

Úlohy 8 (18.4.2018)

Termín: **26. 4. 2018**, vo štvrtok 13:00–10 v H6, alebo cvičiacim, resp. na prednáške . Každý príklad píšete na samostatný papier A4! Nezabudnite sa podpísať a uviesť skupinu kam chodíte na cvičenia (meno cvičiaceho resp. čas cvičenia a miestnosť). Ak máte riešenie jedného príkladu na viacerých papieroch, vhodne ich zopnite.

Vždy uveďte aj zdôvodnenie (dôkaz) vášho riešenia! (Nestačí len áno/nie alebo číslo.)

Opísané riešenia sú za 0b (aj opisované aj opísané, nebudeme zisťovať čo je originál).

Nezabudnite, pri každom TS uviesť v bodoch aj hlavnú myšlienku ako pracuje.

príklad 1

Navrhnite TS, ktorý rozpoznáva jazyk

$$L = \{w \in \{a, b\}^* \mid 3|w|_a = 2|w|_b\}.$$

príklad 2

Na páske je slovo $\zeta x \# y$, kde $x, y \in \{0\} \cup \{1, 2\}^+ \{0, 1, 2\}^*$ a x a y interpretujeme ako čísla v trojkovej sústave - najvyššie rády sú naľavo. Navrhnite TS, ktorý akceptuje práve vtedy keď $x \geq y$.

Príklad 3

Tento príklad je dobrovoľný, za 10b. Predstavte si, že ste dostali veľký textový súbor, z ktorého potrebujete vyfiltrovať webové adresy. Pamätáte si, že kedysi ste počuli o regulárnych výrazoch, tak sa rozhodnete použiť knižnicu RE.

- (a) Najprv treba zistiť, koľko rôznych korektných webstránok sa tam nachádza (za korektnú webstránku považujeme všetko, čo sa začína `www` alebo `http://` alebo `https://`, má v názve aspoň tri znaky, a za bodkou dva až štyri znaky. Nezabudnite, že niektoré webové adresy majú v sebe aj podpriechinky, či adresy typu `/~something`).

V "`www.google.com ww.uniba.sk moja.stranka https://uniba.sk/courses`" sú iba dve korektné stránky.

- (b) Keďže dakedy nás zaujímajú len hlavné adresy, zistíte, koľko rôznych domén druhého rádu sa tam nachádza a vypíšete ich.

Napr. v "`www.edi.fmph.uniba.sk www.uniba.sk/fakulty/ www.euba.sk www.fri.euba.sk/studium www.fns.uniba.sk`" sa nachádzajú zjavne iba dve domény - `uniba.sk` a `euba.sk`

- (c) Zistili ste, že `uniba.sk` má v súbore veľa subdomén, chcete zistiť presne koľko a ktoré.

Napr "`www.edi.fmph.uniba.sk https://fns.uniba.sk www.fphil.uniba.sk`" obsahuje 3 – `fmph`, `fns` a `fphil`

- (d) Na koniec chcete zistiť, koľko zo stránok je osobných, teda sú v tvare `korektna_adresa/~mrkvicka`.

Pozor, zjavne obe adresy v zozname "`http://edi.fmph.uniba.sk/~mrkvicka edi.davinci.fmph.uniba.sk/~mrkvicka`" odkazujú na to isté.

Úlohu riešte pomocou knižnice RE v Pythone(tutoriál). Prípadne v JAVE (tutoriál z prednášky a cvičení Programovania 4). Vzorový súbor nájdete TU. Stringy sú v ňom vo viacerých riadkoch oddelené medzerou, môžete predpokladať, že dostanete dobrý vstup. Po načítaní vstupu program na štandardný výstup (konzolu) vypíše odpovede pre každú z podotázok označenú príslušným písmenom.

Riešenie posielajte ako súbor s vašim priezviskom v názve na [lucia.budinska\(at\)fmph.uniba.sk](mailto:lucia.budinska@fmph.uniba.sk). Nezapudnite svoj kód dobre okomentovať.