Jazyková inkluze regulárních výrazů s čítacími operátory #33

Excel@FIT 2020

Definice problému

Buďte R_1,R_2 regulární výrazy. Problém jazykové inkluze regulárních výrazů R_1 a R_2 je rozhodnutí zda platí $\mathcal{L}(R_1) \subseteq \mathcal{L}(R_2)$. Kde syntax regulární výrazů je:

$$R ::= \emptyset \mid \varepsilon \mid \sigma \mid R_1 R_2 \mid R_1 + R_2 \mid R^* \mid \sigma\{m,n\}.$$

Existující řešení

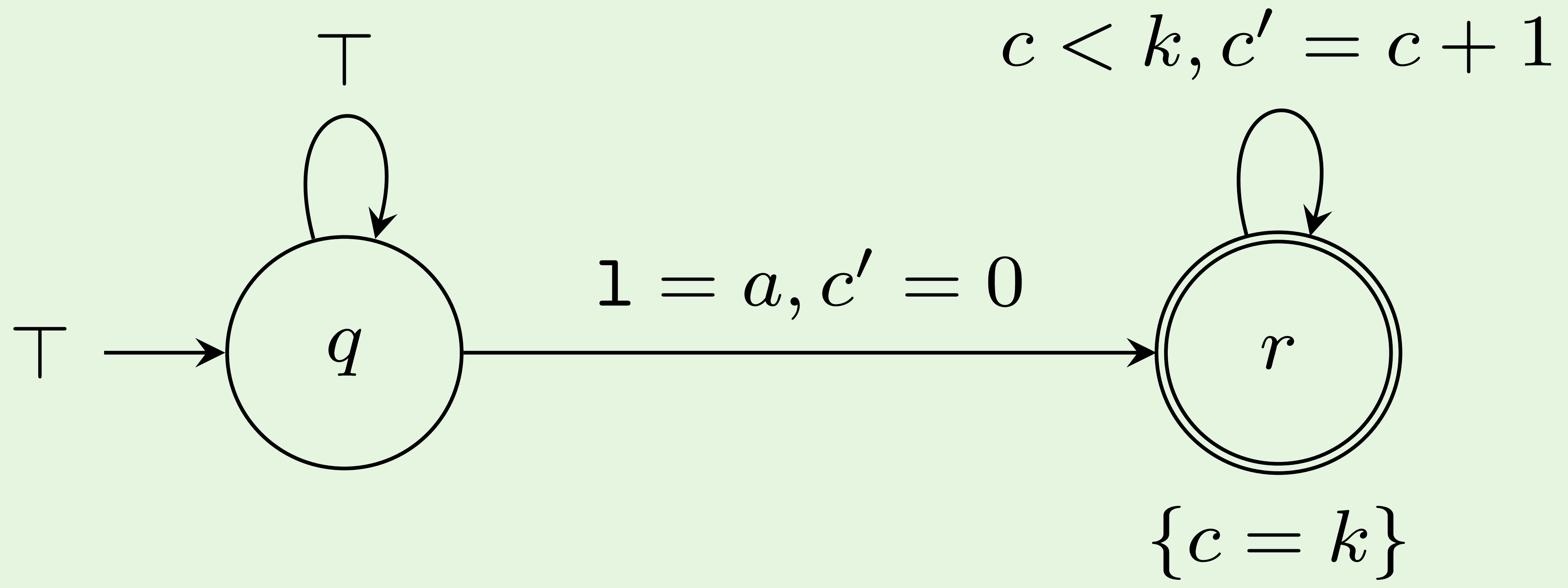
Transformace R_1 a R_2 na nedeterministické konečné automaty (NFAs) N_1 a N_2. Následně použijeme metody pro testovaní jazykové inkluze dvou NFAs. Tyto metody jsou založené na ekvivalenci:

$$\mathcal{L}(N_1) \subseteq \mathcal{L}(N_2) \iff \mathcal{L}(N_1) \cap \mathcal{L}(\overline{N_2}) = \emptyset.$$

NAŠE řešení

Místo NFAs použijeme tzv. monadické čítací automaty (MCAs) [1]—transformujeme R_1 a R_2 na MCAs M_1 a M_2 a kopírujeme výše uvedený postup pro NFAs. Aby jsme tento postup mohli aplikovat musíme najít odpovědi na následující otázky:

- jak vypočítat komplement automatu M_2?
  $\Longrightarrow$ [1] poskytuje algoritmus pro determinizaci MCAs
- jak vypočítat průnik automatů M_1 a $\overline{M_2}$?
  $\Longrightarrow$ analogicky jak průnik dvou NFAs
- jak testovat prázdnost automatu $M_1 \times \overline{M_2}$?
  $\Longrightarrow$ těžší něž v případě NFAs, protože jestli čítací automat může provést přechod nezávisí pouze na vstupním symbolu, ale i na konfiguraci čítačů



Obrázek 1. Ekvivalentní monadický čítací automat M pro regulární výraz $.^*\mathtt{a}.\{\mathtt{k}\}$ kde $\mathtt{k} \in \mathbb{N}$.

Výhody

MCAs kompaktně reprezentují regulární výrazy s výše uvedenou syntaxí
$\Longrightarrow$ vygenerované MCAs jsou obvykle menší než NFAs
$\Longrightarrow$ to nám dovoluje testovat některé regulární výrazy rychleji než původní metodou

Literatura

[1] L. Holík, O. Lengál, O. Saarikivi, L. Turoňová, M. Veanes, T. Vojnar: Succinct determinisation of counting automata via sphere construction. In APLAS'19.

Autor: David Mikšaník (xmiksa05@stud.fit.vutbr.cz)
Vedoucí práce: Ing. Ondřej Lengál, Ph.D
BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY