

Operacijski sistemi



Upravljanje z
datotekami

Vsebina

- Logična organizacija podatkov
 - datoteka, imenik itd.
 - hierarhija in naslavljanje datotek
- Datoteke
 - abstrakcija, tipi datotek
- Kodiranje datotek

Logična organizacija podatkov

- **Datoteka**

- osnovna zaključena zbirka podatkov
 - **dolgotrajna obstojnost** (odvisna od pomnilnega medija)
 - jo lahko naslovimo oz. identificiramo preko imena
- **vsebina** datoteke
 - podatki, ki jih datoteka hrani
- **atributi** (meta podatki) datoteke
 - pripadajoči podatki o datoteki

- ime datoteke
- velikost datoteke
- datum in čas nastanka in spremembe datoteke
- lastnik in skupina datoteke
- dovoljenja za uporabo
- ...

Logična organizacija podatkov

- **Imenik**

- omogoča združevanje (sorodnih) datotek
 - vsebuje datoteke in druge imenike oz.
 - vsebuje seznam datotek (in pripadajočih atributov)
- podobnost z datoteko
 - vsebino imenika OS *razume*
 - vsebina je seznam datotek

imenik, mapa, kazalo,
directory, folder



Logična organizacija podatkov

- 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.



Logična organizacija podatkov

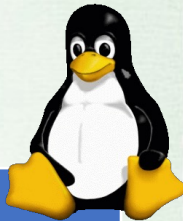
- Naslavljanje datotek
 - **pot** (path) do datoteke
 - seznam imen imenikov od ustreznega izvirnega imenika do dane datoteke
 - **absolutna pot**
 - pot, ki se prične v *korenskem imeniku*
 - **relativna pot**
 - pot, ki se prične v *trenutnem delovnem imeniku*

/	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

Non-file
file types



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

Datoteke

- **Navadna datoteka**

- vsebina datoteke

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

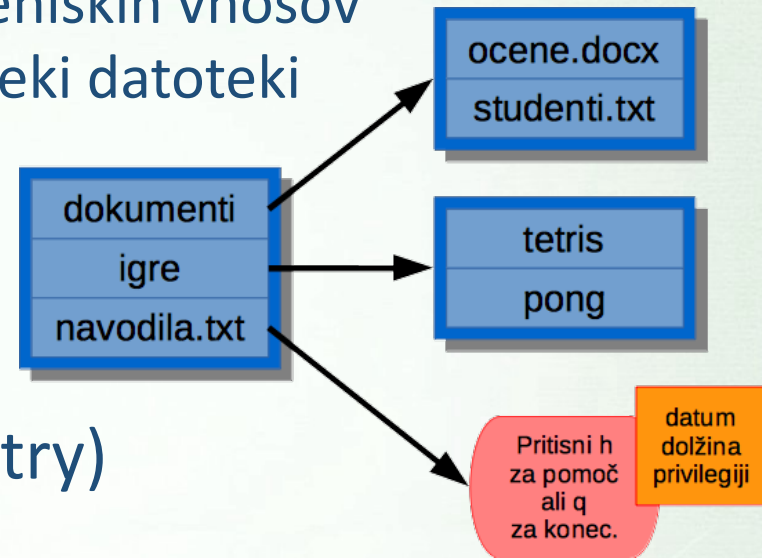
- **vrsta (navadne) datoteke**

- 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

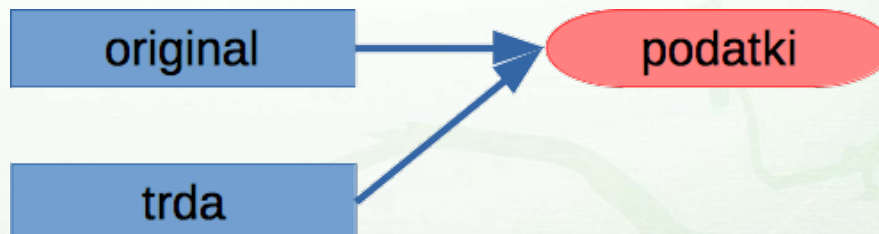
Trda povezava
ni nov tip datoteke.

- Trda povezava

- dodaten **imeniški vnos** (za isto *datoteko*)
 - trda povezava je enakovredna imeniškemu vnosu
 - različni imeniški vnosi lahko predstavljajo isto datoteko (datoteka z več imeni)

- ukaz: `ln original trda`

- Katera je original? Kako datoteko zberemo?



Datoteke

- **Simbolična (mehka) povezava**

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



Datoteke

- **Bločna in znakovna naprava**

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

Datoteke

- **Cev** (pipe) in **vtičnica** (socket)
 - mehanizma za medprocesno komunikacijo
 - naslavljanje preko imen datotek
 - nadzor dostopa preko datotečnih dovoljenj
 - **imenovana cev** (named pipe), *FIFO*
 - ustvarimo z `mkfifo`
 - nudijo enak vmesnik kot datoteke
 - **lokalna vtičnica** (local socket), *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-podobni 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