

WYKŁAD 1

ROZWÓJ SYSTEMÓW OPERACYJNYCH

Próba definicji

System operacyjny jest programem, który działa jako pośrednik między użytkownikiem a sprzętem komputerowym. Jego (systemu) zadaniem jest tworzenie środowiska, w którym użytkownik może wykonywać programy.

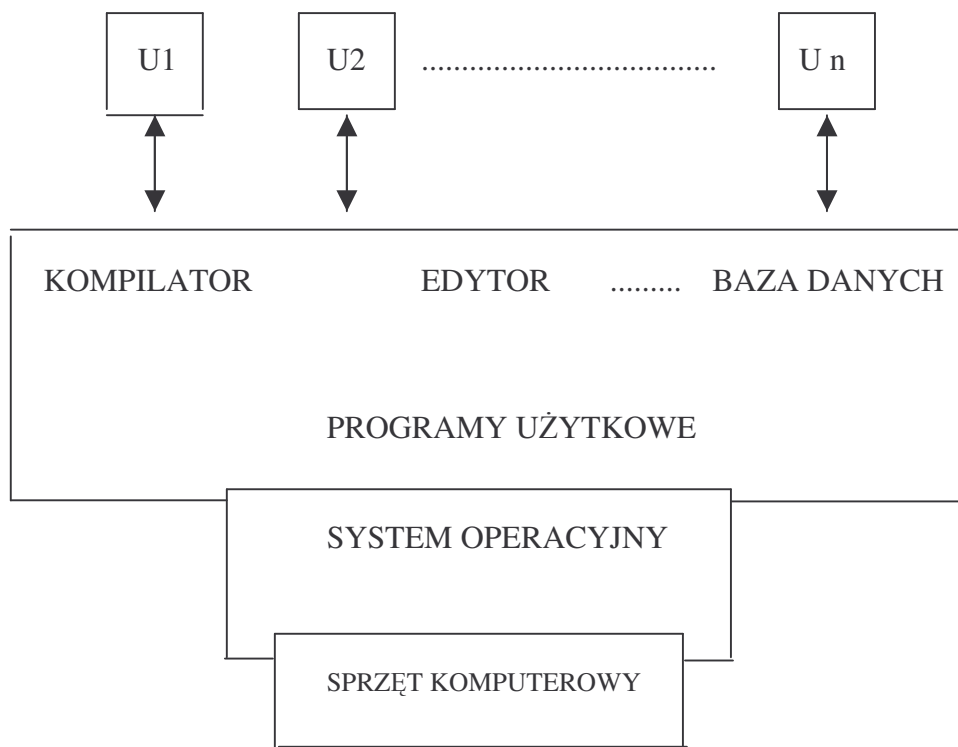
Dwa podstawowe cele systemu operacyjnego to:

- spowodowanie, aby system komputerowy był wygodny w użyciu
- spowodowanie wydajnej pracy (eksploatacji) sprzętu komputerowego

A. Co to jest system operacyjny?

System operacyjny jest częścią prawie każdego systemu komputerowego. System komputerowy składa się z czterech podstawowych części:

- sprzęt
- system operacyjny
- programy użytkowe
- użytkownicy



Rys. 1

System operacyjny nadzoruje i koordynuje posługiwanie się sprzętem przez różne programy działające na zlecenie różnych użytkowników. Możemy też uważać system operacyjny za dystrybutora zasobów takich jak: czas procesora, obszar w pamięci operacyjnej, obszary w systemie plików, urządzenia wejścia_wyjścia itd.

Ponieważ często nie dochodzi do konfliktowych zamówień na zasoby, system operacyjny musi decydować o przydziale zasobów poszczególnym zamawiającym, mając na względzie wydajne i harmonijne działanie całego systemu komputerowego.

Z drugiej strony

System operacyjny jest programem sterującym rozmaitymi urządzeniami wejścia_wyjścia i programami użytkownika. Nadzoruje działanie programów użytkownika, przeciwdziała błędom i zapobiega niewłaściwemu użyciu komputera.

Nie istnieje uniwersalnie akceptowalna definicja tego, co jest, a co nie jest częścią systemu operacyjnego. W uproszczeniu można przyjąć, że system operacyjny to ten program, który działa w komputerze nieustannie (nazywany zazwyczaj JADREM), podczas gdy wszystkie inne są programami użytkowymi.

Dwa podstawowe cele systemu operacyjnego (wygoda, wydajność) stają często sprzeczności. W przeszłości(ze względu na wysoką cenę dużych systemów komputerowych) wydajność była często przedkładana nad wygodę.

W ciągu ostatnich 50 lat systemy operacyjne i architektura systemów komputerowych wywierały na siebie wzajemnie znaczny (a i stymulujący) wpływ. Aby zapewnić wygodę – rozwijano systemy operacyjne. Wraz z rozwojem systemów operacyjnych stwierdzono, że zmiany w sprzęcie mogłyby ułatwić działanie systemom operacyjnym.

