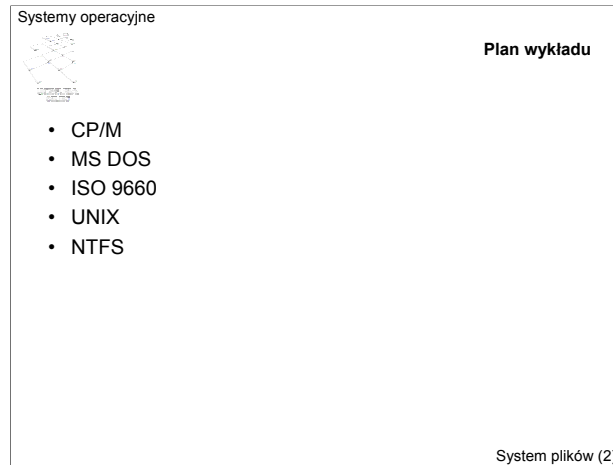


Systemy operacyjne System plików — przykłady implementacji Wykład prowadzą: Jerzy Brzeziński Dariusz Wawrzyniak Nazwa modułu
--

Celem wykładu jest zaprezentowanie wybranych implementacji systemu plików, omówienie struktur najważniejszych metadanych dla tych systemów, oraz przedyskutowanie pewnych konsekwencji przyjętych założeń projektowych.



Geneza systemu plików wiąże się z jakimś systemem operacyjnym, ale współcześnie niektóre systemy operacyjne udostępniają wiele różnych systemów plików. Przykładem może być system Linux, w którym na różnych partycjach mogą współistnieć takie systemy plików, jak ext2, ext3, ReiserFS, itd. Istnieje też możliwość dostępu do systemów plików, opartych na strukturze FAT, zastosowanej w DOS oraz Windows 95/98, NTFS, czy też sieciowych systemów plików, np. NFS. Możliwość współistnienia wielu systemów plików możliwa jest dzięki dodatkowej warstwie w implementacji obsługi systemu plików, zwanej *wirtualnym systemem plików* (VFS).

W module zaprezentowane zostaną systemy plików CP/M, MS DOS (oparty na strukturze FAT), ISO 9660 (podstawowy systemów plików dla CD ROM), typowy system uniksowy, oraz NTFS (stosowany w nowszych rozwiązaniach firmy Microsoft).

Systemy operacyjne



Przykłady implementacji systemu plików ⁽¹⁾

- CP/M — katalog zawiera blok kontrolny pliku (FCB), identyfikujący 16 jednostek alokacji (zawierający indeksy tych jednostek alokacji).
- DOS — wpis katalogowy zawiera indeks pierwszej jednostki alokacji, a pozostałe jednostki wynikają z tablicy FAT.
- ISO 9660 (CD ROM) — bloki zorganizowane są wg. zasady przydziału ciągłego, wpis katalogowy zawiera indeks pierwszej jednostki alokacji oraz rozmiar pliku, wpisy katalogowe są posortowane alfabetycznie.

System plików (3)

Systemy operacyjne



Przykłady implementacji systemu plików (2)

- UNIX — plik opisany jest przez i-węzeł, wpis katalogowy zawiera indeks i-węzła, który z kolei zawiera indeks (kombinowany) jednostek alokacji.
- NTFS — plik identyfikowany jest przez referencję, która jest indeksem rekordu w tablicy MFT, rekord zawiera atrybuty pliku (w szczególności dane) lub odnośniki do bloków z atrybutami.

System plików (4)

Systemy operacyjne



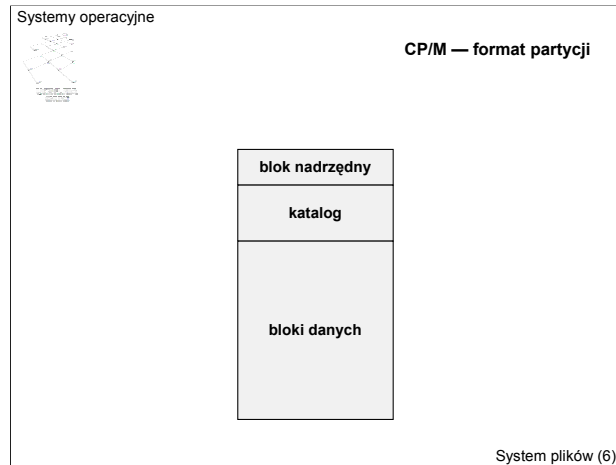
CP/M — informacje ogólne

- Katalog zawiera 32-bajtowe wpisy, tzw. bloki kontrolne plików (FCB), zawierające wartości atrybutów.
- W katalogu znajdują się wpisy wszystkich plików w systemie.
- Dane (zawartość pliku) znajdują się w 128-bajtowych sektorach, alokacji podlegają jednak bloki o rozmiarach 1KB.
- Z każdym wpisem związany jest jednobajtowy kod użytkownika.
- Zarządzanie wolną przestrzenią przy użyciu wektora bitowego, przechowywanego w pamięci.

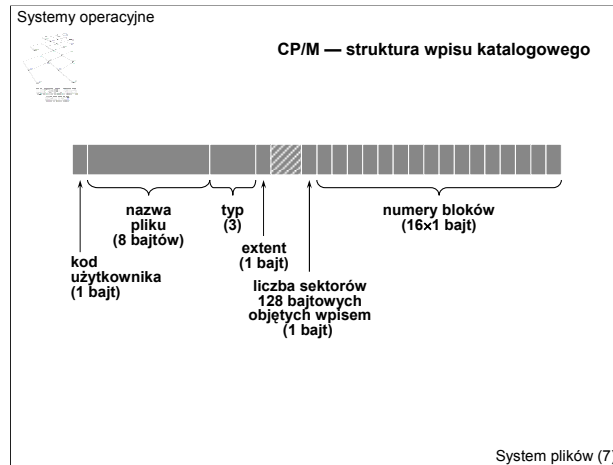
System plików (5)

Plik w systemie CP/M opisany jest przez blok kontrolny (FCB). Bloki FCB wszystkich plików znajdują się w katalogu, przy czym jest tylko jeden katalog dla danej partycji. W systemie CP/M nie ma więc hierarchicznej struktury katalogów. Implementacyjnie struktura ta jest jednopoziomowa, a logicznie dwupoziomowa, gdyż atrybutem każdego pliku jest kod użytkownika. Każdy użytkownik widzi więc tylko swoje pliki. Rozróżnienie plików poprzez kod użytkownika umożliwia zatem nadawanie im tych samych nazw.

