

Praktická aplikácia regulárnych výrazov. Problémy a ťažkosti pri ich písaní a používaní.*

Denis Danilov

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií
`xdanilovd@stuba.sk`

5. október 2023

Abstrakt

Regulárne výrazy, často známe ako regex, je formálny jazyk, ktorý sa používa na priradovanie textových vzorov. Používanie regexov pomáha rýchlo a presne nájsť a extrahovať kľúčové informácie z veľkých objemov textu. Môžu Vám doslova pomôcť nájsť ihlu v kope sena, čo organizáciám umožňuje presmerovať svoje ľudské zdroje do iných oblastí. Regexy sú obzvlášť užitočné na definovanie filtrov, pretože obsahujú postupnosť znakov, ktoré definujú vzor textu, ktorý sa má zhodovať.

Cieľom tejto práce je preskúmať, ako sa regulárne výrazy často používajú v praxi, identifikovať problémy a ťažkosti, s ktorými sa programátori stretávajú pri písaní a používaní regulárnych výrazov.

Metódami na dosiahnutie cieľa bude analýza prieskumu medzi vývojármi a údajov o používaní regulárnych výrazov v otvorených projektoch na GitHub.

Tento článok jasne ukáže výhody používania regulárnych výrazov, ako aj osvedčené postupy pri ich písaní.

1 Úvod

Regulárne výrazy, tiež známe ako regulárne výrazy, sú nástrojom na spracovanie textu, ktorý je integrovaný do všetkých súčasných programovacích jazykov a bežne používaných utilít, ako sú textové editory. Regulárne výrazy sa používajú pri úlohách vyhľadávania a nahrádzania reťazcov, ako je vyhľadávanie slov, úprava textu, analýza súborov. [1] Odhady naznačujú, že viac ako tretina projektov JavaScript a Python obsahuje aspoň jeden regulárny výraz. [2] Tento článok je určený na to, aby vám pomohol pochopiť a efektívne používať regulárne výrazy. Začneme základnými pojmami a postupne prejdeme k pokročilejším príkladom.

Regulárne výrazy sú široko používané v rôznych oblastiach výpočtovej techniky. Tu je niekoľko bežných prípadov použitia:

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2023/24, vedenie: PaedDr. Pavol Baťalík

1. Analýza protokolov: Regulárne výrazy možno použiť na extrahovanie špecifických častí informácií zo systémových alebo aplikačných protokolov.
2. Overenie údajov: Môžu sa použiť na overenie formátu vstupných údajov, ako sú e-mailové adresy, telefónne čísla a ID používateľov.
3. Hľadať a nahradiť: Regulárne výrazy možno použiť na nájdenie a nahradenie konkrétnych vzorov v texte alebo kódovej základni.
4. Premenovanie súborov: Môžu sa použiť na dávkové premenovanie súborov na základe určitých vzorov.
5. Analýza vstupu používateľa: Regulárne výrazy možno použiť na analýzu vstupu používateľa alebo príkazových riadkov.
6. Čítanie konfiguračných súborov: Môžu sa použiť na čítanie a analýzu konfiguračných súborov.

2 Záver

Literatúra

- [1] Carl Chapman, Peipei Wang, and Kathryn T. Stolee. Exploring regular expression comprehension. In *2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, October 2017.
- [2] Louis G. Michael, James Donohue, James C. Davis, Dongyoon Lee, and Francisco Servant. Regexes are hard: Decision-making, difficulties, and risks in programming regular expressions. In *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, November 2019.