

# 1. Pliki i ich organizacja

## (1.1) Pojęcie pliku

Dane bezpośrednio potrzebne procesorowi do wykonywania jego zadań są umieszczane w pamięci operacyjnej systemu. Jest to jednak pamięć ulotna i dane w niej zapisane są tracone po wyłączeniu zasilania. Nie nadaje się więc ona do trwałego magazynowania informacji. Dodatkowo rozmiar pamięci operacyjnej jest stosunkowo niewielki. Dlatego też informacje magazynuje się w sposób trwały na różnych nośnikach takich jak magnetyczne dyski i taśmy lub dyski optyczne.

Plik jest pojęciem logicznym i definiuje podstawową logiczną jednostkę magazynowania informacji. Można też powiedzieć, że plik stanowi zbiór informacji stanowiący pewną logiczną całość i zapisany na trwałym nośniku informacji.

Żadna informacja nie może być zapisana w pamięci pomocniczej inaczej niż w pliku.

## (1.2) Rodzaje i typy plików

Dane zapisane w pliku zależą od jego twórcy i mogą być bardzo różne. Ze względu na charakter zapisanych informacji pliki dzieli się na:

- programy oraz
- dane.

Dane zapisane w plikach mogą być różnego rodzaju. Mogą to być dane liczbowe, teksty, teksty z dodatkowymi informacjami o formatowaniu, obrazy grafiki komputerowej, nagrania dźwiękowe, zapisy sekwencji wideo itd.

Struktura pliku, czyli sposób jego wewnętrznej organizacji, zależą od rodzaju zapisanych informacji. Mówimy, że struktura pliku zależy od jego typu.

Przykłady typy plików zostały zebrane w tabeli *Typy plików*.

<i>Typ pliku</i>	<i>Rozszerzenie nazwy</i>	<i>Funkcja</i>
<b>Wykonywalny</b>	exe, com, bin	Gotowy do wykonania program w języku maszynowym
<b>Kod źródłowy</b>	C, pas, asm, vb, cpp	Kod źródłowy wyrażony w różnych językach
<b>Wsadowy</b>	bat, sh	Polecenia dla interpretera poleceń
<b>Tekstowy</b>	txt, doc	Dane i dokumenty tekstowe
<b>Obraz</b>	gif, jpg	Plik binarny w formacie przeznaczonym do oglądania
<b>Archiwalny</b>	zip, arj, rar	Plik skompresowany odpowiednim algorytmem kompresji

## Tabela 1. Typy plików

Typ pliku jest rozpoznawalny po rozszerzeniu w nazwie pliku.

Wewnętrzna organizacja struktury pliku jest uzależniona od typu pliku i od wymagań programu z niego korzystającego.

Z plikiem związanych jest kilka *atrybutów* (cech). Należą do nich m.in.:

- **Nazwa** – ciąg znaków identyfikujących plik w sposób zrozumiały dla człowieka;
- **Typ** – określa „rodzaj” przechowywanych informacji, na przykład: wykonywalny, wynikowy, kod źródłowy, tekstowy, archiwalny itd. Typ rozpoznawany jest najczęściej przez tak zwane rozszerzenie;
- **Położenie** – pozwala na fizyczna lokalizację na nośniku
- **Rozmiar** – zawiera wielkość pliku najczęściej wyrażoną w bajtach
- **Ochrona** – informacje o tym kto i w jaki sposób może sprawować kontrolę i odczytywać informację z pliku
- **Data** – data utworzenia, modyfikacji, ostatniego użycia, wydrukowania itp.

W różnych systemach mogą występować różne zestawy atrybutów.

### (1.3) Operacje na plikach

System operacyjny zapewnia mechanizmy dostępu i zarządzania plikami.

System operacyjny powinien zapewnić następujące operacje na pliku:

- Tworzenie
- Zapisywanie
- Czytanie
- Usuwanie
- Skracanie
- Zmiana pozycji

W różnych systemach operacyjnych mogą występować dodatkowe operacje na plikach.

