

Systemy operacyjne

Mariusz Żynel

`mariusz@math.uwb.edu.pl`

`http://math.uwb.edu.pl/~mariusz/`

Uniwersytet w Białymstoku

2023/2024

Wyrażenia regularne i programy narzędziowe

- Wyrażenia regularne (regular expressions, RE, REGEX)
- Regular sets (Stephen Cole Kleene 1951), języki regularne i formalne
- PCRE (Perl, PHP, Apache), TCL Advanced Regular Expressions
- Podstawowe wyrażenia regularne
- Ograniczone wyrażenia regularne, programy: sed, grep
- Pełne wyrażenia regularne, programy: egrep, awk, perl
- Użyteczne narzędzia: find, fgrep, diff

Metaznaki w shellu

Przy pomocy metaznaków (wildcard characters) możemy tworzyć wzorce (glob patterns) nazw plików. Shell `sh` rozumie następujące metaznaki:

* zastępuje dowolny ciąg znaków, także pusty

? zastępuje jeden dowolny znak

[abc] zastępuje jeden z wymienionych znaków

[!abc] zastępuje jeden znak spoza wymienionych znaków

[a-z] zastępuje jeden znak z podanego zakresu

[!a-z] zastępuje jeden znak spoza podanego zakresu

Uwaga

Znak `/` oraz `.` jeśli występuje na początku nazwy pliku nigdy nie zostaną zastąpione.

Podstawowe wyrażenia regularne, znaki specjalne

Znaki specjalne w podstawowych wyrażeniach regularnych: \ . [] * ^ \$

\ znak chroniący, pozbawia znak specjalnego znaczenia

. jeden dowolny znak

[abc] jeden z wyszczególnionych znaków

[^abc] jeden znak spoza wyszczególnionych znaków

[a-z] jeden znak z podanego zakresu

[^a-z] jeden znak spoza podanego zakresu

* zero lub więcej wystąpień wyrażenia poprzedzającego

^ początek linii

\$ koniec linii

Podstawowe wyrażenia regularne, przykłady

- `Unix`
- `[Uu]nix`
- `...\.`
- `*.*`
- `2018-[01][0-9]-[0-3][0-9]`
- `^#.*$`
- `^class`
- `[aiowz]$`
- `img[^0-9]\.png`
- `[a-z-..]*@[a-z-..]*\.[a-z]*`
- `[A-Z][A-Z]*`

Znaki specjalne w ograniczonych wyrażeniach regularnych: $\{ \} ()$

$\{m\}$ dokładnie m wystąpień wyrażenia poprzedzającego

$\{m, \}$ co najmniej m wystąpień wyrażenia poprzedzającego

$\{m, n\}$ od m do n wystąpień wyrażenia poprzedzającego

$(...)$ podwyrażenie regularne, odwołania poprzez $\backslash 1, \backslash 2, \dots$

$(...(...)...)$ możliwe zagnieżdżenia podwyrażeń regularnych

Ograniczone wyrażenia regularne, przykłady

- $\text{Un}\{2\}ix$
- $\text{Un}\{2,\}ix$
- $\text{Un}\{2,5\}ix$
- $[0-9]\{1,3\}\backslash.[0-9]\{1,3\}\backslash.[0-9]\{1,3\}\backslash.[0-9]\{1,3\}$
- $[a-z-..]\{3,\}@[a-z-..]\{2,\}\backslash.[a-z]\{2,\}$
- $[0-9]\{3\}-[0-9]\{3\}-[0-9]\{3\}$
- $[1-9][0-9]\{2\}-[0-9]\{3\}-[0-9]\{3\}$
- $(\text{Unix}) \backslash 1$
- $([ABCDE]\{3\}) ([abcde]\{3\}) \backslash 2 \backslash 1$
- $([0-9]\{4\}) - ([0-9]\{2\}) - ([0-9]\{2\})$
- $(([0-9]\{4\}) - [0-9]\{2\}) - [0-9]\{2\}$

Znaki specjalne w pełnych wyrażeniach regularnych: +?|

- + jedno lub więcej wystąpień wyrażenia poprzedzającego
- ? brak lub jedno wystąpienie wyrażenia poprzedzającego
- | alternatywa, jedno z dwóch rozdzielonych wyrażień

Pełne wyrażenia regularne, przykłady

- `Unix files?`
- `Unix files+`
- `Unix (file)?`
- `(Unix|DOS) file`
- `[]+$`
- `"[^"]+"`
- `img-[0-9]{3}\.(jpg|png|tiff|gif)`
- `[0-9]+(EUR|PLN)`
- `[-+]?[0-9]+\.[0-9]+`
- `[-+]?[0-9]*\.[0-9]+`
- `\$[A-Za-z0-9_-]+`