System plików CP/M projektowany był na potrzeby obsługi dyskietek o stosunkowo niewielkich rozmiarach (180 KB). Dane (zawartość pliku) znajdują się w 128-bajtowych sektorach, jednostką alokacji jest jednak blok o rozmiarze 1KB (8 sektorów). Do zarządzania wolną przestrzenią używany jest wektor bitowy, przechowywany w pamięci, zajmujący 23 bajty, wyliczany każdorazowo po restarcie lub zmianie nośnika.



Katalog oraz obszar danych mają w strukturze partycji ustalony rozmiar. Oznacza to, że maksymalna liczba plików oraz ich łączny rozmiar definiowane są na etapie formatowania partycji.



Jak wynika ze struktury wpisu katalogowego, nazwa pliku składa się z 8 znaków, oraz 3 znaków rozszerzenia, kojarzonych z typem pliku. Jeden taki wpis identyfikuje 16 jednostek alokacji, czyli maksymalnie 16 KB. Rzeczywisty rozmiar obszaru danych, objętych wpisem, podany jest z dokładnością do 128 bajtów (liczba sektorów). Jeśli rozmiar pliku nie jest wielokrotnością 128 bajtów, wymagany jest jakiś specjalny bajt w zawartości pliku, sygnalizujący jego koniec.

Koncepcje przydziału kilku pozycji do opisu jednego pliku wykorzystano również na potrzeby przechowywania długich nazw plików w systemach MS Windows 95/98 oraz na potrzeby pomieszczenia dużej liczby atrybutów w systemie plików NTFS.

Systemy operacyjne

MS DOS — informacje ogólne

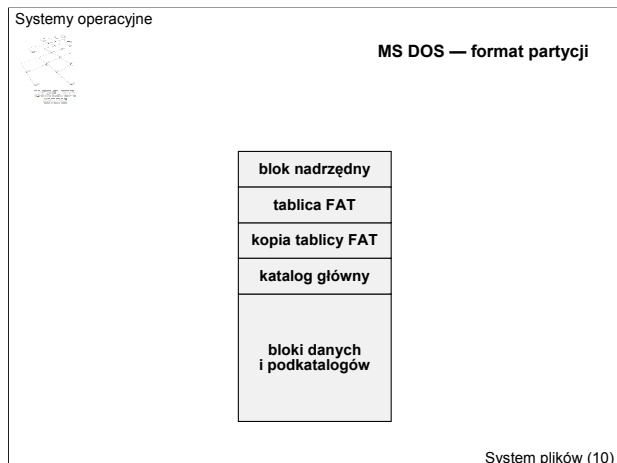
- Katalog zawiera 32-bajtowe wpisy, obejmujące podstawowe atrybuty pliku (nazwa, typ, rozmiar itp.)
- Katalogi tworzą strukturę wielopoziomową
- Dane znajdują się w blokach (jednostkach alokacji) o ustalonym rozmiarze.
- Bloki identyfikowane są z pośrednictwem tablicy FAT, a indeks pierwszego bloku znajduje się we wpisie katalogowym.
- Wolne bloki identyfikowane są poprzez specjalny kod, wpisany na odpowiadających im pozycjach w tablicy FAT.

System plików (9)

Rozmiar wpisu katalogowego jest taki sam, jak bloku kontrolnego w systemie CP/M. Dzięki temu w systemie MS DOS istnieje alternatywna możliwość korzystania z bloków FCB.

System DOS dopuszcza wielopoziomą, hierarchiczną strukturę katalogów. Katalog implementowany jak każdy plik, ale zawiera w treści 32-bajtowe opisy innych plików lub katalogów. Wpis dla katalogu w nadkatalogu zawiera flagę, informującą, że jest to katalog, w związku z czym jego zawartość ma być traktowana i interpretowana w szczególny sposób.

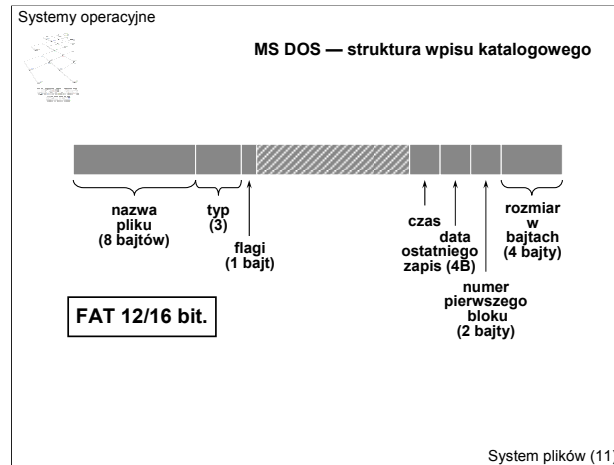
Wielkość jednostki alokacji można dobrać na etapie konfiguracji (formatowania logicznego). Wielkość ta wahała się od 512 B do kilku kilobajtów w miarę zwiększania się wielkości dysków. Bloki plików, a także bloki wolne lub zawierające uszkodzone sektory, identyfikowane są poprzez odpowiednie wartości na odpowiadających im pozycjach w tablicy FAT.