Należy podkreślić, że plik jest strukturą logiczną, tak samo jak jego struktura wewnętrzna. Zapis pliku na dysku musi natomiast być zgodny z fizyczną organizacją nośnika danych. Z tego powodu

zapis pliku na dysku odbywa się całymi blokami, które często nie pokrywają się z logiczną wielkością pliku. W takim wypadku część przestrzeni dyskowej pozostaje niewykorzystana, a zjawisko to znane jest pod nazwą fragmentacji.

[NASTĘPNA](#)

## 2. System plików

### (2.1) Pojęcie systemu plików

*System plików* to mechanizm pozwalający na bezpośredni dostęp oraz przechowywanie informacji (dane i programy) w pamięci pomocniczej.

System plików jest jednym z bardziej widocznych elementów systemów operacyjnego. W jego skład wchodzi zbiór plików oraz struktura katalogów, która pozwala na łatwą organizację i udostępnianie informacji o wszystkich plikach zapisanych w systemie. W niektórych systemach występuje jeszcze dodatkowy element służący do wyodrębnienia w sposób fizyczny lub logiczny dużych zbiorów katalogów.

Głównym zadaniem systemu pliku w systemie komputerowym jest zapewnienie prawidłowego i wydajnego dostępu do plików zapisanych w pamięci pomocniczej oraz zapewnienie prawidłowej ochrony przechowywanej informacji.

### (2.2) Partycja

*Partycja* (ang. *partition*) zwana też czasem *wolumenem* (ang. *volumes*) lub *dyskiem logicznym* stanowi logiczny podział dysku na oddzielne strefy. Podziału na partycje dokonuje się w celu:

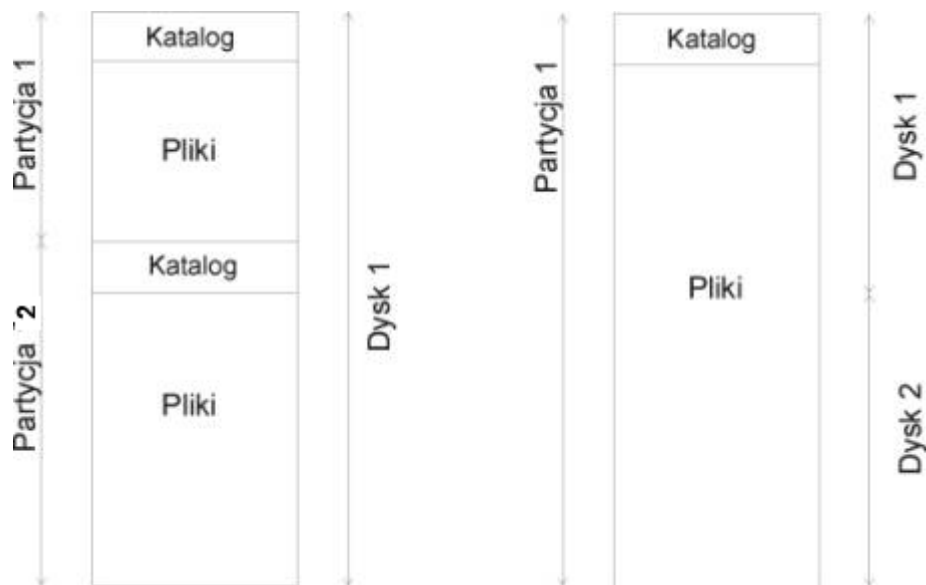
- uzyskania kilku rozłącznych obszarów na jednym dysku fizycznym

lub

- zwiększenia rozmiaru strefy ponad rozmiar pojedynczego dysku fizycznego (w wypadku partycji montowanych na kilku dyskach fizycznych)

Każda partycja zawiera zapisane na niej pliki oraz pełną informację o tych plikach przechowywaną strukturze zwanej w katalogiem (ang. *device directory*).

W katalogu zapisane są najczęściej takie informacje o pliku jak: nazwa, położenie, rozmiar, typ.



