

Operačné systémy

Seminár - Regulárne výrazy

Ing. Martin Vojtko, PhD.

Fakulta informatiky a informačných technológií
STU v Bratislave

2021/2022

RegEx - regulárne výrazy

- Existuje niekoľko štandardov:
 - POSIX definuje basic (BRE) a extended (ERE)
 - ERE rozširuje BRE o ? + a |.
 - **grep** používa prepínače -G BRE (default) a -E pre ERE
 - **sed** používa prepínač -E pre ERE
 - Perl de facto štandard PCRE (Perl Compatible RE)
 - ? umožňuje použitie lazy match
 - **grep** používa -P pre PCRE

RegEx meta znaky nahrádzania

- `.` - ľubovoľný znak
- `[]` - uvedené znaky
- `[^]` - všetky znaky okrem uvedených znakov

RegEx znakové triedy

- `[[:alnum:]]` - `[A-Za-z0-9]`- alfa-numericke znaky
- `[[:alpha:]]` - `[A-Za-z]` - znaky abecedy
- `[[:lower:]]` - `[a-z]`
- `[[:upper:]]` - `[A-Z]`
- `[[:digit:]]` - `[0-9]`
- `[[:xdigit:]]` - `[A-Fa-f0-9]`
- `[[:blank:]]` - `[\t]` - medzera a tab
- `[[:space:]]` - `[\t\r\n\v\f]` - biele znaky
- `[[:punct:]]` - `[!\"#$%&'()*+,-./:;<=>?@\^_`{|}~]`
- `[[:cntrl:]]` - `[\x00-\x1F\x7F]` - control znaky
- `[[:graph:]]` - `[\x21-\x7E]` - viditeľné znaky

RegEx meta znaky opakovania

- $*$ - 0 a viac opakovaní predchádzajúcej položky
- $+$ - 1 a viac opakovaní predchádzajúcej položky
- $?$ - 0 alebo 1 opakovaní predchádzajúcej položky
- $\{n\}$ - n opakovaní predchádzajúcej položky
- $\{n, \}$ - n a viac opakovaní predchádzajúcej položky
- $\{n, m\}$ - n až m opakovaní predchádzajúcej položky
- $\{, m\}$ - 0 až m opakovaní predchádzajúcej položky
- $?$ - lazy match ak použité po opakujúcom znaku. nutný Perl

RegEx meta znaky kotvy

- `^` - začiatok riadku
- `$` - koniec riadku
- `\<` - začiatok slova
- `\>` - koniec slova
- `\b` - začiatok alebo koniec slova

RegEx znaky skupín

- `()` - označenie podskupiny znakov výrazu
- `|` - or, logický súčet dvoch výrazov. nutné ERE

RegEx zmena významu znaku

- \ - ruší význam špeciálnych znakov a pridáva význam normálnym znakom
- BRE vyžaduje použitie \ pred +, {, }, |, (a) ak ich chceme použiť ako meta znaky.
- ERE vyžaduje použitie \ pred +, {, }, |, (a) ak ich chceme použiť ako obyčajné znaky.