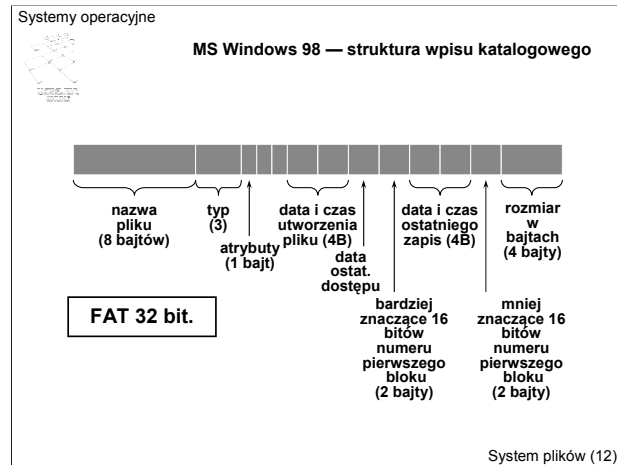
Na etapie formatowania logicznego definiowany jest podział partycji, z którego wynika między innymi rozmiar tablicy FAT oraz katalogu głównego. Rozmiar tablicy FAT określa liczbę jednostek alokacji, pośrednio zatem również rozmiar jednostki alokacji, gdyż wielkość obszaru danych też jest ustalona przy formatowaniu. Oczywiście liczba jednostek alokacji nie może być większa, niż wynika to z rozmiaru elementu tablicy FAT. Element tablicy FAT musi po prostu „pomieścić” numer dowolnego bloku. Przykładowo, w przypadku 16-bitowej struktury FAT do dyspozycji jest $2^{16}=65536$ numerów jednostek (w praktyce kilka numerów mniej, ze względu na kody specjalne). Przeznaczanie zatem więcej niż 128 KB ($65536 \cdot 16$ bitów) na tę strukturę nie ma sensu, bo większej liczby jednostek i tak nie da się obsłużyć.

Ze struktury partycji wynika też, że ustalony jest rozmiar katalogu głównego. Liczba wpisów w katalogu głównym ma zatem górne ograniczenia, a w przypadku nie wykorzystania tych wpisów miejsce jest marnowane. Jeśli chodzi o rozmiar podkatalogu, jest on ustalany dynamicznie z dokładnością do jednostki alokacji w miarę zwiększania lub zmniejszania liczby wpisów.

Tablica FAT jest newralgicznym elementem systemu plików. Jej uszkodzenie może spowodować całkowitą utratę powiązań bloków i tym samym praktycznie zniszczyć system plików. Dlatego utrzymywana jest również kopia tej tablicy. Wątpliwości może budzić lokalizacja takiej kopii. Jeśli tablice sąsiadują ze sobą, to w przypadku zarysowania dysku przez głowicę istnieje ryzyko uszkodzenia obu tablic.



Podobnie, jak w systemie CP/M, nazwa składa się maksymalnie z 8 znaków, oraz 3 znaków rozszerzenia. Flagi są bitami, określającymi specjalne atrybuty pliku: plik ukryty, plik systemowy, plik tylko do odczytu, podkatalog, plik archiwalny, etykieta. W katalogu przechowany jest numer pierwszego bloku, czyli początek listy w tablicy FAT. Przeznaczone jest na to maksymalnie 16 bitów, a więc taka struktura jest właściwa dla rozwiązań opartych najwyżej na 16-bitowej tablicy FAT.



FAT 32-bitowy wymaga pewnych rozszerzeń w strukturze wpisu katalogowego. Wykorzystane zostały wszystkie bajty, nieużywane w wersji opartej na FAT 12/16. Indeks pierwszego bloku podzielony jest na 2 części po 16 bitów w każdej. Dlatego nadaje się do zastosowania w rozwiązaniach opartych na FAT 32.

Systemy operacyjne



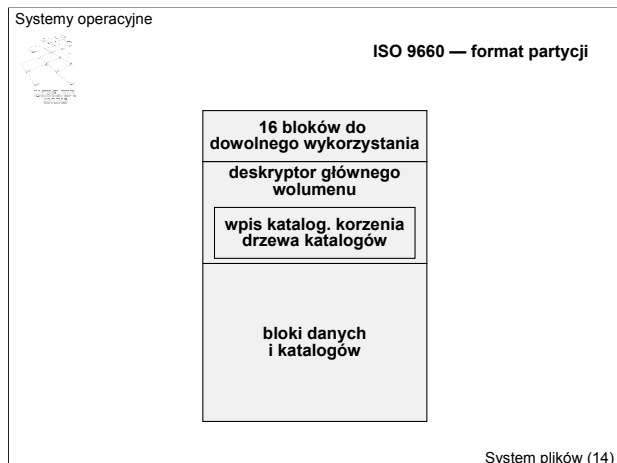
ISO 9660 — informacje ogólne

- Dane na płycie CD nie są rozkładane w cylindrach, lecz tworzą układ spiralny.
- Spirala dzielona jest na bloki logiczne (sektory) o rozmiarze 2352 bajtów (w tym preambuły, kody korekcyjne itp.), z czego na dane pozostaje 2048 bajtów.
- Binarne pola numeryczne są reprezentowane podwójnie: zgodnie z uporządkowaniem grubokońcówkowym oraz cienkokońcówkowym.
- Maksymalny poziom zagnieżdż. katalogów — 8.

System plików (13)

ISO 9660 jest standardem systemu plików dla płyt CD (również DVD), wykorzystującym wszystkie udogodnienia, które wynikają z faktu niezmienności zapisanych danych, takie jak sortowanie katalogów, czy przydział ciągły. Jego powszechnie znanymi rozszerzeniami są standardy *Rock Ridge* oraz *Joliet*.

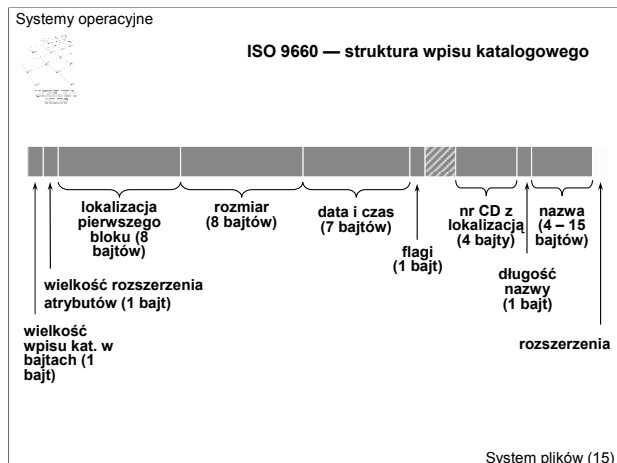
Ciekawostką jest podwójne (grubo- i cienkokońcówkowe) reprezentowanie wszystkich liczb całkowitych — chodziło o to, aby nie dyskryminować jednego czy drugiego uporządkowania, ponieważ mogłoby to utrudnić zaakceptowanie standardu.