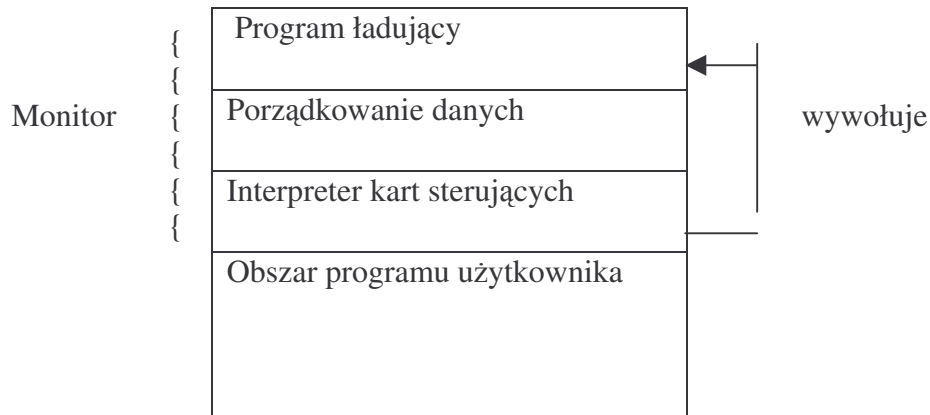
Krótką historia systemów operacyjnych

B. Wczesne systemy operacyjne

- Tryb pracy interakcyjnej (Zapisy na czas komputera – np. 30' ; wykorzystanie procesora w przybliżeniu 5 – 7 %; przepustowość w przybliżeniu 2 – 3 prace / h
- Tryb pracy wsadowej (Zawodowi operatorzy grupujący podobne prace (kompilator, program1, program2, ... , program n); nadal były przerwy związane ze zmianami wsadów; wykorzystanie procesora $\leq 55\%$; przepustowość w przybliżeniu 33 prac / h)

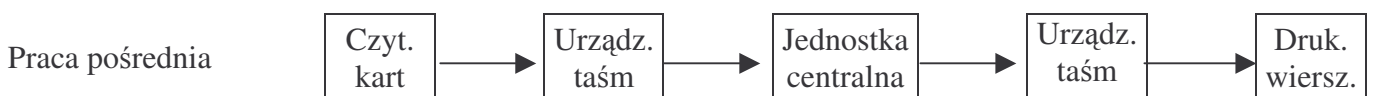
Aby zlikwidować nieuniknione przestoje związane ze zmianą wsadów zastosowano metodę automatycznego porządkowania zadań, która posłużyła jako podstawa do zbudowania pierwszych najprostszych systemów operacyjnych tzw. monitorów (rezydujących).

Można w nim wyróżnić kilka części. Jedną z ważniejszych jest interpreter kart sterujących. Wywołuje on program ładujący, aby załadować do pamięci programy systemowe bądź użytkowe. Program ładujący też jest częścią monitora.



- Praca pośrednia – wprowadzona w trosce o zwiększenie wydajności i wyeliminowanie postojów w pracy procesora (związanych z małą szybkością pracy urządzeń we-wy). Czytnik kart i drukarki wierszowe zastąpiono taśmami magnetycznymi. Wykorzystanie procesora 90 %, przepustowość 55 prac / h

Praca bezpośrednia



Buforowanie i Spooling

- Buforowanie

- zapewnia możliwość jednoczesnego wykonywania obliczeń i operacji we_wy
- w praktyce rzadko utrzymuje jednostkę centralną i urządzenia we_wy w stanie nieustannej aktywności.

Zasada: Gdy średnie prędkości procesora i urządzeń we_wy są podobne – buforowanie przynosi korzyści. Gdy średnie prędkości różnią się znacznie – korzyść będzie niewielka albo żadna!

Są dwa typy zadań:

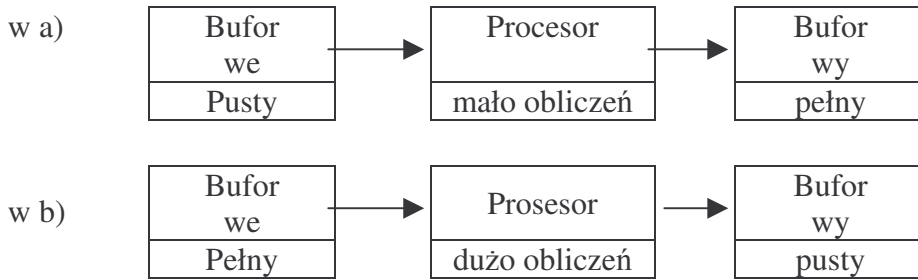
- uzależnione od we_wy
- uzależnione od procesora.