Rysunek 1. Organizacja systemu plików

Organizacja systemu plików przy różnych podziałach partycji między dyski fizyczne została pokazana na rysunku

### (2.3) Struktura katalogów

Katalog zawiera bardzo ważne informacje, gdyż określa położenie pliku na dysku. Operacje wykonywane na plikach takie jak odnajdywanie, odczyt czy usuwanie są wykonywane przy wykorzystaniu katalogu. Dlatego też organizacja struktury katalogu jest dla systemu bardzo istotna.

Najprostszym sposobem organizacji katalogu jest struktura jednopoziomowa. W takiej strukturze informacje o wszystkich plikach ujęte są w jednym poziomie. Katalog o strukturze jednopoziomowej jest łatwo utworzyć i obsługiwać. Ma on jednak sporo wad. Uwypuklają się one przede wszystkim w momencie, gdy rośnie liczba plików lub różnych użytkowników systemu. W takiej strukturze każdy plik musi mieć swoją unikatową nazwę, a różni użytkownicy nie mogą w ten sam sposób nazywać swoich plików. Trudności pojawiają się również w trakcie wyszukiwania informacji o zadanym pliku.

Tych wad pozbawiona jest, najczęściej dziś wykorzystywana, struktura katalogów zorganizowana na kształt drzewa.

W takiej strukturze nie występuje problem jednoznaczności nazw (w ujęciu globalnym) oraz problem oddzielenia plików różnych użytkowników, gdyż każdy z nich może przechowywać swoje pliki w osobnej części struktury drzewiastej.

W katalogu drzewiastym drzewo ma swój katalog główny (ang. *root*), a każdy plik w systemie oprócz swojej nazwy ma jednoznaczną ścieżkę określającą jego położenie w drzewie katalogów.

[NASTĘPNA](#)

## 3. Rodzaje systemów plików

### (3.1) FAT16 i FAT32

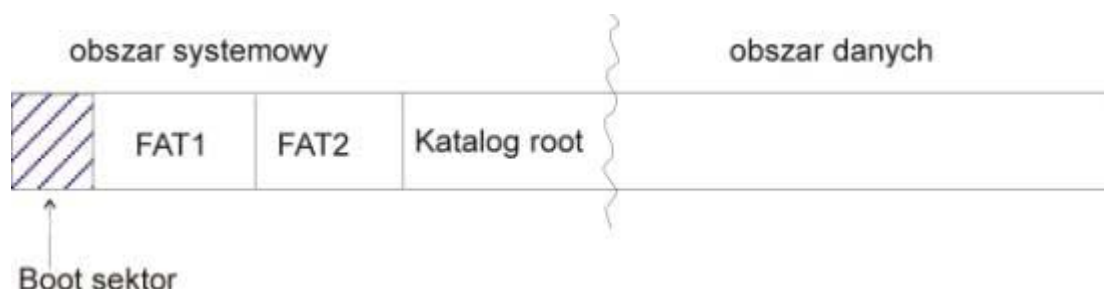
System plików FAT jest pierwszym systemem plików powszechnie zastosowanym w komputerach osobistych. Początkowo do adresowania jednostki alokacji (najmniejszej jednostki organizacji logicznej dysku) projektanci tego systemu przewidzieli zaledwie 12 bitów. Spowodowało to ograniczenie maksymalnego rozmiaru partycji dysku obsługiwanej przez system operacyjny korzystający z FAT zaledwie do 32MB. W późniejszych wersjach systemu DOS poprawiono ten mankament rozszerzając maksymalną wielkość obsługiwanych partycji do 2GB.