Standard nie definiuje przeznaczenia pierwszych 16 bloków. Obszar ten jest do dyspozycji twórcy, może np. służyć do umieszczenia programu rozruchowego w przypadku ładowania systemu operacyjnego z płyty.

Uwzględniając różnorodność zastosowań płyt kompaktowych, w deskrytorze głównego wolumenu znajduje się mnóstwo informacji o jej zawartości, autorze, wytwórcy, prawach autorskich itd. Można tam umieścić dowolne informacje, przydatne raczej tylko dla użytkownika, zapisane z użyciem dużych liter, cyfr i wybranych znaków interpunkcyjnych.

Z zakresu metadanych w deskrytorze znajduje się rozmiar bloku logicznego (2048, ale dopuszcza się więcej, np. 4096, 8192 itd.), liczba bloków na płycie, czas utworzenia i przeterminowania zawartości płyty. Jedną z najważniejszych informacji jest opis katalogu głównego, tzn. blok początkowy i rozmiar, zgodnie z zasadą przydziału ciągłego.



Wpis katalogowy może mieć zmienną długość, co przejawia się w obecności trzech pól: *wielkości wpisu katalogowego*, *wielkości rozszerzenia* oraz *długości nazwy*. Nazwy dłuższe niż 15 znaków należą już do rozszerzeń standardu. W standardzie ISO 9660 przewidziano taką możliwość, umieszczając właśnie pole *wielkości rozszerzenia* w katalogu. Struktura rozszerzenia jest już przedmiotem definicji standardu rozszerzającego.

Atrybuty *lokalizacja* i *rozmiar* są wartościami 32 bitowymi, ale ze względu na podwójną reprezentację zajmują 8 bajtów. Podobnie *nr CD* jest wartością 16-bitową, a zajmuje 4 bajty. Z wartościami całkowitymi 1-bajtowymi nie ma tego problemu, gdyż nie ma tam czego porządkować.

W polu flagi wyróżniono między innymi następujące bity:

- bit, który odróżnia pliki od katalogów,
- bit ukrycia wpisu na listingu zawartości,
- bit atrybutów rozszerzonych,
- bit oznaczający ostatni wpis w katalogu.

Standard ISO uwzględnia operowanie zbiorami płyt, możliwe jest więc, że wpis katalogowy opisuje plik, zlokalizowany na innej płycie. Stąd pole reprezentujące numer płyty CD z lokalizacją pliku.

Systemy operacyjne

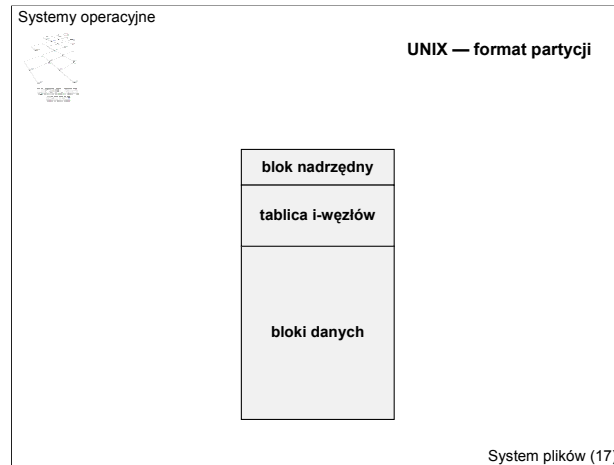


UNIX — informacje ogólne

- Z każdym plikiem związany jest i-węzeł, który przechowuje wszystkie atrybuty pliku z wyjątkiem nazwy.
- Nazwa znajduje się w katalogu obok numeru i-węzła danego pliku.
- Katalogi tworzą strukturę wielopoziomową (katalog zawiera wpis specyfikujący inny katalog).
- Dane (zawartość pliku) znajdują się w blokach (jednostkach alokacji) o ustalonym rozmiarze.
- Bloki danych identyfikowane są za pośrednictwem indeksu kombinowanego.
- Wolne bloki zorganizowane są zgodnie z zasadą grupowania.

System plików (16)

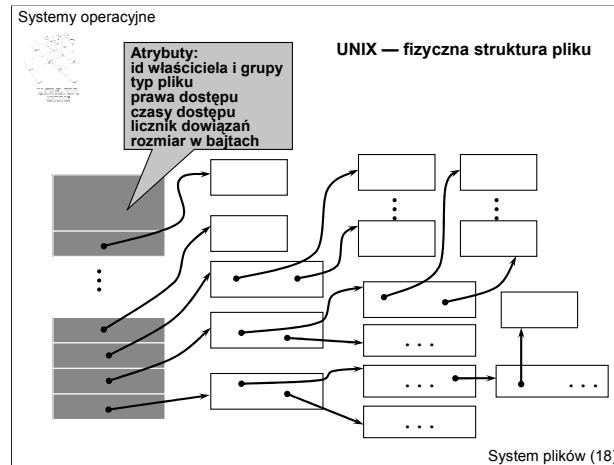
Opis implementacji uniksowego systemu plików ograniczony zostanie do najważniejszych koncepcji, na których opierają się różne odmiany tego systemu, stosowane w rodzinie systemów uniksowych. Podstawą implementacji jest i-węzeł, grupujący wszystkie atrybuty pliku z wyjątkiem nazwy. I-węzły tworzą tablicę, której rozmiar limituje liczbę plików w systemie. I-węzeł pośrednio lub bezpośrednio identyfikuje bloki danych pliku. Wolna przestrzeń jest identyfikowana przez grupowanie, przy czym pierwszy węzeł (blok) z indeksami wolnych bloków zlokalizowany jest w całości w bloku nadrzędnym. W pewnych odmianach stosowany jest wektor bitowy. Wektor bitowy w tych odmianach używany jest również do identyfikacji wolnych i-węzłów.



Dwa główne element partycji stanowią: tablica i-węzłów oraz bloki danych. W niektórych odmianach na jednej partycji wyodrębnia się kilka sekcji, z których każda ma strukturę zbliżoną do przedstawionej na slajdzie. Celem takiego rozdystrybuowania jest zmniejszenie strat w wyniku potencjalnych awarii oraz poprawa efektywności poprzez skrócenie czasu wyszukiwania dla głowic dyskowych.

Jak wynika ze struktury partycji, liczba i-węzłów jest ustalona, podobnie jak liczba bloków danych. Zawartość systemu plików podlega zatem takim samym ograniczeniom, jak w systemie CP/M, ale istnieje hierarchiczna struktura katalogów (drzewo katalogów, a uwzględniając pliki — graf acykliczny).

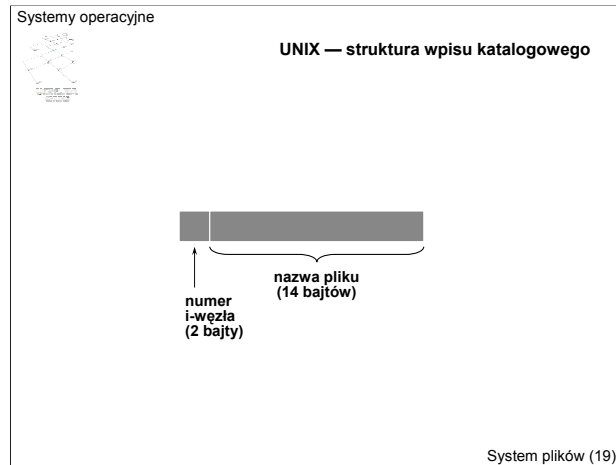
Ograniczenia na rozmiar katalogu wynikają wyłącznie z dostępności bloków danych, gdyż katalog zajmuje przestrzeń dyskową na takich samych zasadach, jak plik. Jego zawartość jest jednak interpretowana przez system zgodnie z zasadami budowy katalogów w danej implementacji. Jeden, ustalony i-węzeł opisuje korzeń drzewa katalogów.



Z każdym plikiem związany jest 1 i-węzeł. Zawiera on podstawowe atrybuty pliku w modelu uniksowym, czyli:

- identyfikator właściciela i grupy,
- typ pliku — plik zwykły, katalog, dowiązanie symboliczne, łącze nazwane, urządzenie znakowe, urządzenie blokowe, gniazdo,
- prawa dostępu — tradycyjne rwx dla właściciela, grupy i pozostałych,
- czasy dostępu — czas modyfikacji pliku, czas modyfikacji i-węzła, czas dostępu,
- licznik dowiązań — liczba różnych nazw, pod jakimi występuje plik w systemie,
- rozmiar w bajtach.

Pozostała część i-węzła wypełniona jest indeksami bloków z danymi. Część indeksów w i-węźle (10 – 12, zależnie od odmiany) wskazuje bezpośrednio bloki z danymi. Ponadto są jeszcze 3 indeksy, z których jeden wskazuje blok indeksowy 1-poziomowy, jeden — blok indeksu 2-poziomowego, a jeden — blok indeksu 3-poziomowego.



Katalog składa się z wpisów, kojarzących nazwę z i-węzłem. W tradycyjnym podejściu nazwa ograniczona była do 14 znaków, ale oczywiście w toku rozwoju systemów uniksowych wielkość tę zwiększano i współcześnie dopuszcza się nawet 256 znaków.

W katalogu istnieją też wpisy specjalne o nazwie `.` (kropka) i `..` (dwie kropki), skojarzone z numerem i-węzła odpowiednio katalogu bieżącego i nadrzędnego. Wyjątkiem w tym zakresie jest korzeń drzewa katalogów, który nie ma nadkatalogu.

Systemy operacyjne



NTFS — informacje ogólne

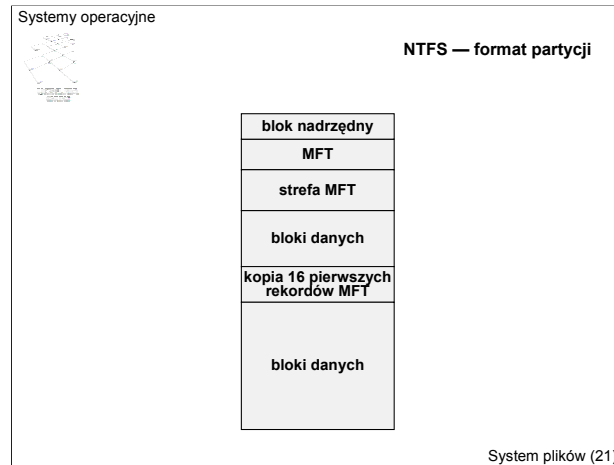
- Plik jest zbiorem atrybutów, obejmującym w szczególności atrybut dane.
- Każdy plik ma swój wpis (rekord) w głównej tablicy plików (MFT), a indeks w tej tablicy jest składnikiem identyfikatora wewnętrznego (referencji pliku).
- Tablica MFT sama jest plikiem, podobnie jak każdy inny obiekt w obrębie systemu plików.
- Wolne bloki identyfikowane są przez wektor bitowy, przechowywany w jednym z plików.

System plików (20)

W projekcie systemu NTFS uwzględniono różne cele projektowe, związane głównie z efektywnością i bezpieczeństwem, takie jak obsługa dużych wolumenów, tolerancja na awarie, kryptograficzne zabezpieczenie danych, czy kompresja danych.

