Szegedi Tudományegyetem

Informatikai Intézet

Reguláris kifejezések ekvivalenciája

Szakdolgozat

Készítette:
Hajagos Károly
programtervező informatikus BSc
szakos hallgató

Témavezető:
Dr. Fülöp Zoltán
egyetemi tanár

Szeged
2017

Tartalomjegyzék

Feladatkiírás

Két reguláris kifejezést ekvivalensnek mondunk, ha az általuk meghatározott nyelvek megegyeznek. Ismert tény, hogy nem adható meg véges sok olyan azonosság (véges axiómarendszer), amelyekből csupán az "egyenlőség szabály" alkalmazásával eldönthető, hogy két reguláris kifejezés ekvivalens-e. Ugyanakkor, A. Salomaa [2] cikkében megadott egy olyan, tizenhárom axiómából és négy következtetési szabályból álló formális bizonyítási rendszert, amely helyes és teljes a reguláris kifejezések ekvivalenciájának bizonyítására vonatkozóan. A bizonyítási rendszer hallgatók számára is érthető, részletesen ismertetésre került az [1] kézikönyvben. A hallgató feladata megismerkedni a reguláris kifejezések azonosságaival és Salomaa formális rendszerével és algoritmusával, továbbá, a rendszer helyességének és teljességének megmutatása és a formális bizonyítás részleteinek kidolgozása.

[1] Dan A. Simovici, R. L. Tenney, Theory of Formal Languages with Applications, World Scientific, 1999.

[2] A. Salomaa, Two complete axiom systems for the algebra of regular events, J. ACM, 13 (1966) 158-169.

Tartalmi összefoglaló

A tartalmi összefoglalónak tartalmaznia kell (rövid, legfeljebb egy oldalas, összefüggő megfogalmazásban) a következőket: a téma megnevezése, a megadott feladat megfogalmazása - a feladatkiíráshoz viszonyítva-, a megoldási mód, az alkalmazott eszközök, módszerek, az elért eredmények, kulcsszavak (4-6 darab).

Az összefoglaló nyelvének meg kell egyeznie a dolgozat nyelvével. Ha a dolgozat idegen nyelven készül, magyar nyelvű tartalmi összefoglaló készítése is kötelező (külön lapon), melynek terjedelmét a TVSZ szabályozza.

Bevezetés

Itt kezdődik a bevezetés, mely nem kap sorszámot.

1. fejezet

Általános fogalmak és jelölések

Ebben a fejezetben azokról az eszközökről lesz szó, amelyek a későbbi definíciók, tételek és példafeladatok építőkövei. Bevezetjük az ábécéket és az azokból képezhető szavakat, nyelveket, továbbá definiálunk különféle függvényeket, melyek segítenek az egyes nyelvi algoritmusok leírásában.

1.1. Általános jelölések

Tetszőleges H véges halmaz esetén $|H|$-val jelöljük a H elemeinek a számát.

1.2. Ábécék és szavak

Szimbólumok egy véges, nemüres halmazát ábécének nevezzük. A továbbiakban az ábécét általában Σ-val jelöljük, elemeit pedig betűknek is hívjuk. Egy $a_1 \dots a_k$ alakú sorozatot, ahol $k \geq 0$ és $a_1, \dots, a_k \in \Sigma$, a Σ ábécé feletti szónak (sztringnek) nevezzünk. Abban az esetben ha $k = 0$, az üres szót kapjuk, melynek jele λ. Az ábécé tehát karakterek halmaza, melyekből szavakat alkothatunk, ezenfelül keretet szab a szóalkotáskor felhasználható karakterek számát illetően.
Ha például $\Sigma = \{a, b\}$, akkor $\lambda, a, b, aa, ab, bba, ababa, \dots$ stb Σ ábécé feletti szavak. Látható, hogy akár egyelemű ábécé esetén is képezhető végtelen számú szó.

A továbbiakban az $a, b, c...$ szimbólumokkal az ábécé betűit, az $...u, v, w, x, y, z$ szimbólumokkal pedig az ábécé feletti szavakat jelöljük.

Legyenek u, v szavak Σ felett. Ekkor az uv szó az u és v konkatenációja, vagyis összeláncolása. A konkatenáció asszociatív, hiszen $(uv)w = u(vw)$, de nem kommutatív. Az u n-edik hatványán az $u^n = u_1 \cdots u_n$ szót értjük, ahol $u_1 = u_2 = \ldots = u_n = u$. Ha $n = 0$, akkor $u^n = \lambda$.
Például, ha $u = abb$ és $v = bab$, akkor $uv = aabbab$ és $u^3 = uuu = abbabbabb$.

Egy w szó hosszán az őt alkotó betűk multiplicitással vett számát értjük, melyet így jelölünk: $|w|$. Formálisan: ha $w = \lambda$, akkor $|w| = 0$, különben, ha $w = av$, akkor $|w| = = 1 + |v|$.

