# GNU/Linux – Obdelava datotek

# *Kodiranje datotek*

- Binarne in tekstovne datoteke.
- Kodiranje ASCII
  - American Standard Code for Information Interchange
  - 1960', 7 bitov
- Razširitve ASCII (8 bitov):
  - ISO-IEC 8859-1 (Latin 1),
  - ISO-IEC 8859-2 (Latin 2),
  - Windows CP-1250,
  - ...

| | 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*

- Standardi Unicode.
  - Nabor UCS – Universal Character Set.
    - Svetoven nabor znakov, tudi ČŠŽčšž.
  - Kodiranja UTF – Unicode Transformation Format.
    - UTF-8, UTF-16, UTF-32.
  - UTF-8 je razširjen ASCII.
    - Porabi od 1 do 4 bajtov/znak.
    - Podpira samo-sinhronizacijo.

# *Kodiranje datotek*

- Znak za novo vrstico.
  - **`LF`** – Multics, Unix, Linux, FreeBSD, Mac OS X;
  - **`CR+LF`** – DEC RT-11, CP/M, DOS, OS/2, Windows, Symbian;
  - **`CR`** – ZX Spectrum, Commodore, Mac OS (< v.9).
- Kontrolna znaka.
  - **`LF`** – naslednja vrstica;
  - **`CR`** – skok na začetek vrstice.

# *Obdelava datotek*

- Izpis vsebine: **`cat`**, **`less`**, **`head`**, **`tail`**.
- Urejanje vrstic: **`sort`**.
- Permutiranje vrstic: **`shuf`**.
- Odstranjevanje duplikatov: **`uniq`**.
- Številčenje vrstic: **`nl`**.
- Obrat vrstic: **`rev`**.
- Spreminjanje znakov: **`tr`**.

# *Obdelava datotek*

- Obdelava stolpcev: **`cut`** in **`paste`**.
- Štetje besed: **`wc`**.
- Primerjava vsebine: **`cmp`**.
- *Sekanje datotek: **`split`**.
- *Ugotavljanje razlik: **`comm`**, **`diff`** in **`patch`**.

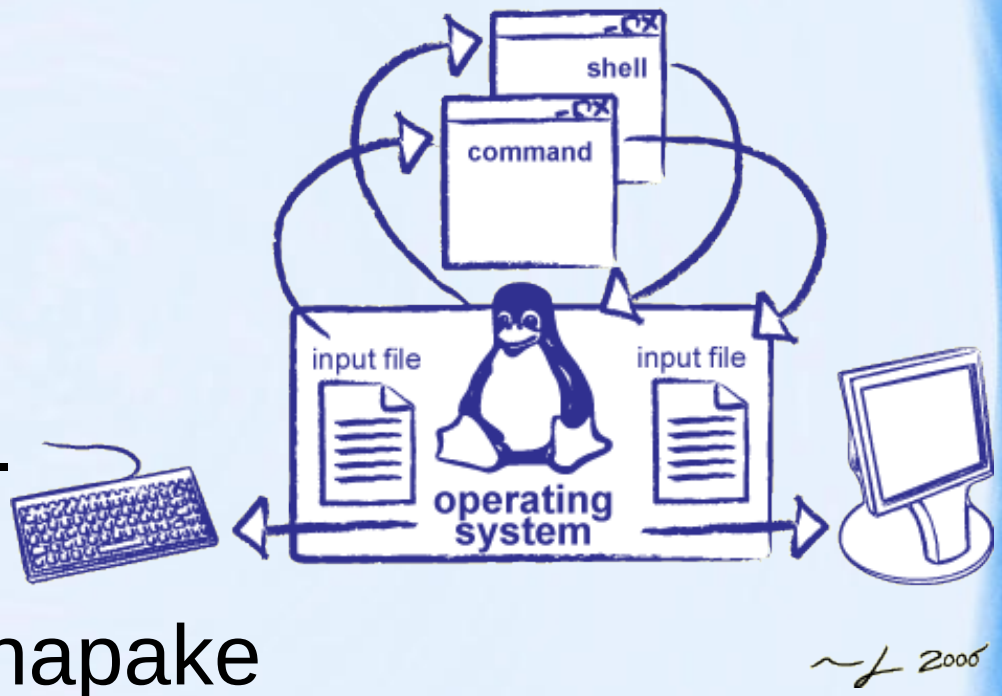
# Skriptiranje v

#!/bin/bash

# Preusmerjanje

# *Preusmerjanje*

- Standardni vhod
  - **`stdin`**, deskriptor 0.
- Standardni izhod
  - **`stdout`**, deskriptor 1.
- Standardni izhod za napake
  - **`stderr`**, deskriptor 2.