#### Budowa FAT

System plików FAT zbudowany jest z

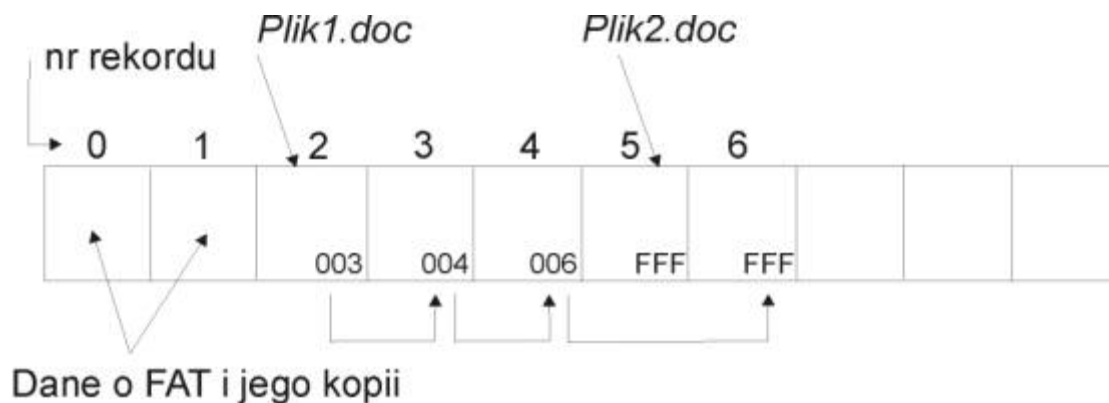
- tablicy alokacji plików;
- boot sektora;
- tablicy partycji;
- katalogu głównego.

**Tablice alokacji plików** (ang. *File Allocation Table*)



Rysunek 2. Struktura systemu FAT

- zawiera informacje o fizycznym położeniu każdego pliku zapisanego na dysku. Pierwsze dwa rekordy tej tablicy, to dane na temat samego FAT-u, a wszystkie pozostałe podają informacje o położeniu plików. Każdy rekord FAT-u opisuje położenie jednego klastra z danymi. W wypadku, gdy plik nie mieści się w całości w jednym klastrze, to w tablicy FAT zapisywana jest informacja o położeniu następnego klastra z danymi tego pliku.



Rysunek 3. Tablica alokacji plików

### Boot sektor

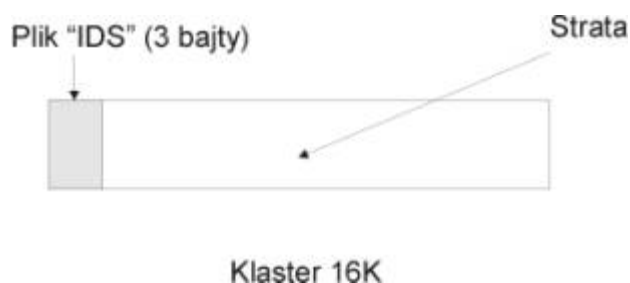
- jest pierwszym sektorem dysku i jest wykorzystywany podczas inicjalizacji ładowania systemu operacyjnego

### Tablica partycji

- przechowuje informacje o partycjach zainstalowanych w systemie

### Katalog główny

- przechowuje dane o zbiorach znajdujących się w poszczególnych katalogach. Każde pole opisujące plik ma 32 bajty i zapisuje się w nim informacje o nazwie pliku, jego atrybutach, dacie modyfikacji, rozmiar i adres pierwszego klastra. Wielkość katalogu głównego jest najczęściej ograniczona. Na przykład dla dyskietki 3,5" sformatowanej w systemie FAT na katalog główny przeznaczone jest 14 sektorów.



Rysunek 4. Fragmentacja dysku

Każdy plik zajmuje na dysku całkowitą liczbę klastrów bez względu na jego wielkość. Przy małych plikach (kilka bajtów) następuje więc znaczna strata miejsca na dysku. Wielkość klastra w systemie FAT16 jest ustalana automatycznie w czasie instalacji systemu i zależy od wielkości partycji, na



której jest zakładany. Zależność ta została pokazana w tabeli. Należy zaznaczyć, że komputery z systemem FAT16 pracujące pod kontrolą systemu operacyjnego DOS nie obsługują partycji większych od 2GB.

<i>Wielkość partycji</i>	<i>Liczba sektorów na klaster</i>	<i>Wielkość klastra</i>
0 MB–32 MB	1	512 bajtów
33 MB–64 MB	2	1 KB
65 MB–128 MB	4	2 KB
129 MB–255 MB	8	4 KB
256 MB–511 MB	16	8 KB
512 MB–1,023 GB	32	16 KB
1,024 GB–2,047 GB	64	32 KB
2,048 GB–4,095 GB	128	64 KB

Tabela 2. Wielkość klastra w systemie FAT16

## FAT32

Zasadnicza różnica w stosunku do systemu FAT polegała na tym, że rozszerza on wielkość partycji, którą można wykorzystać, z 2 GB do prawie 127 GB. Dzieje się tak dzięki zastosowaniu do przechowywania rekordów i numerów sektorów bloków o wielkości 32 bitów.

