

MATERIAŁY Z ARCHIWALNYCH NUMERÓW CZASOPISMA „CHIP”

CHIP nr 05/2000

CHIP nr 08/2001

BIOS lub system operacyjny czasami zadziwiają nas informacjami o wolnej lub zajętej przestrzeni dyskowej. Przyczyny takiego stanu rzeczy mogą być bardzo różne. Czasami zdarza się, że pojemność dysku jest zaniżana. Jeśli natomiast dysk ma jedynie nieznacznie mniejszy rozmiar, niż wynika to z jego specyfikacji, prawdopodobnie różnica ta powstaje w wyniku niewłaściwego przeliczania kilobajtów na megabajty i megabajtów na kilobajty.

Podczas gdy w nauce i technice przedrostek kilo oznacza mnożenie przez 1000, w informatyce, w której dominuje system binarny, kilo odpowiada wartości 1024 (1 kilobajt = 1024 bajty). Ta sama reguła dotyczy mega i gigabajta. Pojemność 9-gigabajtowego dysku równa jest więc 9 663 676 416 bajtom ($9 \times 1024 \times 1024 \times 1024$ bajty). Tymczasem producenci czy dystrybutorzy ze względów marketingowych często podają rozmiar, stosując przeliczenia dziesiętne - w ten sposób dziewięciogigabajtowy dysk oznaczany jest jako 9,6 GB (9 663 676 416 bajtów)

Inne przyczyny straty pojemności

Pojemność dysku może też ulec zmniejszeniu w wyniku jego przeadresowania logicznego (translacji). Wyobraźmy sobie dysk o następujących parametrach fizycznych: 10 210 cylindrów, 16 głowic i 63 sektory na każdej ścieżce. W sumie dysk ma 10 291 680 sektorów, czyli 5 269 340 160 bajtów. W ramach przeadresowania logicznego całkowita liczba sektorów dzielona jest przez liczbę sektorów przypadających na jedną ścieżkę, co daje wartość 163 360. Wynik ten dzielony jest przez maksymalną liczbę sektorów dopuszczalną przez BIOS, czyli 1 024. Iloraz 159,5 jest teoretycznie wyznaczoną liczbą głowic. W praktyce jednak ilość głowic musi być równa potęgze liczby 2 (16, 32, 64, 128 lub 256). Ostatecznie jako liczba głowic podawana jest najmniejsza wartość większa od wyliczonego ilorazu, w tym przypadku 256.

Aby obliczyć ostateczny parametr przeadresowania dysku, algorytm translacji dzieli wartość 163 360 przez 256, co w wyniku daje 638 cylindrów logicznych. Jeśli teraz na nowo obliczymy objętość tak przeadresowanego dysku, okaże się, że mamy do dyspozycji 10 289 664 sektory, czyli o 2 016 sektorów mniej niż przed translacją. Na szczęście różnica ta odpowiada jedynie 1 008 KB, a więc w wyniku translacji tracimy mniej niż 1 MB przestrzeni dyskowej.

Wielkość klastrów a rozmiar partycji

Jedną z przyczyn utraty sporej części użytecznej przestrzeni dyskowej tkwi w budowie systemu plików. Pojedyncze sektory, będące fizycznie najmniejszymi jednostkami pamięci dyskowej, grupowane są przez system plików w klastry. Z punktu widzenia systemu operacyjnego najmniejszą jednostką alokacji jest właśnie klaster. Takie postępowanie pozwala poprawić wydajność (do zaadresowania danych potrzebne jest mniej bitów), jednak w przypadku zapisywania wielu małych plików sporo miejsca na dysku pozostaje nie wykorzystane (każdy plik musi mieć początek w nowym klastrze). Jeśli na partycji z klastrami o wielkości 32 KB zapiszemy pięciokilobajtowy plik, zajmie on cały klaster, a więc stracimy w ten sposób pozostałe 27 KB. W skali całego dysku możemy z tego powodu stracić nawet setki megabajtów.

Rozwiązaniem tego problemu są mniejsze klastry, dlatego począwszy od wersji Windows 95 OSR2, najlepiej używać systemu plików FAT32 zamiast FAT16. Podczas gdy na partycji 2 GB FAT16 tworzy klastry wielkości 32 KB, FAT32 zużywa jedynie 4 KB na klaster. Użytkownicy starszej wersji Windows 95 (skazani na FAT16) muszą dobrze skalkulować optymalny rozmiar partycji, pamiętając o tym, że im większa partycja, tym większe na niej klastry. Często opłaca się podzielić dysk na kilka mniejszych części (patrz: tabela "Liczba i rozmiar klastrów..." na s. 123).

