

4 Datoteke (File store)

U Glavi 1 datoteke su definisane kao način za memorisanje većih skupina podataka. U datotekama se čuvaju kako podaci tako i programi. Informacije se čuvaju u datotekama a ne u glavnoj memoriji zbog toga što je glavna memorija mnogo manjeg kapaciteta nego što je količina podataka i programa sa kojima se obično radi. Većina informacija koje se čuvaju u datotekama su potrebne u glavnoj memoriji samo u intervalima vremena kada se obrađuju.

Datoteke imaju i tu prednost što su permanentnog karaktera tj. i kada je računar isključen informacije su sačuvane, što obično nije slučaj sa glavnom memorijom. Programi se mogu izvršavati samo kada se i program i podaci nalaze u glavnoj memoriji. Zato se računari moraju konstruisati tako da obezbijede brz i efikasan prenos programa i podataka iz datoteka u glavnu memoriju i obrnuto. Sada će biti opisane metode organizacije datoteka kao i uređaji na kojima se datoteke smještaju.

4.1 Datoteke, slogovi i podaci

Datoteke su slične kartotekama koje su prije pojave računara bile osnovni način čuvanja podataka. Razmotrimo na primjer slučaj fakulteta; datoteka studenata sadrži pojedinačne slogove o svakom studentu, a svaki slog sadrži niz podataka kao na primjer: identifikacioni broj, ime, adresu itd. Univerzitet ima više fakulteta pa za evidenciju studenata može postojati više datoteka.

Medijumi na kojima se memorišu datoteke se dijele na više djelova tako da svaki dio sadrži podatke iz jedne datoteke. Datoteka je skup slogova iste strukture, to jest, unutar jedne datoteke svi slogovi sadrže istu vrstu informacija. Struktura sloga je različita za različite datoteke. Jedan slog unutar datoteke je sastavljen od vrijednosti podataka tako da se svaki podatak lako pronalazi u trenutku obarade datoteke. Podatak ne nosi sa sobom značenje, već se njegovo značenje određuje po kontekstu u kojem se nalazi. Tako broj 1976 sam po sebi ne znači ništa i postaje informacija tek kada se kaže da predstavlja godinu nečijeg rođenja. U slogu je potrebno memorisati samo broj 1976, jer se na osnovu strukture sloga unaprijed određuje značenje svakog podatka.

Papirni obrasci (formulari) koji sadrže podateke (recimo o studentima) mogu činiti neku kartoteku. Po analogiji, datoteku čine slogovi (koji odgovaraju pojedinačnim obrascima), a struktura sloga je analogna rasporedu pojedinačnih podataka na obrascu (broj indeksa, ime i prezime, godina studija itd.). Kao kod papirnih kartoteka, tako i u datotekama slogovi se obično ređaju jedan za drugim po nekom logičkom redosljedu da bi se lakše pronašli kada su potrebni za obradu. Redosljed se pravi na osnovu jednog ili više podataka u slogu (po alfabetskom redosljedu prezimena, po broju indeksa itd.). Oni podaci na osnovu kojih se pravi redosljed slogova u datoteci nazivaju se ključnim podacima, ili prosto ključevima.

Datoteke koje se koriste u računarskim sistemima mogu biti organizovane na više načina, od kojih su tri dominantna: sekvencijalan, indeks-sekvencijalan, random (slučajan). Izbor zavisi od načina obrade datoteke i utiče na efikasnost obrade.

4.1.1 Sekvencijalne datoteke

U sekvencijalnim datotekama slogovi se smještaju jedan za drugim u unaprijed zadatom redosljedu. Kod takvih datoteka ne postoji efikasan način za dodavanje novog sloga koji po redosljedu treba da bude smješten negdje u sredini datoteke. Ako se baš mora dodati novi slog, to je jedino moguće uraditi tako da se u novu datoteku kopiraju slogovi do pozicije na kojoj treba da dođe dodati slog, zatim se smješta dodati slog i kopira ostatak datoteke. Ovo je naravno vrlo neefikasna procedura. Sličan problem nastaje i kada se iz sekvencijalne datoteke briše slog.

Kod pretraživanja slogova, da se pronadje traženi slog, može se primijeniti više strategija: traženje od početka do željenog sloga, polovljenjem intervala traženja i sl.