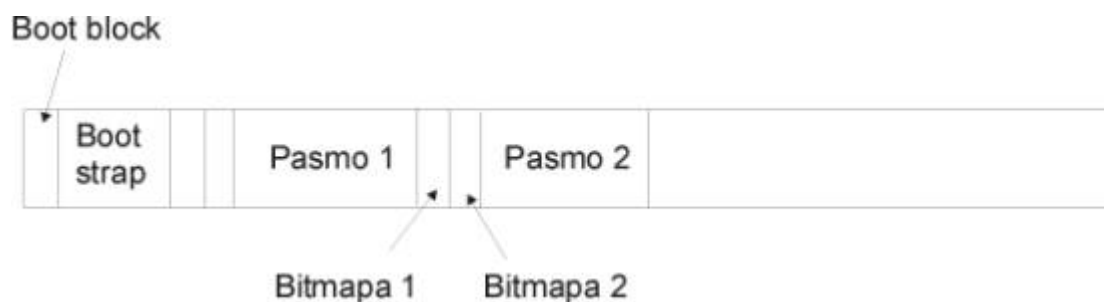
Wielkość klastra w systemie FAT32 pokazano w tabeli

<i>Wielkość partycji</i>	<i>Rozmiar klastra</i>
Poniżej 8 GB	4 KB
Od 8 GB do 16 GB	8 KB
Od 16 GB do 32 GB	16 KB
od 32 GB	32 KB

Tabela 3. Wielkość klastra w systemie FAT32

Dodatkowo system plików FAT32 pozwala na stosowanie długich nazw plików oraz wykorzystanie w ich nazwach znaków regionalnych (w tym polskich znaków diakrytycznych) dzięki zastosowaniu standardu Unicode.

## (3.2) HPFS



Rysunek 5. Struktura systemu HTFS

System HPFS (ang. *High Performance File System*) stosowany jest w systemie operacyjnym OS/2. System HPFS zapisuje na dysku boot block, super block i spare block. Struktury te są wykorzystywane do startu systemu oraz zarządzania systemem plików. Ponadto system rezerwuje obszar tzw. Bitmapy (2KB), które są rozmieszczane w odstępach co 8MB. Każda bitmapa zawiera jeden bit dla każdej jednostki alokacji w paśmie. System operacyjny zapisuje odpowiednie bitmapy i zaznacza, które sektory są wolne, a które zajęte. HPFS planuje rozmieszczenie plików na dysku w ten sposób, aby jego ewentualne powiększenie nie powodowało defragmentacji dysku. Dodatkowo nazwy plików i katalogów są przechowywane w postaci struktury b-tree. Struktura b-tree jest strukturą, której organizacja zbliżona jest do struktury drzewa. W wypadku organizacji danych w strukturę drzewiastą nie trzeba przeszukiwać całej zawartości magazynu informacji, a tylko odpowiednią ścieżkę drzewa. Tego typu rozwiązanie zwiększa znakomicie wydajność wyszukiwania plików na dysku.

### (3.3) NTFS

System NTFS został zaprojektowany w ten sposób, aby umożliwiać obsługę bardzo dużych dysków, zapewnić wysoki poziom bezpieczeństwa i dużą szybkość dostępu do danych.

NTFS wykorzystuje 64 bitów do numerowania klastrów. Wielkość klastra w systemie NTFS nie jest stała i może być definiowana przez użytkownika.

