

Téma: Souborový systém

Předmět: Informační technologie (IKT)

Přidal(a): White 002

Obsah

- Souborový systém
- Příklady souborových systémů
- Znakové sady
- Formátování disku

1. Souborový systém

1.1. Co je to souborový systém a k čemu slouží?

Všechny datové soubory a programy musí být někde uloženy: na pevném disku, CD disku nebo nějakém jiném médiu. Jejich uskladnění má nějaký systém, nejsou tam uloženy chaoticky. Tento systém určuje, jak se mají soubory ukládat, přístupová práva k nim, jak se mají jmenovat, jak jednotlivé soubory hledat, jak se pozná komu patří apod.

Existuje spousta souborových systémů. Většina souborových systémů vznikla současně s vývojem nějakého operačního systému (FAT16, FAT32, NTFS pro Windows), jiné jsou

standardem pro zápis na určité speciální médium.

1.2. Alokační soubory

Alokační bloky neboli clustery. Disk se skládá z pole soustředěných kružnic nazývaných stopy. Každá stopa je rozdělena ještě na sektory. Systém si z jednotlivých sektorů skládá alokační jednotky nazývané clustery. Cluster je nejmenší použitelné množství dat pohromadě. Menší clustery zefektivňují využívání prostoru na disku, větší clustery zrychlují práci s diskem. [3]

1.3. Fragmentace

Rozeznáváme dva druhy - vnitřní a vnější.

Vnitřní fragmentace vzniká tehdy, když zůstávají nevyužity velké části alokačního disku. Vnější fragmentace znamená že je jeden soubor rozmístěn na mnoha místech pevného disku a tím zpomaluje rychlost čtení a zápisu. [2]

1.4. Symbolický systém souborů

Na úrovni symbolického systému se používá jména souborů k jednoznačnému určení odpovídajícího zápisu v adresáři souborů.

1.5. Logický systém souborů

Provádí transformaci struktury logických vět na lineární řetězec slabik, s nímž se pracuje na

úrovni fyzického systému souborů.

1.6. Fyzických systémů

Fyzický systém souborů zajišťuje přidělováním diskového prostoru souborům. Transformuje logické adresy na adresy fyzických bloků.

1.7. Způsoby přidělování diskového prostoru souborům

1.7.2. Přidělování souvislých oblastí

Je to nejjednodušší způsob. U této metody zabírá konkrétní soubor množinou sousedních bloků disku. Výhodou je menší potřeba posunů diskové hlavy, nevýhodou je délka hledání volného prostoru.

1.7.2. Přidělování nesouvislých oblastí disku

1.7.2.1. Metodou Zřetězení

Je používána například u souborového systému FAT. Jde o jednoduchý a rychlý princip, jelikož stačí uchovávat pouze počáteční adresu.

1.7.2.2. Využitím mapy souboru

Ke každému souboru existuje mapovací tabulka souboru MTS. Hodnota -1 značí neobsazeno.

2. Příklady souborových systémů

2.1. FAT

FAT znamená, File Allocation Table neboli Tabulka rozložení souboru. Jedná se o fyzickou oblast na oddílu, ve které najdeme informace o alokaci jednotlivých clusterů. Položky mohou nabývat délek 12, 16 nebo 32 bitů. Systém FAT má na začátku diskového oddílu sektor s nultou pozicí nazývaný boot sektor. Na konci tohoto sektoru je obslužný program, který se spouští při bootování, aby zavedl a spustil systémové soubory. Za boot sektorem následuje oblast FAT. Ta má obvykle dvě kopie, protože při chybě v první kopii, ji můžeme pomocí druhé rekonstruovat a data tak zachránit. Za touto oblastí navazuje oblast kořenové složky – root. Po rootu již následuje volné místo použitelné pro data našich souborů. Její délka není jednoznačná a závisí na konkrétním disku. V tabulce FAT je zaznamenáno umístění všech souborů na disku. Při požadavku na přečtení souboru je nejdříve v tabulce zjištěno fyzické umístění souboru na disku, a poté je soubor přečten. [1]

2.1.1. FAT 12

Tento systém se používá pro jednotky FAT, které mají méně než 4096 clusterů (212bitů). V současné době je nalezneme pouze na disketách (disky s velikostí oddílu pod 32 MB se již v dnešní době nevyskytují). [1]

