

Operacijski sistemi



Upravljanje z
datotekami

Vsebina

- Datotečni sistem
 - organizacija podatkov, datoteka, imenik
 - hierarhija in naslavljanje datotek
- Datoteke
 - abstrakcija, tipi datotek
- Kodiranje datotek
- Pripenjanje naprav
- Nadzor dostopa in zaščita datotek

Datotečni sistem

- Pomnilni medij oz. pomnilna naprava
 - omogoča dolgotrajno hranjenje podatkov
 - veliko različnih vrst medijev
 - disk, disketa, zgoščanka, magnetni trak, usb ključ
- naloga OS
 - medij hrani bite oz. bajte
 - uporabnik hrani dokumente, slike itd.
 - OS premošča vrzel med medijem in uporabnikom

Datotečni sistem

- Logična organizacija podatkov
 - visoko nivojski pogled
 - uporabnikov pogled na podatke
 - logični ali konceptualni pogled
 - OS nudi enotno abstrakcijo podatkov
 - ne glede na to kje in kako so shranjeni
 - med različnimi OS je veliko podobnosti
 - osnovni koncept je datoteka

Datotečni sistem

- Fizična organizacija podatkov
 - nizko nivojski pogled
 - za uporabnika je ta pogled skrit
 - način in oblika zapisa podatkov in metapodatkov
 - gre za dejanski fizični zapis na pomnilnem mediju
 - veliko različnih načinov fizične organizacije
 - prilagoditev pomnilnemu mediju
 - gonilnik datotečnega sistema
 - program, ki skrbi za specifičen datotečni sistem

Datotečni sistem

- Datoteka

- osnovna zaključena zbirka podatkov
 - hrani neko vsebino (podatke)
 - dolgotrajna obstojnost (odvisna od pomnilnega medija)
 - souporaba s strani več uporabnikov
- vsebina datoteke
 - podatki, ki jih datoteka hrani
- atributi (meta podatki) datoteke
 - pripadajoči podatki o datoteki, npr.
 - ime datoteke, velikost datoteke
 - datum in čas nastanka in spremembe datoteke
 - lastnik in skupina datoteke
 - dovoljenja za uporabo

Datotečni sistem

- Imenik

- omogoča združevanje (sorodnih) datotek
 - vsebuje datoteke in druge imenike oz.
 - vsebuje seznam datotek
- omogoča hierarhično strukturo
- podobnost z datoteko
 - vsebina je seznam datoteke
 - pripadajoči atributi

imenik, mapa, kazalo,
directory, folder



Datotečni sistem

- Hierarhična struktura imenikov
 - gnezdenje imenikov in datotek
 - poljubna globina gnezdenja
 - odnosi med imeniki
 - **podimenik** (sub-directory)
 - imenik vsebovan v danem imeniku
 - **nadimenik** (super-directory, parent directory)
 - imenik, ki vsebuje dani imenik
 - **korenski imenik** (root directory)
 - imenik, ki nima nadimenika (ni nikjer vsebovan)

Nekateri starejši sistemi nimajo poljubnega gnezdenja.



Datotečni sistem

- Naslavljanje datotek
 - **pot** (path) do datoteke
 - seznam imen imenikov od ustreznega izvirnega imenika do dane datoteke
 - **absolutna pot**
 - pot, ki se prične v korenskem imeniku
 - **relativna pot**
 - pot, ki se prične v *delovnem imeniku*

/	ločilo v poti (Unix)
\	ločilo v poti (Windows)
.	trenutni imenik
..	imenik starša trenutnega imenika
~	domači imenik uporabnika

Datoteke

- Abstrakcija datoteke
 - datoteka je pogosto splošnejši koncept
 - različni **tipi datotek**
 - OS nekatere tipe posebej obravnava



oznaka	opis	primer
-	navadna datoteka (regular file)	/etc/passwd
d	imenik (directory)	/bin
s	simbolična povezava (symbolic link)	/usr/src/linux
b	bločno orientirana naprava (block special device)	/dev/sda
c	znakovno orientirana naprava (character special device)	/dev/tty
p	poimenovana cev (named pipe)	
s	vtičnica (socket)	

Datoteke

- **Navadna datoteka**

- vsebina datoteke

- poljubna interna struktura oz. format datoteke
 - **binarne** in **tekstovne** datoteke
 - OS omogoča notranje operacije (nad vsebino)
 - odpiranje, zapiranje, branje, pisanje, ...