4.1.2 Indeks-sekvencijalne datoteke

Indeks-sekvencijalna datoteka je slična sekvencijalnoj s tom razlikom što se poseban prostor ostavlja za dodavanje slogova. Osim toga poseban prostor se rezerviše i za indekse kojima se postiže da se pretraživanje slogova vrši ne od vrha datoteke već od pozicije na koju ukazuje indeks. Korišćenjem indeksa se značajno smanjuje vrijeme traženja sloga. U tu svrhu, indeks-sekvencijalna datoteka se dijeli u manje sekcije gdje je svaka sekcija uređena po sekvencijanoj organizaciji. Indeks upućuje na sekciju u kojoj se nalazi traženi slog. Tako se traženje sloga izvršava u dva koraka: najprije se pretražuje indeksni dio da se odredi sekcija u kojoj se traženi slog nalazi, a zatim se unutar sekcije sukcesivno prolazi kroz slogove dok se ne nađe na traženi slog. Kod velikih datoteka može se formirati i čitava hijerarhija indeksa za potrebe povećanja efikasnosti traženja.

Indeks-sekvencijalna datoteka se može obrađivati na dva načina: sekvencijalno i pojedinačno. Kada se datoteka obrađuje sekvencijalno postupa se na isti način kao kod sekvencijalne organizacije, a indeksi se zanemaruju. Kod pojedinačne (slučajne) obrade, traženom slogu se pristupa uz pomoć indeksa kako je već opisano. Indeks-sekvencijalna organizacija je u velikoj praktičnoj upotrebi zbog efikasnosti i jednostavnosti korišćenja.

4.1.3 Random (slučajna) organizacija datoteka

Random datoteke se zasnivaju na činjenici da svaki slog u datoteci ima jedinstvenu memorijsku adresu. Adrese slogova su brojevi od 1 do broja slogova u datoteci. Ako se na pogodan način

utvrdi veza između ključa sloga i adrese na kojoj se slog nalazi, tada se pretraživanje datoteke može učiniti vrlo efikasnim.

Transformacija kojom se na osnovu vrijednosti ključa sloga određuje adresa nazivaju se hešing algoritmima.

Najjednostavniji slučaj je kada je transformacija tipa 1:1, to jest, svakom slogu (ključu) se pridružuje jedna adresa. Ovakvu transformaciju, u praksi, često nije moguća, pa je najčešći slučaj tip n:1, to jest više slogova (ključeva) hešing algoritmom dobijaju istu adresu. Zbog toga hešing algoritmi, pored preslikavanja ključa u adresu, moraju sadržati i metod za razrješavanje problema jedna adrese za više ključeva (takozvanih hešing sinonima).

Postoji više načina za rješavanje ovog problema. Dva jednostavna rješenja su sljedeća: (1) sinonim slogovi se smještaju na prvu slobodnu adresu od vrha datoteke; (2) sinonim slogovi se smještaju na najbližu slobodnu adresu počev od adrese na kojoj su trebali da budu smješteni na osnovu hešing transformacije.

Random organizacija je veoma pogodna kada se slogovima datoteke pristupa u slučajnom redoslijedu, kao na primjer kada korisnici kreditnih kartica slučajno dolaze do kompjuterizovanih šaltera na kojima ulažu ili podižu novac.

4.1.4 Sistemi datoteka i zaštita pristupa

Računarski sistemi su u stanju da smjeste i rade sa mnogo datoteka i sa mnogo različitih korisnika podataka u njima. Zato je potrebno u takvom sistemu obezbijediti adekvatnu zaštitu podataka od neautorizovanog pristupa. Uobičajeni način za provjeru prava pristupa podacima je pomoć pasvorda (password) koji se dodjeljuju korisnicima i definišu njihova prava za korišćenjem podataka u datotekama.

U trenutku formiranja datoteke, obično se definišu prava za njeno korišćenje i pasvordi kojima se ta prava provjeravaju. Tipična prava korišćenja datoteka su: privatna datoteka (može joj pristupiti samo kreator-vlasnik), datoteka samo za čitanje (read only), datoteka čiji se sadržaj može mijenjati (read-write), datoteka koja sadrži program koji se može samo izvršavati (execute only). Jednoj istoj datoteci može biti pridruženo više pasvorda sa različitim pravima pristupa.

4.1.5 Baze podataka

Pod bazama podataka uobičajeno je da smatramo skupove uzajamno povezanih datoteka (organizovanih na prethodno opisani način). Potreba za organizacijom baza podataka je nametnuta praktičnim problemima gdje se integriše više procesa obrade podataka u jedinstven

proces. Na primjer, pomenuta datoteka studenata nije dovoljna za obradu podataka na jednom fakultetu. Pored podataka o studentima potrebno je imati podatke o profesorima, predmetima, ispitima, učionicama itd. Kada se želi obrada podataka za čitav univerzitet problem se još više komplikuje.