Az összes Σ feletti szavak halmazát Σ^*-gal jelöljük, továbbá $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$. Tehát

$$\Sigma^* = \{a_1 \dots a_k \mid k \geq 0, a_1, \dots, a_k \in \Sigma\} \text{ és } \Sigma^+ = \{a_1 \dots a_k \mid k \geq 1, a_1, \dots, a_k \in \Sigma\}.$$

Legyen $\Sigma = \{a, b\}$, ekkor

$$\Sigma^* = \{\lambda, a, b, aa, bb, ab, ba, aaa, bbb, aab...\},$$
$$\Sigma^+ = \{a, b, aa, bb, ab, ba, aaa, bbb, aab...\}.$$

A w szóban található a betűk számát $\mathrm{n}_a(w)$ jelöli. $w = abba$ esetén $\mathrm{n}_a(w) = 2$.

Egy w szó prefixe minden olyan u szó, amelyhez van olyan v, hogy $w = uv$. Továbbá w szuffixe minden olyan u szó, amelyhez van olyan v, hogy $w = vu$. A w szó összes prefixének halmazát $\mathrm{pre}(w)$, az összes szuffixének halmazát $\mathrm{suf}(w)$ jelöli. Tehát

$$\mathrm{pre}(w) = \{u \in \Sigma^* \mid \exists(v \in \Sigma^*) : w = uv\},$$
$$\mathrm{suf}(w) = \{v \in \Sigma^* \mid \exists(u \in \Sigma^*) : w = uv\}.$$

Nyilvánvaló, hogy $|\mathrm{pre}(w)| = |\mathrm{suf}(w)|$.
Legyen $u = abaa$, ekkor $\mathrm{pre}(u) = \{\lambda, a, ab, aba, abaa\}$, illetve $\mathrm{suf}(u) = \{abaa, baa, aa, a, \lambda\}$.
Egy w szó valódi prefixe minden olyan $u \in \mathrm{pre}(w)$ szó, amelyre $u \neq \lambda$ és $u \neq w$. A w valódi szuffixeit hasonló módon definiáljuk.

1.3. Nyelvek

Σ feletti nyelvnek nevezzük Σ^* tetszőleges részhalmazát. Például, $\Sigma = \{a, b\}$ feletti $\{a, ba, aa\}$ nyelv véges, míg az ugyancsak Σ feletti $\{a^n \mid n \geq 0\}$ és $\{a^n b^n \mid n \geq 0\}$ nyelvek végtelenek. A továbbiakban egy nyelvet általában L-lel jelölünk.
A Σ feletti nyelvek halmaza tartalmazza az üres nyelvet ($\emptyset$), a teljes nyelvet (Σ^*) és az egység nyelvet, azaz $\{\lambda\}$-t. Egy L nyelv λ-mentes, ha $\lambda \notin L$. Ha L λ-mentes, akkor $L \subseteq \Sigma^+$.
Legyen $\Sigma = \{a, b\}$ és legyen L azon három hosszúságú szavak halmaza, melyeknek középső betűje különbözik a többi betűtől. Ekkor $L = \{aba, bab\} \subseteq \Sigma^+$, ezért az is igaz, hogy $L = \{aba, bab\} \subseteq \Sigma^*$.
Az összes Σ feletti nyelvek halmaza $\mathcal{P}(\Sigma^*)$, vagyis Σ^* összes részhalmazainak halmaza.

Legyen $L_1, L_2 \subseteq \Sigma^*$, ekkor a halmazelméleti műveletek: $L_1 \cup L_2$, $L_1 \cap L_2$, $L_1 \setminus L_2$ és az L_1 komplementere: $\overline{L_1} = \Sigma^* \setminus L_1$. Továbbá, L_1 és L_2 konkatenációja, L_1 iterációja és L_1 λ-mentes iterációja sorrendben:

$$L_1 L_2 = \{uv \mid u \in L_1, v \in L_2\},$$
$$L_1^* = \{\lambda\} \cup L_1 \cup L_1 L_1 \cup L_1 L_1 L_1 \cup \ldots$$
$$L_1^+ = L_1 \cup L_1 L_1 \cup L_1 L_1 L_1 \cup \ldots$$

Legyen $\Sigma = \{a, b\}$, $L_1 = \{ab, bb, bab\} \subseteq \Sigma^*$ és $L_2 = \{bb, aab, bab\} \subseteq \Sigma^*$. Ekkor

- $L_1 \cup L_2 = \{ab, bb, bab, aab\}$;
- $L_1 \cap L_2 = \{bb, bab\}$;
- $\overline{L_1} = \{\lambda, a, b, aa, ba, aaa, aba, baa, aab, abb, bbb, bba, \ldots\}$;
- $L_1 L_2 = \{abbb, abaab, abbab, bbbb, bbaab, bbbab, babbb, babaab, babbab\}$

– $L_1^3 = L_1L_1L_1 = \{ababab, ababbb, ababbab, stb...\}$, továbbá $L_1^0 = \{\lambda\}$;

– $L_1^* = \{\lambda, ab, bb, bab, abab, abbb, abbab, bbab, bbbb, bbbab, babab, \ldots\}$;

– $L_1^+ = \{ab, bb, bab, abab, abbb, abbab, bbab, bbbb, bbbab, babab, \ldots\}$.