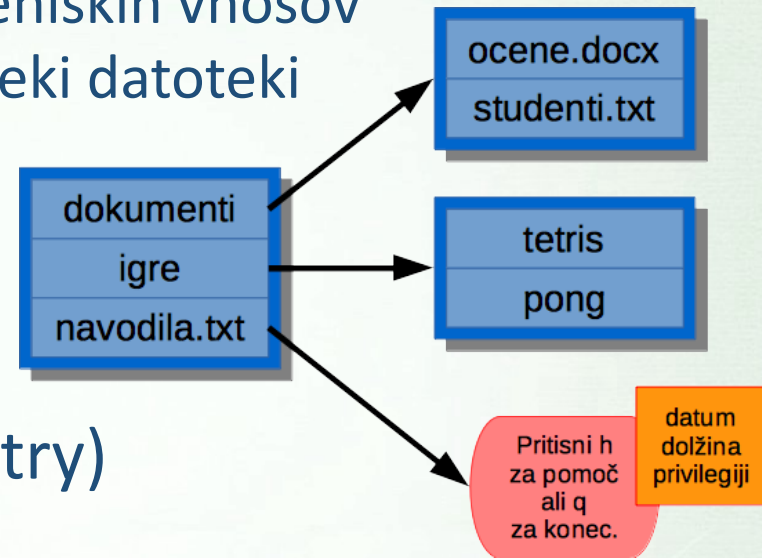
- **vrsta datoteke**

- različne vrste datotek glede na vsebino
 - pogosto del imena (npr. končnica) podaja vrsto
 - veliko različnih vrst
 - odvisno od uporabniških in programskih potreb
 - npr. tekstovna datoteka, izvršljiva datoteka, slika, avdio, video, ...

Datoteke

- **Imenik**

- datoteka posebnega tipa
 - njena vsebina je seznam imeniških vnosov
 - vsak imeniški vnos ustreza neki datoteki



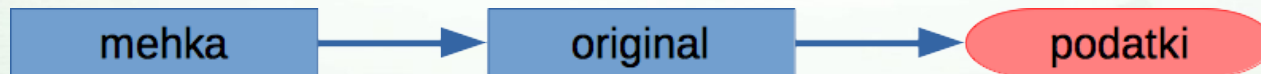
- **imeniški vnos (directory entry)**

- ime datoteke
- kazalci na ostale podatke (vsebina, meta-podatki)
- znotraj imenika ni mogoče imeti imeniških vnosov z enakim imenom

Datoteke

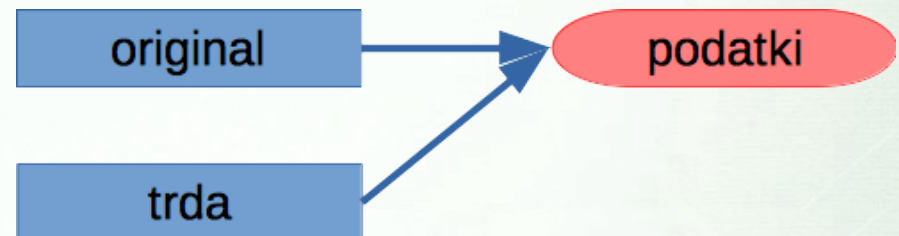
- **Simbolična (mehka) povezava**

- datoteka posebnega tipa
- vsebina datoteke (simbolične povezave)
 - naslov (pot do) ciljne datoteke
- večina ukazov avtomatsko sledi povezave
 - operira na ciljno datoteko
- ukaz: ln -s original mehka



Datoteke

- Trda povezava
 - dodaten imeniški vnos (za isto datoteko)
 - datoteka z več imeni
 - različni imeniški vnosi lahko predstavljajo isto datoteko
 - ukaz: `ln original trda`



- večkratne trde povezave
 - trda povezava je torej ekvivalent imeniškemu vnosu
 - Katera je original? Kako datoteko zberemo?