2.1.2. FAT 16

Systém FAT 16 schopen adresovat až 65536 clusterů (216bitů). Libovolně velký disk je tedy při formátování rozdělen na maximálně 65536 částí, velikost jedné části se pohybuje od 512 B do 32 kB. S rostoucí kapacitou disku roste i podíl nevyužitého místa, protože i ten nejmenší

soubor (např. o velikosti 1 kB) zabere celý cluster, což v případě 32 kB clusteru znamená ztrátu 31 kB. Dalším problémem je fragmentace obsahu disku. Při zápisu souboru na disk je využito první volné místo bez ohledu na to, jestli jeho velikost dostačuje zápisu celého souboru. Často dochází k situaci, kdy jsou jednotlivé části jednoho souboru uloženy na různých místech disku. [1]

2.1.3. FAT 32

32bitová alokace clusterů umožňuje adresovat disky až do velikosti 2 TB. Pro disky do 8 GB se použijí clustery 4 kB, což znamená zvýšení efektivity využití prostoru disku asi o 10 až 15 % vzhledem k velkým diskům na FAT16. Užívá záložní kopii FAT na místo běžné druhé kopie. Boot sektor je rozšířen a je v něm uložena záloha kritické struktury dat, to znamená že FAT32 je méně náchylná na poruchy než FAT16. Problémy jsou především s absencí komprese, šifrováním a bezpečnostních vlastností složek a souborů na úrovni souborového systému. [1]

2.2. NTFS

Jedná se o souborový systém Microsoftu podporovaný jejich operačními systémy. Jinými operačními systémy není s plnou funkcí podporován.

Posunuje velikost nejmenší alokovatelné jednotky až na 512 bajtů. Používá 16bitové kódování Unicode, což umožňuje pojmenovat soubor v libovolném jazyce. Logické disky i soubory mohou být rozprostřeny přes více fyzických disků a poškozené soubory mohou být rekonstruovány. Z dalších novinek (oproti FAT) jmenujme například zabezpečení, protokolování, automatickou kompresi dat a šifrování.

Stejně jako každý systém dělí NTFS využitelné místo na clustery - bloky bytů použity pohromadě. NTFS podporuje skoro všechny velikosti clusterů - od 512 B až do 64 kB. 4kB cluster je považován za standard. Disk NTFS je symbolicky rozdělen do dvou částí. Kromě uživatelských dat přidává NTFS na disk i tzv. METADATA. Ta jsou ukládána do 16 speciálních skrytých souborů. Jedním z nich je i Master File Table (MFT). Každý soubor začíná záznamem

atributu v MFT.

MFT zabírá oblast začátku disku, přičemž si pro sebe vyhradí větší. Nedostává-li se na disku místa pro soubory, volné místo určené MFT se na nějaký čas zredukuje. Uprostřed disku je záložní kopie nejdůležitějších souborů MFT. [1]

2.2.1. Metasoubory

Prvních 16 souborů (metasouborů) jsou systémové soubory zodpovědné za systémové operace. Metasoubory jsou v rootu NTFS, jejich název začíná znakem „\$“, takže je obtížné pomocí běžných prostředků získat o nich nějaké informace. V následující tabulce je pro úplnost soupis metasouborů a jejich funkce:

\$MFT – Sám MFT

\$MFTmirr – Kopie prvních 16 záznamů MFT umístěná do středu disku

\$LogFile – soubor pro protokolování

\$Volume – interní informace – název oddílu, verze systému souborů atd.

\$AttrDef – soupis standardních atributů souborů na oddílu

\$. – složka rootu

\$Bitmap – bitová mapa volného místa oddílu

\$Boot – boot sektor (bootovatelný oddíl)

\$Quota – soubor, ve kterém jsou uložena práva uživatele na užití místa disku (od NT5)

\$Ucase – Soubor, tabulka shody malých a velkých písmen v názvech souborů na daném oddílu. Je důležitá, jelikož názvy souborů NTFS jsou zapsány v Unicodu, který tvoří 65 tis. různých znaků a není jednoduché vyhledávat kvůli ekvivalenci malých a velkých písmen.

\$Extend – Používá se pro různá volitelná rozšíření jako jsou kvóty, data bodů změny zpracování, a identifikátory objektů. [1]

2.2.1.1. Protokolování

NTFS je systém zajištěný proti chybám, který dokáže sám opravit prakticky jakoukoli reálnou poruchu. Kterýkoliv moderní souborový systém je založen na koncepci transakcí – akce provedená zcela a správně nebo vůbec neprovedena. [1]

2.2.1.2. Komprese

NTFS je postavena s podporou diskové komprese. Jakýkoliv soubor či složka mohou být jednotlivě uloženy na disk v komprimované formě a tento proces nijak neomezuje aplikace. Komprese souborů má velmi vysokou rychlost a jen jednu negativní vlastnost – velkou virtuální fragmentaci komprimovaných souborů.