Az eddigiekből következik, hogy minden L nyelvre $\lambda \in L^*$, és $L^* = \{\lambda\} \cup L^+$, illetve $\emptyset^* = \{\lambda\}$. Az $\cup$, $\cap$ és a komplementer műveleteket Boole műveleteknek, míg az $\cup$, konkatenáció és iteráció műveleteket reguláris műveleteknek nevezzük.

Legyen L, K két nyelv Σ ábécé felett. Ekkor a jobb és bal hányados sorrendben a következő:

$$LK^{-1} = \{u \in \Sigma^* \mid uv \in L \text{ és } v \in K\}$$
$$K^{-1}L = \{u \in \Sigma^* \mid vu \in L \text{ és } v \in K\}$$

Legyen $a, b, c \in \Sigma$, továbbá $L = \{\lambda, a, ba, bca\}$, $K_1 = \{a, b\}$ és $K_2 = \{b, a\}^*$ pedig Σ feletti nyelvek. Ekkor

$$LK_1^{-1} = \{\lambda, b, bc\}$$
$$LK_2^{-1} = \{\lambda, a, b, bca\}$$
$$K_1^{-1}L = \{\lambda, a, ca\}$$

1.1. Lemma. *Legyenek L_1 és L_2 tetszőleges nyelvek Σ felett. Ha $\lambda \notin L_2$, akkor az $X = L_1 \cup L_2X$ egyenletnek az $L_2^*L_1$ nyelv az egyetlen megoldása.*

Bizonyítás. Vezessük be az $L = L_2^*L_1$ rövidítést. Könnyű ellenőrizni, hogy $L = L_1 \cup L_2L$, tehát L megoldása az egyenletnek. Tegyük fel, hogy egy másik $L' \neq L$ nyelv is megoldása ugyanannak az egyenletnek, tehát $L' = L_1 \cup L_2L'$ is teljesül. Ekkor az $L \setminus L'$ és az $L' \setminus L$ halmazokból legalább az egyik nemüres. Feltesszük, hogy $L' \setminus L \neq \emptyset$, és azt, hogy w a minimális hosszúságú szó az $L' \setminus L$ halmazban. Világos, hogy $w \notin L_1$, mert különben a $w \in L$ tartalmazás is teljesülne. Ugyanakkor $L' = L_1 \cup L_2L'$, tehát $w \in L_2L'$. Ezért $w = uv$, ahol $u \in L_2$ és $v \in L'$. Mivel $\lambda \notin L_2$, $u \neq \lambda$, tehát $|v| < |w|$. Az is világos, hogy $v \notin L$, hiszen ekkor igaz lenne, hogy $w = uv \in L_2L \subseteq L$. Tehát $v \in L' \setminus L$, ez azonban ellentmond annak, hogy w a minimális hosszúságú szó az $L' \setminus L$ halmazban. Az $L \setminus L' \neq \emptyset$ eset ezzel szimmetrikusan bizonyítható. □

1.4. Helyettesítések és morfizmusok

2. fejezet

Formális következtetési rendszer reguláris kifejezésekre

Ebben a fejezetben megadunk egy formális következtetési rendszert, amely egy deduktív megközelítést ad reguláris kifejezések ekvivalenciájára vonatkozóan. A rendszer tizenhárom axiómából és négy következtetési szabályból áll. A rendszerrel $R_1 \sim R_2$ alakú elemeket következtethetünk (vagy: bizonyíthatunk be), ahol R_1 és R_2 reguláris kifejezések. Bebizonyítjuk, hogy a rendszer helyes és teljes a reguláris kifejezések ekvivalenciájára nézve: bármely két R_1 és R_2 reguláris kifejezésre teljesül, hogy $R_1 \sim R_2$ akkor és csak akkor bizonyítható, ha $R_1 \equiv R_2$. A rendszert Arto Salomaa prezentálta az 1966-ban megjelent [2] cikkében.

2.1. Definíció. *Legyenek R_1, R_2 reguláris kifejezések. Ha R_1' és R_2' helyettesítő példányai az R_1 és R_2 kifejezéseknek, akkor az $R_1 \sim R_2$ helyettesítő példánya: $R_1' \sim R_2'$.*

A helyettesítés két ábécé esetén egy leképezés (más szóval függvény), melynek során az első ábécé minden elemének egy, a másik ábécé feletti nyelvet feleltetünk meg. Tehát a helyettesítés egy $s : \Sigma \to \mathcal{P}(\Theta^*)$ leképezés, ahol Σ és Θ tetszőleges ábécék.

Irodalomjegyzék

[1] Dan A. Simovici, R. L. Tenney, Theory of Formal Languages with Applications, World Scientific, 1999.

[2] A. Salomaa, Two complete axiom systems for the algebra of regular events, J. ACM, 13 (1966) 158-169.