Datoteke

- **Bločna in znakovna naprava**
 - dva posebna tipa datotek
 - branje in pisanje takšne datoteke se neposredno nanaša na napravo
 - **bločne** naprave
 - dostop po blokih (npr. 4096 B)
 - **znakovne** naprave
 - dostop po znakih oz. bajtih

Datoteke

- **Cev (pipe) in vtičnica (socket)**
 - mehanizem za medprocesno komunikacijo
 - naslavljanje preko imen datotek
 - nadzor dostopa preko datotečnih dovoljenj
 - **imenovana cev (named pipe)**
 - tudi *FIFO*
 - ustvarimo z `mkfifo`, sicer enak vmesnik kot datoteke
 - **lokalna vtičnica (local socket)**
 - tudi *Unix domain socket*
 - poseben vmesnik
 - BSD knjižnica vtičnic nudi enoten API tako za lokalne kot internetne vtičnice
 - ročica vtičnice je datotečni deskriptor

Kodiranje datotek

- Kodiranje znakov
 - preslikava *zaporedja* bajtov v znake
 - kodiranja iz sorodnih področij
 - Morsejeva koda, Braillova pisava, Baudotjeva koda
 - IBM BCD & EBCDIC
- pomembno predvsem za tekstovne datoteke
 - tekst = neformatirano besedilo

International Morse Code

1. The length of a dot is one unit.
2. A dash is three units.
3. The space between parts of the same letter is one unit.
4. The space between letters is three units.
5. The space between words is seven units.

A	• —	U	• • —
B	— • • •	V	• • • —
C	— • — •	W	— • — •
D	— • •	X	— • • —
E	•	Y	— • — —
F	• • — •	Z	— — • •
G	— — •		
H	• • • •	1	• — — — —
I	• •	2	• • — — —
J	• — — —	3	• • • — —
K	— • —	4	• • • • —
L	• — • •	5	• • • • •
M	— —	6	— • • • •
N	— •	7	— — • • •
O	— — —	8	— — — • •
P	• — — •	9	— — — — •
Q	— — • —	0	— — — — —
R	• — •		
S	• • •		
T	—		

Kodiranje datotek

American Standard Code for
Information Interchange

- Kodiranje ASCII

- izhaja iz telegrafskih kodiranj
- nastalo v 1960'
- 7 bitno kodiranje
 - 95 vidnih znakov
 - 33 ne-vidnih, kontrolnih kod

	0	1	2	3	4	5	6	7
0	NUL	DLE	space	0	@	P	'	p
1	SOH	DC1 XON	!	1	A	Q	a	q
2	STX	DC2	"	2	B	R	b	r
3	ETX	DC3 XOFF	#	3	C	S	c	s
4	EOT	DC4	\$	4	D	T	d	t
5	ENQ	NAK	%	5	E	U	e	u
6	ACK	SYN	&	6	F	V	f	v
7	BEL	ETB	'	7	G	W	g	w
8	BS	CAN	(8	H	X	h	x
9	HT	EM)	9	I	Y	i	y
A	LF	SUB	*	:	J	Z	j	z
B	VT	ESC	+	;	K	[k	{
C	FF	FS	,	<	L	\	l	
D	CR	GS	-	=	M]	m	}
E	SO	RS	.	>	N	^	n	~
F	SI	US	/	?	O	_	o	del

Kodiranje datotek

- Izzivi internacionalizacije

- ASCII oz. US-ASCII

- podpora latinski abecedi in peščici simbolov
 - manjkajo znaki ne-angleških jezikov

- standard YUSCII, 7 bitov

- <https://en.wikipedia.org/wiki/YUSCII>

- ostale razširitve ASCII, 8 bitov

- ISO-IEC 8859-1 (Latin 1),
 - ISO-IEC 8859-2 (Latin 2),
 - Windows CP-1250,
 - ...



koda	ASCII	YUSCII
40	@	Ž
5B	[Š
5C	\	Đ
5D]	Ć
5E	^	Č
60	`	ž
7B	{	š
7C		đ
7D	}	ć
7E	~	č

Kodiranje datotek

- Standard Unicode

- nabor UCS – universal character set
- svetoven repertoar znakov, tudi ČŠŽčšž ☺
- 21-bitni kodni prostor (codespace)
 - kodne točke od U+0000 do U+10FFFF
 - 1.114.112 kodnih točk
 - 2¹¹ rezerviranih (od U+D800 do U+DFFF)
 - možnih torej 1.112.064 kodnih točk
 - razdeljen na 17 ravnin (plane) od 0 do 0x10
 - 0 - osnovna ravnina, ...
- kodirani znaki
 - opisuje 143.859 znakov in pokriva 154 pisav (2020)
 - prvih 256 znakov je identičnih ISO-IEC 8859-1