2.2.1.3. Zabezpečení

NTFS skýtá mnoho prostředků k rozlišení práv objektů. Práva systému souborů NTFS jsou úzce spojené se samotným systémem a to znamená, že nejsou povinné při fyzickém přístupu na disk jiným operačním systémem.

2.2.1.4. Šifrování

Každý soubor složky může být zašifrován a pak není možné, aby ho někdo přečetl jinou instalací NT. V kombinaci s velice bezpečným heslem tato možnost nabízí bezpečnost vybraných důležitých dat pro většinu aplikací.

2.2.1.5. Kvóty

Administrátor může každému uživateli nastavit, kolik může do dané složky uložit dat. Výhodné pro servery nebo stanice s přístupem více lidí. [2]

2.2.1.6. Typy svazků NTFS

2.2.1.6.1. Jednoduché svazky

Jde o svazky vytvořené na jednom fyzickém disku přes jeden nebo libovolný počet oddílů NTFS, které spojíme dohromady.

2.2.1.6.2. Rozložené svazky

Pokud jednoduchý svazek rozšíříme přes několik fyzických disků, stává se jednoduchý svazek rozloženým svazkem.

2.2.1.6.3. Prokládané svazky

Data takového svazku jsou ukládána střídavě na dva nebo více fyzických disků. Data těchto svazků jsou alokována střídavě a rovnoměrně na disky prokládaného svazku. Tyto svazky značně urychlují přístup k harddisku.

2.2.1.6.4. Zrcadlené svazky

Jde o svazky vylučující chyby, které ukládají data ve dvou kopiích na dva fyzické disky. Poskytují zálohu dat tím, že používají kopii (zrcadlení) svazku, aby duplikovali informace uložené na svazcích. Zrcadlo je uloženo na jiný fyzický disk. Pokud data na jednom disku začnou vykazovat chyby a stanou se nepoužitelnými, systém pokračuje v operaci a používá nepoškozená data druhého disku. [1]

2.2.2. Použitelnost NTFS

Všeobecně je NTFS vysoce výkonná pro jakýkoliv disk, avšak doporučovaná je až od velikosti oddílu 512 MB. Není problém ji však použít na oddíly menší. Výhoda pak spočívá, v možnosti si navolit velikost clusterů. Můžeme na ni aktivovat účinnou a rychlou kompresi, nebo také šifrování.

3. Znakové sady

Zatímco znaková sada je prostý souhrn znaků, kódování znaků definuje navíc i přiřazení kódů jednotlivým znakům. Kódem může být například Morseova abeceda. Pro kódování znaků se někdy používá termín kódovaná znaková sada.

3.1. Moderní přístup ke kódování

Cílem této dekompozice je ustavit univerzální znakovou sadu, která může být kódována množstvím způsobů.

3.1.1. Znakový repertoár

Znakový repertoár je plná sada znaků podporovaná daným systémem. Může být otevřený nebo uzavřený.

3.1.2. Kódovaná znaková sada

Kódovaná znaková sada určuje, jak má být zobrazený repertoár znaků užitím nezáporných celých (1,2,3,4,5...) čísel nazývaných kódové body.

3.1.3. Forma kódování znaků

Forma kódování znaků popisuje převod kódů z kódované znakové sady na sadu celočíselných hodnot kódu pro ukládání v systému reprezentujícím čísla v binární formě za použití omezeného počtu bitů. [4]

3.1.4. Systém kódování

Systém kódování definuje rozklad kódových hodnot na byty. Srovnává je za sebou od nejvyššího po nejmenší a naopak.

3.2. Lexikální řazení

Většina jazyků má různé výjimky (např. české Ch), proto nejde realizovat lexikální řazení přidaných znaků do znakové sady.

3.3. Znakové sady a HTML

Aby nemusel uživatel u každé stránky hledat v menu znakovou sadu vhodnou pro její zobrazení, je možné v HTML hlavičce uvést znakovou sadu, ve které byla daná stránka vytvořena. [4]

3.4. ASCII

ASCII je základní kódovací sadou, z které vychází v euro-americkém prostoru ostatní sady. Skoro všechny sady pouze rozšiřují ASCII.

3.5. ISO 8859

Definuje osmibitové znakové sady, používané zvláště v Linuxu.