Plik w modelu NTFS jest zbiorem atrybutów, w szczególności atrybutu dane. Wartość każdego atrybutu traktowana jest jako strumień bajtów. Potencjalnie w pliku może być kilka strumieni danych.

Podstawową strukturą w implementacji jest *główna tablica plików* (MFT – master file table)



Tablica MFT ma ustalone położenie na partycji NTFS, ale ponieważ sama jest plikiem, może być powiększana w ramach strefy MFT. Strefa MFT potencjalnie dostępna jest też dla bloków danych, ale wykorzystywana jest dopiero wówczas, gdy brakuje miejsca we właściwym obszarze danych. Dla bezpieczeństwa 16 rekordów MFT, identyfikujących ważne dla systemu pliki (zawierające metadane) ma swoją kopię w środkowej części partycji.

Przestrzeń dyskowa podzielona jest na bloki (zwane klastrami lub gronami), których rozmiar jest ustalony i jest wartością z zakresu 512 B – 64 KB, zależnie od wielkości wolumenu. Typowa wartość to 4 KB. Każdy blok identyfikowany jest przez pewien logiczny numer (LCN), związany z fizyczną lokalizacją bloku.

Systemy operacyjne

NTFS — struktura rekordu MFT

- Rekord MFT ma ustalony rozmiar 1KB – 4KB
- Rekord MFT składa się z nagłówka rekordu i tabeli atrybutów.
- Atrybut składa się z nagłówka atrybutu oraz wartości.
- Wartość znajduje się albo bezpośrednio w rekordzie MFT (atrybut rezydentny), albo w bloku danych.

nagłówek rekordu MFT	
nagłówek atrybutu	wartość atrybutu
nagłówek atrybutu	wartość atrybutu
⋮	

System plików (22)

Rozmiar rekordu MFT ustalany jest na etapie tworzenia systemu plików (formatowania logicznego) na ogół ma wielkość 1 KB. Rekord MFT poza nagłówkiem, składa się z atrybutów. Atrybut traktowany jest jako strumień bajtów, oczywiście odpowiednio interpretowany. Nagłówek atrybutu znajduje się zawsze w rekordzie MFT. Jeśli wartość atrybutu jest niewielka, może również zmieścić się w samym rekordzie. Jest to tzw. *atrybut rezydentny*. Jeśli wartość jest zbyt duża, umieszczana jest w obszarze danych, a w rekordzie MFT zamiast wartości znajduje się wskaźnik do tego obszaru. W przypadku niewielkiej objętości danych jest szansa umieszczenia całego pliku w rekordzie MFT.

Systemy operacyjne



NTFS — zawartość nagłówka rekordu

- Magiczna liczba — wartość używana do sprawdzania poprawności
- Numer sekwencyjny — wartość zwiększana o 1 za każdym razem, gdy rekord używany jest dla nowego pliku
- Liczba odniesień do pliku
- Liczba wykorzystywanych bajtów przez rekord
- Identyfikator rekordu bazowego (referencja) w przypadku rekordu rozszerzeń

System plików (23)

Numer sekwencyjny jest składnikiem referencji pliku i ma na celu eliminację pomyłek, wynikających z przechowania jakiejś starej referencji do już usuniętego pliku. Zwiększenie numeru sekwencyjnego o 1 oznacza, że stara referencja, nawet jeśli odnosi się do tego samego rekordu, będzie nieaktualna, gdyż będzie zawierać już nieobowiązujący numer sekwencyjny.

Pojęcia *rekord bazowy* i *rekord rozszerzeń* odnoszą się do przypadku, w którym jeden rekord MFT jest niewystarczający do przechowania wszystkich atrybutów i na potrzeby tego samego pliku przydzielany jest następny rekord.

Systemy operacyjne



NTFS — zawartość nagłówka atrybutu

- Identyfikator atrybutu (typ i opcjonalna nazwa)
- Długość wartości atrybutu w bajtach
- Lokalizacja wartości atrybutu
- Flagi

System plików (24)

Systemy operacyjne



NTFS — wybrane atrybuty pliku

- Informacje standardowe (\$STANDARD_INFORMATION)
- Nazwa pliku (\$FILE_NAME)
- Dane (\$DATA)
- Korzeń indeksu (\$INDEX_ROOT)
- Alokacja indeksu (\$INDEX_ALLOCATION)
- Mapa bitowa indeksu (\$BITMAP)
- Lista atrybutów (\$ATTRIBUTE_LIST)

System plików (25)

Nie wszystkie atrybuty są obecne w każdym rekordzie MFT. Niektóre z atrybutów są zawsze rezydentne, w przypadku innych uzależnione jest to od wielkości strumienia.

Atrybut *informacje standardowe* jest obligatoryjnym atrybutem każdego rekordu MFT, zawierającym informacje o właścicielu, flagi, czasy, licznik dowiązań itp. Ma on ustalony rozmiar i jest atrybutem rezydentnym.

Nazwa pliku zapisana jest kodzie Unicode (UTF-16). W rekordzie może wystąpić kilka atrybutów *nazwa pliku*, np. w przypadku twardych dowiązań lub dla zachowania zgodności nazw z systemem MS DOS. Atrybut ten jest zawsze rezydentny.

W pliku zwykłym istnieje anonimowy atrybut *dane*, który oznacza domyślny strumień danych, udostępniany w ramach operacji na pliku. Można jednak utworzyć dodatkowy, nazwany atrybut *dane*, którego strumień dostępny będzie przez jawne podanie nazwy tego atrybutu. Atrybut ten nie występuje w przypadku katalogu, chyba że zostanie jawnie utworzony nazwany atrybut tego typu.

Atrybuty *korzeń indeksu*, *alokacja indeksu* oraz *mapa bitowa indeksu* wykorzystywane są w implementacji katalogów.

Atrybut *lista atrybutów* zawiera listę atrybutów wraz z ich lokalizacją w rekordach rozszerzeń w MFT w przypadku, gdy jeden rekord MFT jest niewystarczający do pomieszczenia informacji o wszystkich atrybutach.