Kodiranje datotek

- Različna kodiranja
 - UTF – unicode transformation format
 - Kodiranje UTF-32
 - 4 bajti, začetni biti so 0
 - vsaka vrednost predstavlja natančno kodno točko s to vrednostjo
 - prostorsko neučinkovit, zelo redka uporaba
 - Kodiranje UTF-16
 - ena ali dve dvobajtni besedi
 - uporaba znotraj Windows, JVM, CLI, JavaScript, KDE, macOS
 - tudi Microsoft opušča ta način
 - Kodiranje UTF-8
 - Unix-like sistemi, svetovni splet

Kodiranje datotek



- Standard UTF-8
 - razširjen ASCII, se ujema v prvih 128 znakih
 - porabi od 1 do 4 bajtov/znak.
 - podpira samo-sinhronizacijo.

# bajtov	začetna kodna točka	zadnja kodna točka	bajt 1	bajt 2	bajt 3	bajt 4
1	U+0000	U+007F	0xxxxxxx			
2	U+0080	U+07FF	110xxxxx	10xxxxxx		
3	U+0800	U+FFFF	1110xxxx	10xxxxxx	10xxxxxx	
4	U+10000	U+10FFFF	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx

Kodiranje datotek

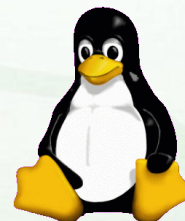
- Skok v novo vrstico
 - kontrolni znaki
 - LF – naslednja vrstica
 - CR – skok na začetek vrstice
 - ipd., glej tudi WP:Newline



- CR+LF (zaradi kompatibilnosti s “tiskalniki”)
 - TENEX, DEC OS (RT-11, OS/8, RSX-11, ..., TOPS-10), CP/M,
 - DOS, OS/2, Windows, Symbian, ARPANET/Internet (telnet, ...)
- LF
 - Multics, Amiga, BeOS, RISC OS, Unix, Unix-like
 - macOS, Linux, FreeBSD
- CR
 - ZX Spectrum, Commodore C64/C128, Mac OS (< v.9), Oberon

Pripenjanje naprav

- Pripenjanje več pomnilnih naprav oz. medijev
 - dostop do datotek v različnih napravah
 - vsaka naprava ima svojo imeniško strukturo
- Ločevanje med datotekami različnih naprav
 - več ločenih imeniških struktur
 - Windows
 - ena enotna imeniška struktura
 - Linux, Unix, macOS



Pripenjanje naprav



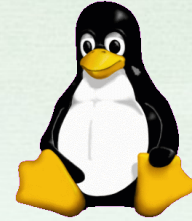
- Več imeniških struktur
 - črkovne oznake naprav
 - disketni enoti: A:, B:, diskovne enote: C:, D:, ...
 - ob pripetju naprave se ji avtomatsko dodeli prosta črka
 - vsaki oznaki ustreza imeniška struktura naprave
 - dostop je preko oznake naprave in poti do datoteke
 - polno ovrednoteno ime
 - vsebuje napravo in absolutno pot
 - npr. C:\Users\Jure\OS-skrivnosti.docx

Pripenjanje naprav



- Enotna imeniška struktura
 - ena imeniška struktura
 - naprave so dostope preko nekaterih imenikov
 - celotno imeniško strukturo dane naprave pripnemo kot pod-strukturo na podani imenik
- Korenski datotečni sistem
 - osnovna imeniška struktura izbrane naprave
 - se priklopi ob zagonu OS na korenski imenik enotne imeniške strukture
 - se pripne ob zagonu OS
 - vsebuje datoteke, pomembne za zagon OS

Pripenjanje naprav



- Pripenjanje (montiranje, mount)
 - pripenjanje dodatnega datotečnega sistema danega pomnilnega medija v obstoječo imeniško strukturo
 - **točka pripenjanja** (mount point)
 - ciljni imenik v obstoječi imeniški strukturi, kamor želimo pripeti dodatni datotečni sistem
 - npr. navadno nek prazen imenik, /mnt/usb
 - točka pripenjanja se **prekrije** s korenskim imenikom dodatnega datotečnega sistema
 - prvotne datoteke za čas priklopa niso dostopne
 - ukaza: `mount` in `umount`

Nadzor dostopa

STROGO ZAUPNO

- Nadzor dostopa do datoteke
 - določa **kdo** lahko **kaj** počne **s čim**
 - kdo? subjekt
 - uporabnik, skupina, proces, območja zaščite
 - npr. lastnik datoteke, skupina datoteke, ostali
 - kaj? dovoljenje
 - dovoljenja za operacije nad datoteko
 - npr. nič, preverjanje obstoja, izvajanje, branje, dodajanje, pisanje, spreminjanje zaščite, brisanje
 - s čim? objekt
 - objekti zaščite so navadno datoteke, vendar lahko tudi naprave, pomnilnik itd.