Za organizaciju baza podataka, pored organizacije datoteka, potrebno je obezbijediti i dodatne veze među datotekama. Skup hardverskih i softverskih rješenja koja omogućavaju organizaciju baza podataka nazivaju se sistemima za upravljanje bazama podataka (SUBP). Razvijeni su i posebni programski jezici kojima se definise kako struktura baze tako i procedure za obradu podataka (ORACLE, INFORMIX, ACCESS i sl.).

4.2 Medijumi za smještaj (memorisanje) datoteka

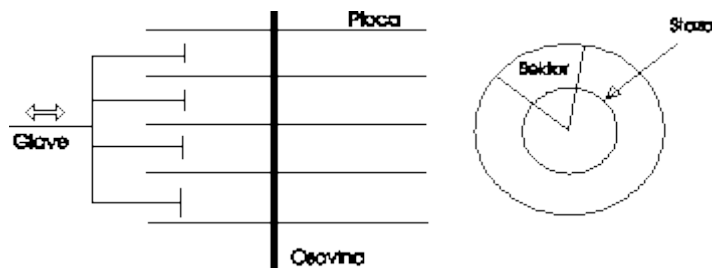
U ovom odjeljku biće opisane tehnička rješenja uređaja koji služe za smještaj datoteka. Takođe će biti razmotreno kako se na takvim uređajima mogu smještati datoteke organizovane na gore navedene načine. Većina uređaja koji se danas koriste za smještaj datoteka koriste ili magnetni ili optički medijum.

Kod magnetnih medijuma iskorišćena je mogućnost da se pomoću magnetnog polja "zapisuju" podaci na magnetni materijal. Čitanje se vrši tako što namagnetisana podloga prolazi pored magnetnog detektora.

Optički uređaji koriste lasersku tehnologiju za upisivanje podataka na optičku podlogu, a čitanje se vrši optičkim detektorima (foto diodama).

4.2.1 Magnetni disk

Sistemi sa magnetnim diskovima se sastoje od dva glavna dijela. Prvi dio čine okrugla **ploča** smještena na centralnu osovinu. Ploča je presvučena magnetnim materijalom obično sa obje strane. Prečnik ploče varira od 3.5 do 15 inča kod različitih disk sistema. Drugi dio disk sistema je upisno/čitajuća **glava** smještena unutar kućišta. Glava može da se pomjera koaksijano u odnosu na centar okrugle ploče koja tokom rada rotira. Kod manjih sistema imamo samo po jednu ploču i jednu upisno/čitajuću glavu. Međutim, kod većih sistema postoji više ploča na istoj osovinu a svaka pojedinačna ploča ima svoju upisno/čitajuću glavu. Na slici 4.1 prikazan je jedan takav disk sistem, gdje su sve glave povezane tako da imaju istu relativnu poziciju u odnosu na ploču sa kojom rade.



Slika 4.1 Magnetni disk

Da bi se razumio rad diska potrebno je još da pojasnimo neke termine. **Stazom** (track) se naziva dio jedne disk ploče koji može biti pročitan jednom glavom bez njenog pomjeranja (u jednom okretanju disk ploče). Na jednoj disk ploči se tako nalazi više koncentričnih staza. Broj staza je najčešće u opsegu od 30 do 400. U slučaju disk sistema sa više ploča govorimo o **cilindru** kao o skupu staza koje glave mogu čitati bez pomjeranja. Sistem prikazan na slici 4.2 ima osam staza po cilindru. Broj podataka koji može biti memorisan na jednoj stazi je obično veliki (ponekad i veći od kapaciteta glavne memorije). Zato se staza dijeli na određeni broj **sektora**, tako da se čitav sektor prenosi u memoriju računara. Sektor može sadržati jedan ili više slogova neke datoteke, a nekada i samo dio sloga.

Disk sistem pored disk ploče i glave, ima i sistem za upravljanje pozicijom glava (jednostavna robotska ruka). Pored toga svaki sektor ima oznaku koja je smještena u prostoru između dva uzastopna sektora. Na taj način sistem za upravljanje uvijek zna na kojem dijelu diska se nalazi i u mogućnosti je da pročita pravi sektor na osnovu njegove adrese.

Učitavanje traženog sloga neke datoteke na disku obavlja se tako u dva koraka: glava se najprije pomjera do staze na kojoj se slog nalazi, a zatim čeka da pored nje prođe sektor sa traženim slogom. To znači da je za učitavanje sloga potrebno znati tri podatka : cilindar, stazu i sektor. O njihovim adresama vodi računa softver za upravljanje datotekama. Najviše vremena pri pristupu slogu troši se na pomjeranje glave do željenog cilindra (staze) i čekanja na okretanje ploče do željenog sektora. Ova vremena su reda veličine milisekunde.