3.5.1. ISO 8859-1

Používaná pro západoevropské jazyky. Je osmibitová. Prvně byla vytvořena znaková sada ISO 8859, která poté byla základem pro ISO 8859-1. Skládá se ze 191 znaků.

3.5.2. ISO 8859-2

Používaná pro východoevropské jazyky. Je osmibitová. Byla vytvořena 1987. Skládá se ze 191 znaků. V této znakové sadě má největší význam angličtina a němčina.

3.5.3. ISO 8859-3

Používaná pro jihoevropské jazyky. Je osmibitová. Byla vytvořena v roce 1988.

3.6. Windows

Operační systémy rodiny Microsoft Windows používají několik osmibitových kódování, více nebo méně podobných standardům ISO 8859.

3.6.1. Windows-1250

Používá se při reprezentaci textů ve středoevropských jazycích používajících latinku. Je velmi podobná ISO 8859-2.

3.6.2. Windows-1251

Populární 8bitové kódování znaků, navrženo tak, aby pokrylo jazyky, které používají Azbuku

3.6.3. Windows-1252

Znaková sada používaná implicitně ve starších verzích Microsoft Windows pro angličtinu a některé další západoevropské jazyky.

3.7. Vícebytové sady

3.7.1. UTF-8

Je to zkratka pro Unicode Transformation Format. UTF-8 používá proměnnou délku znaku, od 1 do 4 (resp. 6) bajtů. Je zpětně kompatibilní s ASCII. Zakódované znaky mohou být dlouhé až 4 bajty. Nejvýznamnější bajt ukládá jako první

3.7.2. UTF-16

Kóduje řetězce do posloupností 16bitových slov (2 bajty). Zabírá mnohem více místa a je náročnější.

3.7.3. UTF-32

Kóduje řetězce do posloupností 32bitových slov (4 bajty). Vzniká Přímá reprezentace znaků.

3.8. Kódování češtiny

3.8.1. Kód Kamenických

V podstatě se jedná o sadu CP437. Osmibitové kódování češtiny. Zachovává semigrafické znaky. Znaky 128-171 nahradil českými znaky.

3.8.2. Unicode

Současná celosvětová znaková sada, používá se v současných OS. Značí mezinárodní standardizaci pro kódování znaků s cílem zajištění dostupnosti a kompatibility pro co nejvíce světových jazyků a znaků. Poskytuje prostor pro 1 114 112 znaků. Základní počet znaků je 65 536. Ke kódování používá UTF-8, UTF-16, UTF-32 a UCS-2.

3.8.3. CP852

Osmibitová znaková sada používaná pod operačním systémem MS-DOS pro středoevropské jazyky využívající latinku.

4. Formátování disku

Postup, při kterém je vnější paměť uspořádána tak aby byla připravena k prvnímu použití. Formátování zapíše na disk metadata popisující prázdné místo ve tvaru zvoleného souborového systému. Disk můžeme rozdělit na oddíly, podle toho jaký souborový systém obsahuje. [5]

4.1. Vysokoúrovňové formátování

Vysokoúrovňové formátování je proces, při kterém jsou na disk zapsány informace o

rozdělení disku na oddíly a o souborových systémech.

4.2. Nízkoúrovňové formátování

Základní formátování, které provádí výrobce pevného disku ihned po jeho fyzickém vytvoření. Nejprve jsou vytvořeny značky na začátku a konci stopy, následně jsou označeny začátky sektorů, je označen konec sektoru.

Zdroje

- [1] Studentská sms: Souborový systém [online]. [cit. 2020-12-10].
Dostupné z: <http://studentka.sms.cz/seminarka/souborove-systemy>
- [2] Wikipedie: Souborový systém [online]. [cit. 2020-12-10].
Dostupné z: https://cs.wikipedia.org/wiki/Souborov%C3%BD_syst%C3%A9m
- [3] ČVUT – oddělení kybernetiky: Souborový systém [online]. [cit. 2020-12-10].
Dostupné z: <http://labe.felk.cvut.cz/vyuka/A3B33OSD/Tema-07-Soubory-OSD.pdf>
- [4] Wikipedie: Kódování znaků [online]. [cit. 2020-12-10]. Dostupné z:
https://cs.wikipedia.org/wiki/K%C3%B3dov%C3%A1n%C3%AD_znak%C5%AF
- [5] Wikipedie: Formátování [online]. [cit. 2020-12-10]. Dostupné z:
https://cs.wikipedia.org/wiki/Form%C3%A1tov%C3%A1n%C3%AD_disku