W tabeli podane są absolutne wartości graniczne dla systemów FAT. FAT12 nie jest już używany dla dysków twardych i można się z nim spotkać jedynie na dyskietkach.

Ograniczenia wynikające z maksymalnej liczby klastrów			
	FAT12	FAT16 i VFAT	FAT32

Rozmiar wpisu w tablicy FAT	12 bitów	16 bitów	28 bitów*
Maks. liczba klastrów na jednej partycji	4 086	65 536	268 435 456
Min. rozmiar klastra	512 B	2 kB	4 kB
Maks. rozmiar klastra	4 kB	32 kB	32 kB
Maks. rozmiar partycji	16 MB	2 GB	ok. 2 TB
* dodatkowe 4 bity zarezerwowane są dla specjalnych informacji.			

Alternatywą może być formatowanie

Mało znaną metodą określania rozmiarów klastrów jest nieudokumentowany parametr komendy format, dostępny już w Windows 95. Opcja **/z:n** tworzy klastry o rozmiarze n kilobajtów. Standardowo wielkość ta ustawiona jest na 4 KB. Aby sformatować drugą partycję dysku twardego, moglibyśmy wydać na przykład polecenie **format d: /z:4**. W Windows NT mamy natomiast do dyspozycji komendy **format d: /f:NTFS /a:4k** i **format d: /f:FAT /a:4k**. Musimy jednak pamiętać o dwóch rzeczach. Po pierwsze, nie możemy przekroczyć maksymalnej dopuszczalnej liczby klastrów dla partycji. Po drugie, duża liczba małych klastrów negatywnie wpływa na wydajność, o czym przekonamy się najpóźniej przy pierwszej próbie uruchomienia programu do defragmentacji dysku.

Klastry występują także w Linuksie i zarządzane są przez drzewo tzw. l-nodes. Po zainstalowaniu na przykład dystrybucji SuSE i po podziale dysku na partycje właściwości systemu plików możemy przeglądać za pośrednictwem programu *YaST*

Liczbę l-nodes możemy określić ręcznie, jednak samodzielna konfiguracja tego ustawienia ma sens i jest potrzebna tylko w niewielu przypadkach, na przykład w trakcie instalacji serwera grup dyskusyjnych. Do typowych zastosowań wystarcza całkowicie domyślna wartość l-nodes.

W obu tabelach znajdują się standardowe rozmiary klastrów w zależności od zadanego rozmiaru partycji i wybranego systemu plików.

Klastry w FAT16			Klastry w FAT32		
Rozmiar partycji	Rozmiar klastra	Liczba sektorów w klastrze	Rozmiar partycji	Rozmiar klastra	Liczba sektorów w klastrze
16-128 MB	2 kB	4	512-8 192 MB	4 kB	8
128-256 MB	4 kB	8	8-16 GB	8 kB	16
256-512 MB	8 kB	16	16-32 GB	16 kB	32
512-1 024 MB	16 kB	32	Ponad 32 GB	32 kB	64
1 024-2 148 MB	32 kB	64			

Symbolom MB i GB przypisywane są wielkości według dwóch różnych systemów. W zeszłym roku w wielu krajach świata weszła w życie norma mająca wyeliminować tę niejednoznaczność.

Rozmiary pamięci elektronicznej bywają często ograniczone liczbą bitów użytych do jej adresowania. Przykładowo: jeśli bitów tych jest 10, to zaadresować możemy 1024 komórki pamięci. Wielkość ta niewiele różni się od 1000. W powszechnie stosowanym systemie SI mnożnik 1000 oznacza się małą literą "k", dlatego też mnożnik 1024 przyjęło się zapisywać za pomocą przedrostka podobnego, ale jednak różnego - poprzez wielkie "K". Obydwa przedrostki czytają się jednak tak samo, czyli "kilo".

Dziwaczna definicja gigabajta, nieprawdaż?! Ale przynajmniej wiadomo, jaka jest rzeczywista pojemność tego "twardziela".

Dawno minęły czasy, gdy wielkość pamięci podawano w kilobajtach. Pojawiły się kolejno megabajty i gigabajty, a ostatnio także terabajty. No i mamy problem - jak to zapisać? Zapewne z rozpędu do kolejnych potęg liczby 1024

zaczęto stosować przedrostki z układu SI, które odpowiadają kolejnym potęgom liczby 1000. W układzie SI "M" oznacza okrągły milion, a tymczasem w sensie "komputerowym" "M" nagle zaczęło też oznaczać 1 048 576, czyli prawie 5% więcej. Jakby tego nie było dość, pojawiło się trzecie znaczenie symbolu "MB". Mianowicie dyskietki "1,44 MB" z niezrozumiałego powodu mają rozmiar podawany w jednostkach MB wynoszących... 1 024 000 bajtów. W taki oto sposób zaistniała sytuacja, w której tym samym przedrostkom przypisywane są różniące się od siebie wielkości. Krótko mówiąc, przedrostki skalujące dopadła schizofrenia. Nie ominęła ona nawet przedrostka "kilo". Zdarza się, że duże "K" wypiera małe "k" z jego roli. Bywa też - kto wie, czy nie częściej - odwrotnie, gdy małe "k" stosowane jest w znaczeniu 1024.