Prije upotrebe disk medijum mora biti pripremljen za upotrebu - **formatizovan**; to je najčešće proces kojim se upisuju adrese cilindra, staza i sektora u među- sektorski prostor. Ove adrese će se kasnije koristiti pri smještaju datoteka na disk.

Disk sistemi se u praksi pojavljuju u više formi: fleksibilni (**floppy**) i tvrdi (**hard**) diskovi. Razlika među njima je samo u ploči na kojoj se nanosi magnetni materijal. Kod floppy diskova ona je plastična i savitljiva, a kod tvrdih diskova metalna.

Floppy diskovi

Floppy sistemi su vrlo jednostavni i jeftiniji po cijenu tolerancija pri izradi. Međutim imaju mnogo manju brzinu transfera podataka i manji kapacitet od hard diskova. Najčešće se koriste za razmjenu podataka i programa među računarima koji nijesu povezani u mrežu. Postoje dva standardna formata floppy diskova: 5,25 i 3,5 inča, od kojih je drugi format dominantan. Floppy diskovi se najčešće nazivaju disketama (deminutiv od disk). Disketa se ubacuje u disk uređaj (drajv) koji sadrži glave za čitanje/pisanje. Kapacitet disketa varira u zavisnosti od tipa: single-side (ploča je sa magnetnim materijalom samo sa jedne strane), double-side (magnetni materijal nanesen na obje strane ploče), single-density (jednostruka gustina zapisa) ili double-density (dvostruka gustina). Tipičan kapacitet disketa je 1,44 mega bajta (Mbyte).

Diskete formata 3,5 inča su danas dominantne zbog više pogodnosti u odnosu na stariju verziju formata 5,25 inča:

Bolja zaštita u plastičnom omotaču (za razliku od papirnog kod 5,25 formata),

Olakšan transport (može da stane u džep na košulji),

Veći kapacitet,

Jednostavnija zaštita od upisivanja (mali prozorčić u jednom uglu diskete).

Hard (tvrđi) diskovi

Tvrđi diskovi se pojavljuju u dva oblika: izmenljivi i neizmenljivi. Kod izmenljivih diskova magnetni disk se po želji može ubaciti u disk drajv. Na taj način imamo neogranižen kapacitet za smještaj datoteka, ali su samo neke od njih raspoložive u datom trenutku. Kod PC računara najčešće su u potrebi neizmenljivi diskovi kod kojih su i medijum (ploče) i uređaj za čitanje fiksirani u jednom kućištu. Kapacitet diskova je danas dostigao vrijednosti reda gigabajta (1000 megabajta).

4.2.2 Optički diskovi

Osnovni pravac razvoja diskova je danas u oblasti optičkih diskova uz korišćenje laserske tehnologije i svi su izgledi da će u računarskoj tehnologiji izazvati sličan bum kao što su CD ploče izazvale u Hi-Fi tehnici. Trenutno su raspoloživi samo za jednokratno upisivanje, ali se uskoro očekuju i optički diskovi sa mogućnošću brisanja/pisanja. Disk ploča je napravljena od tankog kompozitnog sastava metala i polimera na kojoj se laserskim "spaljivanjem" zapisuju informacije u vidu mikroopskih rupica ispod plastične ili staklene površine ploče. Za

čitanje se ponovo koriste laserski detektori koji optički zapis konvertuju u digitalni. Konstrukcija optičkih diskova (gdje je metalni film pokriven providnom plastikom ili staklom) daje dugotrajnost i zaštitu od oštećenja pri povišenim temperaturama, vlažnošću, prašini i sličnim uticajima. U poređenju sa magnetnim medijumima informacija na optičkom disku je praktično neuništiva. Optički diskovi se najčešće prave u veličini od 4,5 inča i poznati su kao CD-ROM (compact disk read only memory) sa tipičnim kapacitetom od 640 Mbajta.

Razvoj CD-ROM plejera vodi ka unifikaciji načina čitanja audio, video i datoteka sa istim uređajem. Ovo još više otvara put multimedijalnim računarima gdje se kako je već rečeno kombinuje tekst, stereo zvuk, grafika i animacija.

Primejna optičkih diskova (CD-ROM) ide od zabave, preko simulacionih sistema za obuku (recimo pilota) do arhiviranja u novinarstvu, bibliotekama, muzejima, drzavnoj administraciji.