a) – mało obliczeń – dużo operacji we_wy

b) – dużo obliczeń – mało operacji we_wy

W zadaniach typu: a) prędkość obliczeń jest uzależniona od we_wy

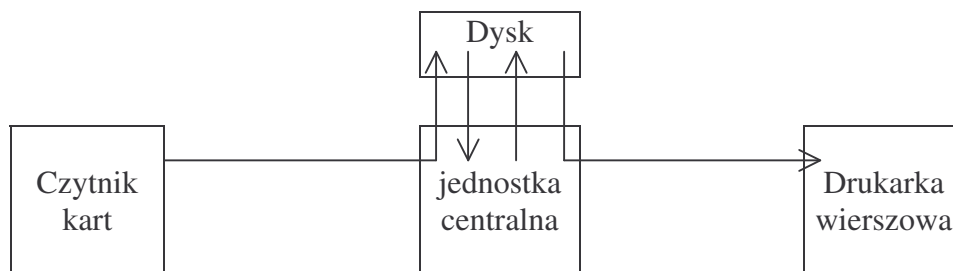
b) prędkość obliczeń jest uzależniona od procesora



w a) – procesor czeka na dane we

w b) – procesor nie nadąża z przetwarzaniem i bufor we się zapełnia!

- **Spooling** (SPOOL – Simultaneous Peripheral Operation On – Line)



W istocie jest to buforowanie z bardzo dużym buforem w postaci dysku.

- Buforowanie umożliwia jednoczesne obliczanie i realizację we_wy dla tego samego zadania
- Spooling umożliwia równoczesne wykonywanie operacji we_wy jednego zadania i obliczeń dla innych zadań.

Ważną cechą spoolingu jest wytwarzanie istotnej struktury danych – puli zadań.

Spooling powoduje, że pewna liczba zadań jest zawczasu przeczytana i czeka na dysku na realizację. Ze względu na swobodny (niesekwencyjny) dostęp do dysku, pula zadań pozwala systemowi operacyjnemu wybierać następne zadania do wykonania tak, aby zwiększyć wykorzystanie jednostki centralnej.

Czyli przy wykorzystaniu dysku możliwe jest planowanie zadań. (Należy pamiętać, że przy zadaniach nadchodzących z kart lub taśmy nie ma możliwości zmiany ich porządku (FIFO)).

Wybrane aspekty planowania zadań:

a) wieloprogramowość – SO wybiera i rozpoczyna wykonywanie zadania Z1. Prędzej czy później Z1 musi zaczekać na coś. Wówczas SO zaczyna realizować Z2. Dopóki są jakieś zadania do wykonania to procesor nie będzie bezczynny. Takie systemy są jednak dość skomplikowane.

Problemy:

- aby zadanie było gotowe do pracy, to musi być w pamięci. SO musi mieć w pamięci wiele zadań;
- jeśli jest kilka gotowych zadań, to system musi wybrać takie, które dostanie czas procesora (planowanie przydziału czasowego);
- wykorzystanie współbieżności wielu zadań wymaga, by ich potencjalne niekorzystne oddziaływania znajdowały się pod stałą kontrolą SO w zakresie: planowania zadań, planów procesora, magazynowania danych na dysku i zarządzania pamięcią;

b) podział czasu (wielozadaniowość): to już jest skomplikowane, bo wykonują się tu na przemian wiele zadań, przy czym przełączenia następują tak często, że użytkownicy mogą współdziałać z każdym programem podczas jego wykonywania. SO z podziałem czasu sprawia, że wielu użytkowników dzieli równocześnie jeden komputer

c) systemy rozproszone

- ściśle powiązane (rozdzielenie obliczeń na wiele procesów; komputery wieloprocessorowe; współdzielenie pamięci i zegara);
- luźno powiązane (każdy procesor ma własną pamięć lokalną)

Przyczyny budowania systemów rozproszonych:

- podział zasobów (np. wspólne korzystanie z drukarki);
- przyspieszenie obliczeń (założenie - jeśli obliczenia da się podzielić na zbiór niezależnych obliczeń cząstkowych);
- niezawodność;
- łączność (poczta, zdalne dyski - stacje lokalne)

IV Systemy czasu rzeczywistego

- sterowanie procesami technologicznymi (np. system wtrysku paliwa)
- czas reakcji systemu.

Przetwarzanie musi się zakończyć przed upływem określonego czasu.

V Systemy operacyjne jedno stanowiskowe – komputery osobiste

Jeszcze niedawno – procesory takich komputerów pozbawione były cech niezbędnych do ochrony systemu oper. w tym podwójnego trybu wykorzystywania rozkazów. SO tych komputerów nie były zatem ani wielozadaniowe ani niedostępne.

Postawiono na maksymalną wygodę użytkownika i właściwego czasu reakcji i po pojawieniu się procesów z trybem dwuzadaniowym (tryb jądra i tryb użytkownika) pojawiły się też ????? SO z wielozadaniowością i wielowarstwowością (OS/2, W 95, 98, NT).