<i>Wielkość partycji</i>	<i>Liczba sektorów na klaster</i>	<i>Wielkość klastra</i>
Do 512 MB	1	512 bajtów
513 MB–1024 MB	2	1 KB
1025 MB–2048 MB	4	2 KB
2049 MB–4096 MB	8	4 KB
4097 MB–8192 MB	16	8 KB
8193 MB–16 384 MB	32	16 KB
16 385 MB–32 768 MB	64	32 KB
>32 768 MB	128	64 KB

Tabela 4. Wielkość klastra w systemie NTFS

System ten jako standard kodowania znaków stosuje standard Unicode. Pozwala to na używanie w nazwach plików narodowych znaków diakrytycznych, których długość może mieć 255 znaków (w tym spacje i kropki).

Po sformatowaniu nośnika w systemie NTFS tworzone są specjalne pliki zwane plikami metadaty, a wśród nich Master File Table. MFT zawiera rekordy o wielkości 1KB, w których umieszczana jest informacja o rozmieszczeniu na dysku poszczególnych plików. W zależności od rozmiaru zbioru jest mu przyporządkowana różna liczba rekordów. Pierwszych 17 rekordów MFT ma specjalne znaczenie dla systemu, gdyż w nich zapisane są informacje bezpośrednio związane z samym systemem plików. Wszystkie pozostałe zawierają już informacje o plikach i katalogach.

System NTFS 5.0 wykorzystany w Windows 2000 posiada kilka interesujących cech. Do najważniejszych należy niewątpliwie możliwość zabezpieczania pojedynczego pliku lub katalogu dzięki zapisywaniu informacji o uprawnieniach dostępu w specjalnej bazie zwanej ACL.

Do nowego systemu plików dodano możliwość określania rozmiaru katalogu, wolumenu lub dysku wykorzystywanego przez danego użytkownika lub grupę.

Innym udoskonaleniem jest mechanizm tzw. Sparse File, polegający na alokowaniu takiej ilości miejsca na dysku, jaka jest rzeczywiście potrzebna konkretnemu plikowi. Jeżeli aplikacja żąda od systemu np. 5 MB miejsca na dysku celem zapisania pliku, to system sprawdzi, czy faktycznie taka ilość miejsca jest niezbędna i jeśli okaże się to możliwe zostanie przydzielony mniejszy obszar (tylko taki, jaki jest potrzebny do przechowania pliku).

NTFS 5.0 posiada także mechanizm śledzący i zapamiętujący zmiany w plikach nie tylko podczas pracy komputera. Po jego restarcie zmiany na dyskach są nadal pamiętane. Ma to ułatwić odzyskiwanie przypadkowo utraconych danych.

Nowy NTFS może również montować do dowolnego pustego katalogu inny katalog znajdujący się na dysku lokalnym lub zdalnym. (Mechanizm ten znany jest od dawna m.in. w systemach unixowych.)

Partycje NTFS 5.0 (inne nie) mogą również być obsługiwane przez mechanizm EFS (Encryption File System) pozwalający na szyfrowanie danych na partycji i dostęp do zaszyfrowanych danych na podstawie identyfikatora elektronicznego. Przy tym dane (oraz ich ewentualne kopie zapasowe i pliki pomocnicze) pozostają zakodowane podczas całego czasu pracy z szyfrowanymi zbiorami.

[NASTĘPNA](#)

## Podsumowanie

Bez pamięci masowej trudno byłoby sobie wyobrazić pracę systemu komputerowego. Dane w pamięci masowej przechowywane są w postaci plików. Za kontrolę i obsługę plików jest odpowiedzialny system plików. Od niego zależy sposób i jakość wykorzystania nośnika pamięci oraz komfort i sposób pracy użytkownika. Niektóre systemy operacyjne mogą współpracować tylko z jednym systemem plików, inne potrafią wykorzystać wiele z nich. W najpopularniejszych systemach rodziny Windows stosowanych w komputerach klasy PC stosowane są przede wszystkim systemy FAT i NTFS. System NTFS ma znacznie większe możliwości i pozwala m.in. na zastosowanie odpowiednich mechanizmów zabezpieczeń systemu na poziomie pojedynczego pliku i katalogu.

Obok wymienionych systemów plików często można spotkać również systemy CDFS i UDF przeznaczone do zapisu i odczytu danych z nośników CD i DVD.