132
 - sieciowy · 120
 - TCP · 132
 - TCP/IP · 122
 - UDP · 132
- przełącznik, switch · 126
- przeplot w monitorze · 84
- przerwanie · 42; 82
 - programowe · 82
 - sprzętowe · 82
- przetwornik
 - analogowo/cyfrowy · 75
 - cyfrowo/analogowy · 75
- PS2 · 78; 96
- ramdysk · 63
- ramka · 113
- real · 47
- RealAudio · 140
- real-time · 107
- RealVideo · 140
- rejestr
 - maski · 83
 - rozkazów · 34
- rekord
 - fizyczny · 168
 - logiczny · 167
- repeater · 126
- RevDNS · 136
- RGB · 83
- RISC · 51
- RLL · 67
- ROM · 55
- router · 126
- routing, rutowanie, trasowanie · 115
- rozdzielczość
 - druku · 85
 - monitora · 84
- rozkaz · 146; 147
 - skoku · 37
 - warunkowy · 38
- rozpoznawanie mowy · 178
- RS-232 · 78; 81

R

- RAID · 73
- RAM · 55

S

- SCO Unix · 110
- SCSI · 100
- segment sieci · 116

- sektor · 67
- semantyka · 145
- serwer · 17; 127
- katalogowy · 140
 - usług internetowych · 127
- serwis news · 138
- sieć
- komputerowa · 113
 - lokalna LAN · 113
 - miejska MAN, metropolitalna · 113
 - neuronowa · 179
 - rozległa WAN · 113
- simplex · 91
- skalarny · 37
- skalowalność · 105
- skaner · 75; 87
- ręczny · 87
- skok warunkowy · 38
- słowo komputera · 27
- software · 103
- Solaris · 110
- Sparc · 52
- społeczeństwo informacyjne · 181
- spooler · 170
- SPP · 80
- sprzęg · 79
- sprzężenie
- decyzyjne · 95
 - informacyjne · 95
- SRAM · 56
- ssh · 136
- SSI · 16
- SSL · 142
- stacja robocza · 17
- stałopozycyjny zapis · 45
- standard V · 94
- sterownik sprzętowy · 100
- stopa błędów · 91
- stos · 41
- streamer · 71
- Sun Microsystems · 52
- superkomputer · 17
- switch · 126
- sygnał dwustanowy · 21
- syntaktyka · 145
- SyQuest · 68
- system
- czasu rzeczywistego · 107
 - ekspercki · 177
 - komputerowy · 103
 - operacyjny · 161
 - ósemkowy · 23
 - plików · 165; 168
 - szesnastkowy · 23
 - wejścia/wyjścia · 75
 - wieloprocesorowy · 108
- sztuczna inteligencja · 176
- szybkość
- druku · 85
 - modulacji · 91
 - taktowania procesora · 27
 - transmisji · 91
-
- S*
- ścieżka · 67
- światowa pajęczyna informacyjna · 136
-
- T*
- technika multimedialne · 176
- tekstowy interfejs · 170

telewizja na żądanie · 176
 teleworking · 181
 telnet · 136
 terminal emulowany · 138
 THE · 162
 topologia drzewa · 119

- gwiazda-magistrala · 120
- gwiazdy · 117
- liniowa · 117
- lokalnych sieci komputerowych · 117
- pierścień-gwiazda · 120
- pierścienia · 118

 touchpad · 75; 97
 touchscreen · 75; 97
 trackball · 75; 96
 translacja adresów · 134
 translator · 145; 170
 transmisja

- asynchroniczna · 76; 91
- jednoczesna · 91; 115
- jednokierunkowa · 91
- naprzenienna · 91
- równoległa · 76
- synchroniczna · 76; 91
- szeregowa · 76
- z potwierdzeniem · 80

U

układ scalony · 16
 Unikod · 78
 UNIVAC · 14
 UNIX · 107; 110
 UPS · 31
 urządzenie

- logiczne · 169
- wejścia/wyjścia · 26

 USB · 78; 99

uzupełnienie dwójkowe · 44

V

V.32 · 94
 V.32bis · 94
 V.34 · 94
 V.42 · 94
 V.42bis · 94
 V.90 · 94
 Ventura · 77
 VLSI · 16
 von Neumann · 14

W

WAN · 116
 WAP · 141
 warstwa

- aplikacji · 115
- fizyczna · 114
- łącz danych · 114
- prezentacji · 115
- sesji · 115
- sieciowa · 115
- transportowa · 115

 wielodostępny · 107
 wieloprogramowość · 105
 wielowątkowość · 105
 Winchester · 67
 wirtualne urządzenie wejścia/wyjścia · 169
 wskaźnik stosu · 41
 współbieżność · 108; 165
 współdzielenie · 163
 WWW · 135

wzmacniacz, repeater · 126

X

X-windows · 170

Z

zakleszczenie · 164

zarządzanie

 czasem procesora · 165
 operacjami wejścia/wyjścia · 165

 pamięcią · 165

 zasilacz impulsowy · 31

 zasób · 116; 163

 zegar · 27

 zewnętrzna częstotliwość taktowania · 49

 zgodność · 19

 ZIP · 68

 zmiennopozycyjny zapis · 47

 znacznik stanu · 38

 znak-moduł · 44