Nadzor dostopa

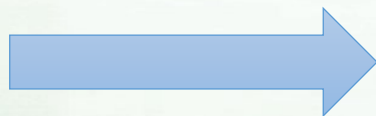
- Matrika dostopa
 - stolpci: objekti nadzora dostopa
 - vrstice: subjekti, ki dostopajo
 - elementi: nabor dovoljenj

	Datoteka 1	Datoteka 2	Datoteka 3	Datoteka 4
Uporabnik A	lastnik R, W	R, X	lastnik R	
Uporabnik B		lastnik R, W, X	R	
Uporabnik C	W			lastnik R

Nadzor dostopa

- Nadzorni seznam dostopa (access control list)
 - dekompozicija matrike dostopa po stolpcih
 - stolpec matrike predstavlja seznam dostopa

	Datoteka 1	Datoteka 2	Datoteka 3	Datoteka 4
Uporabnik A	lastnik R, W	R, X	lastnik R	
Uporabnik B		lastnik R, W, X	R	
Uporabnik C	W			lastnik R



Datoteka 1: (A, lastnik, R, W), (C, W)

Datoteka 2: (A, R, X), (B, lastnik, R, W, X)

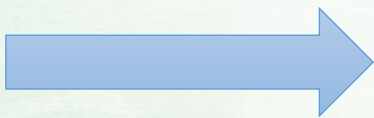
Datoteka 3: (A, lastnik, R), (B, R)

Datoteka 4: (C, lastnik, R)

Nadzor dostopa

- Seznam zmožnosti (capabilities list)
 - dekompozicija matrike dostopa po vrsticah
 - vrstica matrike predstavlja zmožnosti

	Datoteka 1	Datoteka 2	Datoteka 3	Datoteka 4
Uporabnik A	lastnik R, W	R, X	lastnik R	
Uporabnik B		lastnik R, W, X	R	
Uporabnik C	W			lastnik R



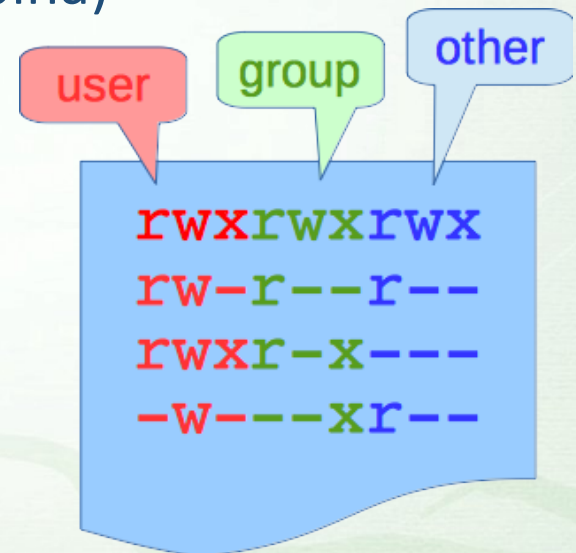
Uporabnik A: (1, lastnik, R, W), (2, R, X), (3, lastnik, R)

Uporabnik B: (2, lastnik, R, W, X), (3, R)

Uporabnik C: (1, W), (4, lastnik, R)

Nadzor dostopa

- Zaščita datotek v Unix / Linux sistemih
 - vrste dovoljenj
 - r – read, w – write
 - x – execute, - – prazno
 - sklopi uporabnikov
 - u – user (lastnik), g – group (skupina)
 - o – other (ostali), a – all (vsi)



Nadzor dostopa

- Zaščita datotek v Unix / Linux sistemih
 - pomen dovoljenj

Obstajajo še posebnosti: setuid, setguid in omejeno brisanje – več o tem na vajah.

znak	datoteka	imenik
r	branje datoteke	izpis vsebine datoteke
w	pisanje oz. spreminjanje datoteke	spreminjanje imenika (ustvarjanje in brisanje datotek v imeniku)
x	izvajanje oz. datoteke je izvršljiva	vstop v imenik (sistemski klic chdir oz. ukaz cd)