Płaciłem za 30 giga, a mam 28!

Wobec takiej dwuznaczności, czytając wyrażenie w rodzaju "30 GB", nie możemy już mieć pewności, o jaką dokładnie ilość bajtów chodziło piszącemu. Zgodnie z układem SI podaje się zwykle liczby bajtów tam, gdzie nie występuje adresowanie binarne lub nie ma ono istotnego znaczenia, np. przy podawaniu szybkości transmisji, pojemności dysków twardych czy pamięci taśmowych. W jednostkach binarnych natomiast podaje się głównie wielkość pamięci, gdzie dają się we znaki ograniczenia wynikające z liczby bitów adresowych. Dotyczy to np. pamięci półprzewodnikowych. Nie wiedzieć czemu, również rozmiary plików oraz pojemność płyt CD-ROM podawane są zazwyczaj w takich jednostkach. Ich stosowanie mogłoby sugerować, że użytkownicy komputerów lubią przeliczać wielkości swoich plików na jednostki alokacji, sektory i pozycje w FAT-cie, bo tylko do takich zastosowań jednostki te mogą być wygodne.

Do przykrych nieporozumień prowadzi niekiedy podawanie wielkości plików w jednostkach binarnych. Zdarzają się np. reklamacje dysków twardych, bo po sformatowaniu kupiony dysk "30 GB" w Menedżerze plików objawia się jako dysk mający tylko 28 GB. Na szczęście sprawę wyjaśnia odczyt wielkości dysku bezpośrednio w bajtach, których jest - zgodnie z tym, co producent napisał - 30 miliardów.

Mebibajty

Szansą na wyeliminowanie takich dwuznaczności jest upowszechnienie się opublikowanego na początku 1997 r., a w połowie ub.r. uznanego za obowiązujący w Stanach Zjednoczonych oraz krajach Unii Europejskiej, systemu przedrostków dla mnożników będących potęgami liczby 1024. System ten, opracowany przez IEC, z założenia różny od układu SI, jest jednak do niego bardzo podobny i na nim wzorowany. To podobieństwo pozwala na bardzo łatwe jego zapamiętanie, jak też na intuicyjne odczytanie przez osoby, które pierwszy raz się z nim zetkną.

Przedrostki IEC od przedrostków SI różnią się tylko małą literką "i", oznaczającą słowo "binary". Ich nazwy tworzy się, łącząc pierwszą sylabę przedrostka SI z pierwszą sylabą słowa "binary". Przykładowo: jeden megabajt binarny zapisuje się jako 1 MiB, a wymawia "mebibajt" (skrót od "mega binary byte"). Oznacza on 1 048 576 bajtów. Dla porównania: 1 MB, czyli 1 megabajt według układu SI, oznacza 1 000 000 bajtów.

SI niech znaczy, co ma znaczyć

Unormowanie zapisu przedrostków binarnych nie wystarczy do wyeliminowania niejednoznaczności. Konieczne jest jeszcze ich prawidłowe stosowanie. Ale nie tylko ich. Przedrostków SI również należy używać zgodnie ze znaczeniem nadanym im wraz z wprowadzeniem systemu metrycznego ponad 200 lat temu, a następnie "uświęconym" przez system SI. Chodzi m.in. o podawanie rozmiarów plików. Czy pytał ktoś współczesnych użytkowników komputerów (księgowych, sekretarek, tłumaczy, dziennikarzy, młodzieży...), jakie jednostki bardziej im odpowiadają? Jakiej wydają się im bardziej naturalne i przyjazne? Czy nie woleliby kojarzyć napisu "MB" z milionem bajtów niż z milionem czterdziestoma ośmioma tysiącami pięciuset siedemdziesięcioma sześcioma bajtami?

Jaki los czeka nowe przedrostki? Czy norma zostanie zapomniana, podobnie jak to było z normą precyzującą zapis daty, która z wieloletnim wyprzedzeniem mogła zapobiec "pluskwie milenijnej"? Niestety, nadal często widuje się datę w postaci "01.02.03". Czy z przedrostkami IEC będzie podobnie? Miejmy nadzieję, że jednak nie. Najważniejsze, że została oto wreszcie stworzona możliwość jednoznacznego wyrażania liczby bajtów z wykorzystaniem przedrostków skalujących.