# *Preusmerjanje*

- Preusmerjanje standardnega vhoda
  - *ukaz <datoteka*
- Preusmerjanje standardnega izhoda
  - *ukaz >datoteka*
  - *ukaz >>datoteka*

```
ls > spisek.txt
cat spisek.txt
cat < spisek.txt
ls -lp >> spisek.txt
head < spisek.txt
tail < spisek.txt
```

# *Preusmerjanje*

- Splošno preusmerjanje
  - *ukaz deskriptor<datoteka*
  - *ukaz deskriptor>datoteka*
  - *ukaz deskriptor>>datoteka*
  - *ukaz deskriptor1>&deskriptor2*
- Preusmeritev **`stderr`**
  - *ukaz 2>datoteka*
  - *ukaz* **`1>`***datoteka* **`2>&1`**
  - *ukaz* **`&>`***datoteka*

```
mkdir test 2>/dev/null
mkdir test &>/dev/null
```

# *Preusmerjanje*

- Tukaj dokument
  - *ukaz <<ločilo...*
  - *'ločilo'*
  - *-ločilo (začetni tabulatorji)*

```
cat <<'EOF' >skripta.sh
#!/bin/bash
read -p "Vnesi ukaz: " line
echo Vnesel si: $line
exit
EOF
exit
```

- Tukaj niz
  - *ukaz <<< niz*

```
tr "13579" "02468" <<< $(seq 0 9)
```

# *Preusmerjanje*

- *Cevovod*
  - *ukaz1* **|** *ukaz2* **|** *…* **|** *ukazN*
  - Zajem v datoteko **`tee`**.
  - *Izvajanje ukaza z argumenti **`xargs`**.

```
cat /etc/passwd | cut -d: -f7 | sort | uniq -c | sort -gr | head -3
```

# Skriptiranje v


# Regularni izrazi

# *Regularni izrazi*

- Vzorci za opisovanje nizov.
- Osnovni in razširjeni RI.
  - Razširjeni nabor je potrebno posebej vklopiti!
- Iskanje v datotekah: **`grep`** in **`egrep`**.
  - **`egrep`** "*regularni izraz*"
- Urejevalnik toka podatkov: **`sed`**.
  - **`sed -r`** '*naslov***`s`***`/ri/niz/`*'.
- *Ukaz **`awk`**.

# *Regularni izrazi*

- Opis enega znaka.

| ***RI*** | ***Pomen*** | ***Primer*** |
|---|---|---|
| *znak* | opisuje samega sebe | |
| **^** | začetek vrstice | |
| **$** | konec vrstice | |
| **\<** | začetek besede | |
| **\>** | konec besede | |
| **\b** | rob besede | |
| **\B** | ne rob besede | |
| **.** | poljuben znak | |
| **[***znaki***]** | nabor znakov | **[a-dijs-z]** |
| **[^***znaki***]** | komplement nabora znakov | **[^0-9]** |

# *Regularni izrazi*

- Stik in izbira.
  - **`abcd`**, **`a|b|c|d`**, **`00(a|b|c)11`**
- Ponavljanje.

| ***RI*** | ***Št. ponovitev*** | ***Primer*** |
|---|---|---|
| **`?`** | 0 ali 1 | **`ab?`** |
| **`*`** | # >= 0 | **`a*b`** |
| **`+`** | # > 0 | **`a+b+`** |
| **`{`*`n`*`}`** | # == *n* | ***`ab{3}c`*** |
| **`{`*`n`*`,}`** | # >= *n* | ***`a{2,}b`*** |
| **`{`*`n`*`,`*`m`*`}`** | *n* <= # <= *m* | ***`ab{2,4}c`*** |