Systemy operacyjne



NTFS — identyfikacja rekordu MFT

- Rekord MFT i tym samym plik identyfikowany jest wewnętrznie przez 64-bitową referencję.
- 48 mniej znaczących bitów referencji jest numerem rekordu w tablicy MFT.
- 16 bardziej znaczących bitów referencji jest numerem sekwencyjnym zwiększanym o 1 po każdym przydzieleniu rekordu do nowo utworzonego pliku, co zapobiega błędnym odwołaniom do plików z użyciem zdezaktualizowanych referencji.

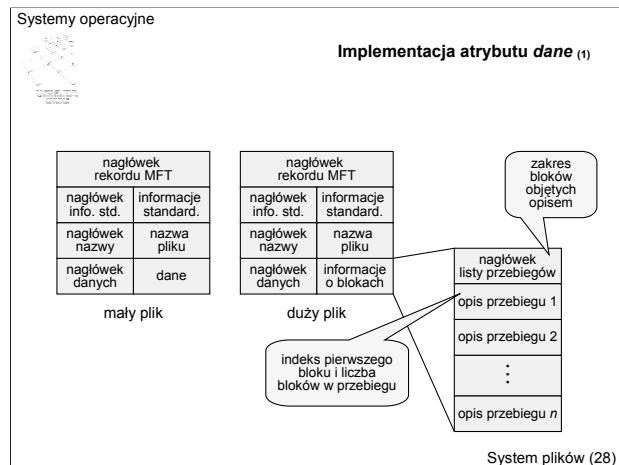
numer sekwencyjny	numer rekordu MFT
-------------------	-------------------

System plików (26)

Systemy operacyjne

**NTFS — rekord bazowy**

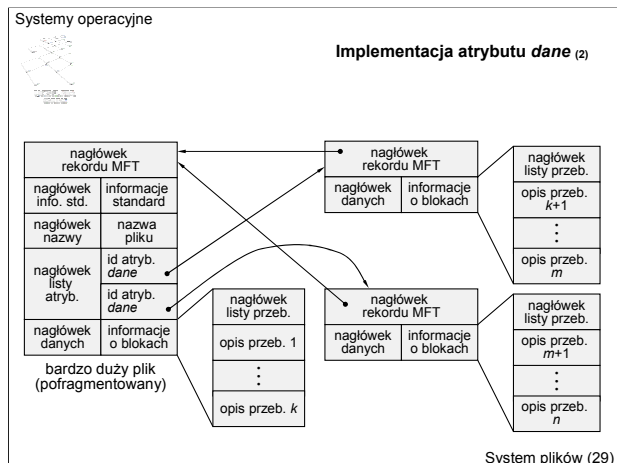
- Jeśli lista atrybutów pliku nie mieści się w jednym rekordzie MFT, do pliku przydzielane są dodatkowe rekordy.
- Pierwszy rekord MFT nazywany jest bazowym (lub podstawowym).
- Dodatkowe rekordy MFT danego pliku określane są jako rekordy rozszerzeń.
- Identyfikacja atrybutów, umieszczonych w rekordach rozszerzeń, znajduje się w atrybucie *lista atrybutów*, znajdującym się w rekordzie bazowym.



Mały plik to taki, dla którego **wartość** atrybutu *dane* mieści się w rekordzie MFT. Rekord taki w najprostszym przypadku składa się z atrybutów: *informacje standardowe*, *nazwa pliku*, *dane*. W przypadku większego pliku, wartość atrybutu *dane* musi zostać przeniesiona do bloków danych poza tablicą MFT, a bloki te muszą zostać opisane odpowiednio w rekordzie MFT.

Opis polega na odwzorowaniu *wirtualnych numerów bloków* danego pliku (virtual cluster number — VCN), czyli numerów kolejnych bloków stanowiących strumień danych, na logiczne numery bloków dyskowych (logical cluster number — LCN). Nagłówek tabeli takiego odwzorowania zawiera zakres numerów VCN objętych opisem w tabeli, a poszczególne pozycje samej tabeli zawierają opisy tzw. *przebiegów* (ang. run). Przebieg jest ciągiem kolejnych bloków wg. numeracji LCN. Opis takiego przebiegu składa się zatem z numeru LCN pierwszego bloku oraz liczby bloków wchodzących w skład przebiegu.

Przykład: strumień danych pliku umieszczony jest w 30 blokach, porozkładanych w 3 przebiegach różnej wielkości. Pierwszy przebieg obejmuje 8 bloków zlokalizowanych pod numerami LCN 1500 – 1507, drugi 10 bloków o numerach 1800 – 1809, a trzeci 12 bloków o numerach 2000 – 2011. Numery VCN dla takiego pliku są z zakresu 0 – 29 i taka informacja znajdzie się w nagłówku tabeli. Pozycje tej tabeli zawierają zatem: pierwsza 1500 i 8, druga 1800 i 10, trzecia 2000 i 12.



W przypadku mocno pofragmentowanego pliku liczba przebiegów może być tak duża, że ich opis nie zmieści się w jednym rekordzie MFT. W rekordzie bazowym musi zatem zostać utworzony atrybut *lista atrybutów* ze wskazaniem na dodatkowe rekordy MFT (rekordy rozszerzeń), w których umieszczony zostanie opis kolejnego zakresu przebiegów. W przedstawionym przykładzie rekord bazowy uzupełniono o 2 rekordy rozszerzeń.

Potencjalnie mogłoby się pojawić potrzeba utworzenia tak dużej liczby rekordów rozszerzeń, że w rekordzie bazowym nie wystarczy miejsca na ich opisanie. Rozwiązaniem jest wówczas przeniesienie atrybutu *lista atrybutów* z rekordu MFT do bloku dyskowego. Atrybut ten stanie się wówczas nierezydentny.

Atrybut *dane* jest tym, który zajmuje zwykle najwięcej miejsca. Każdy inny atrybut, którego wielkość jest znacząca, przechowywany jest według tej samej zasady.

Systemy operacyjne



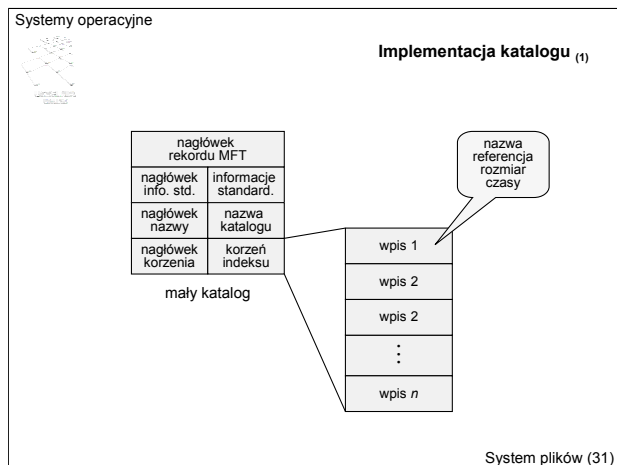
NTFS — katalogi

- Katalog zawiera sekwencję wpisów (rekordów), z których każdy odpowiada jednemu plikowi i obejmuje jego następujące atrybuty:
 - referencja pliku
 - nazwa (wraz z długością)
 - rozmiar
 - czasy dostępu
- Implementacja dużych katalogów oparta jest na strukturze indeksowej (B⁺-drzewo).

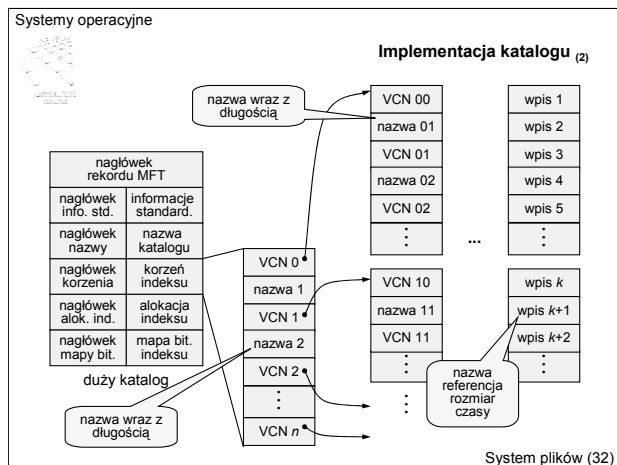
System plików (30)

Katalog jest sekwencją wpisów, zawierających najczęściej wyświetlane atrybuty pliku. Atrybuty te kopiowane są z rekordów MFT plików, wchodzących w skład katalogu. Przyspiesza to sporządzenie listingu zawartości katalogu, gdyż nie trzeba „sięgać” do rekordu MFT każdego pliku.

Istotnym atrybutem rekordu MFT, opisującego katalog, jest *korzeń indeksu* (index root). W przypadku małych katalogów atrybut ten zawiera po prostu sekwencję wpisów. W przypadku większych katalogów, wpisy zorganizowane są w strukturę B⁺-drzewa, w której kluczem jest nazwa pliku. Atrybut *korzeń indeksu* jest korzeniem tego drzewa (nazwa atrybutu staje się adekwatna do jego znaczenia), a dodatkowe atrybuty *alokacja indeksu* oraz *mapa bitowa indeksu* opisują odpowiednio lokalizację bloków z węzłami struktury indeksowej oraz ich wypełnienie.



W przypadku małego katalogu, wpisy katalogowe, posortowane według nazwy, umieszczona są w samym rekordzie MFT.



W przypadku dużego katalogu atrybut *korzeń indeksu*, jako korzeń B⁺-drzewa zawiera klucze (nazwy plików) oraz wskaźniki na węzły B⁺-drzewa z następnymi kluczami. Zgodnie z przykładem, *VCN 0* jest wskaźnikiem na blok wypełniony nazwami, które w porządku leksykograficznym są wcześniejsze niż *nazwa 1*. Podobnie, *VCN 1* jest numerem bloku-węzła z nazwami w zakresie leksykograficznym od *nazwa 1* do *nazwa 2* itd. Węzły na kolejnych poziomach zbudowane są według tej samej zasady. Liście z kolei wypełnione są właściwymi wpisami, obejmującymi oprócz nazw, referencje, rozmiary oraz czasy dostępu.

Odwzorowanie numerów VCN bloków-węzłów na obszar danych zapewnia następny atrybut — *alokacja indeksu*, zbudowany tak samo, jak to opisano dla atrybutu dane.

Systemy operacyjne



NTFS — wybrane pliki metadanych

- \$MFT — główna tablica plików
- \$MTFMIRR — kopia MFT
- \$LOGFILE — plik dziennika
- \$VOLUME — plik wolumenu
- \$DEFATTR — definicja atrybutów plików
- \$ — katalog główny (korzeń drzewa)
- \$BITMAP — wektor bitowy (informacja o wolnych/zajętych blokach)

System plików (33)

Wszystkie informacje, w tym metadane na temat samego systemu plików, przechowywane są w plikach. Jak już wspomniano, sama tablica MFT jest plikiem i pierwszy rekord tej tablicy opisują ją samą jako plik o nazwie \$MFT. \$MFTMIRR jest plikiem, zawierającym kopię wpisów dla metadanych i zlokalizowanym w środkowej części partycji.

Na wypadek konieczności odtwarzania spójności systemu plików wszystkie modyfikacje metadanych (również tych, zlokalizowanych w innych rekordach MFT, opisujących normalne pliki i katalogi) rejestrowane są w tzw. dzienniku (kronice), przechowywanym w pliku \$LOGFILE. W pliku tym rejestrowane są właściwie wszystkie zmiany atrybutów, z wyjątkiem atrybutu *dane*.

Plik \$VOLUME zawiera ogólne informacje o wolumenie (np. rozmiar, etykieta, wersja systemu plików).

Plik \$DEFATTR zawiera definicję typów atrybutów, możliwych do stosowania w systemie plików.

Wpis o nazwie \$ zawiera opis korzenia drzewa katalogów.

Plik \$BITMAP zawiera mapę bitową z informacją o wolnych i zajętych blokach dyskowych.