0. príklad - opakovanie

(a) Koľko je postupností nad $\Sigma = \{a, b\}$ dĺžky 10?

(b) Koľko z nich obsahuje rovnaký počet a a b?

(c) Koľko slov w nad touto abecedou, $|w| \leq 10$, obsahuje podslovo ab?

  (i) Chceme generovať takéto slová. Pomocou akých (*dvoch*) pravidiel vznikajú?

  (ii) Aký je všeobecný vzťah?

  (iii) Koľko prvkov musíme definovať, aby sme vedeli tento (rekurentný) vzťah použiť?

Riešenie 0. príkladu

(a) Máme 10 pozícií, na každej môžu byť dva symboly, preto 2^{10}.

(b) Rovnaký počet áčok a béčok znamená, že oboch symbolov bude 5. Preto len potrebujeme obsadiť 5 miest jedným symbolom, a na zvyšné dosadíme druhý symbol. To znamená, že takýchto slov je $\binom{10}{5}$.

(c) Najprv sa zamyslime, ako tieto slová vyzerajú. Poďme po jednotlivých dĺžkach:

- dĺžka 0: $\emptyset$, nevyhovujúce: $\{\lambda\}$
- dĺžka 1: $\emptyset$, nevyhovujúce: $\{a, b\}$
- dĺžka 2: $\{ab\}$, nevyhovujúce: $\{aa, ba, bb\}$
- dĺžka 3: $\{aba, abb, aab, bab\}$, nevyhovujúce: $\{aaa, baa, bba, bbb\}$
- dĺžka 4: $\{abaa, abba, aaba, baba, abab, abbb, aabb, babb, aaab, baab, bbab\}$, nevyhovujúce: $\{aaaa, baaa, bbaa, bbba, bbbb\}$
- ...

Vieme povedať, ako tieto slová vznikajú? Keď sa pozrieme na slová dĺžky 3, tak vidíme, že k slovu dĺžky 2 sme na koniec pridali a, b, resp. že sme tam pridali a, b na začiatok. Dobrý generátor však vytvára slová iba z jednej strany (buď zľava, alebo sprava), aby nevytváral nejaké duplicitne. Takže rátajme s tým, že tento slová generuje zľava doprava, čiže vždy niečo pridá na koniec. Akým spôsobom takto mohli vzniknúť slová aab, bab? Čo majú spoločné?

Obe majú sufix ab, ich prefixy sú a a b. Kde nájdeme tieto dve slová? V nevyhovujúcich slovách dĺžky 2.

Skúsme sa pozrieť na slová dĺžky 4, aby sme overili, či tento postup funguje. Prvých 8 slov zo slov dĺžky 4 sú slová dĺžky 3 zreťazené s a alebo b. A čo posledné tri slová? Vidíme, že tam je na konci vždy ab a ich prefixy sú tiež nevyhovujúce slová dĺžky 2 (čiže o dva kratšej ako výsledné slovo).

Označme jazyk slov obsahujúcich podslovo ab dĺžky n ako S_n a jazyk nevyhovujúcich slov dĺžky n ako $S_n^{'}$. Potom vieme vytváranie nových slov zapísať pomocou týchto pravidiel:

1. $S_{n-1}.\{a\} \cup S_{n-1}.\{b\}$
2. $S_{n-2}^{'}.\{ab\}$

Čiže $S_n = (S_{n-1}.\{a\} \ \cup \ S_{n-1}.\{b\}) \ \cup \ S_{n-2}^{'}.\{ab\}$.

Aby sme vedeli tento rekurentný vzťah použiť, je potrebné mať definované aspoň prvé dva členy. Čiže slová dĺžky 0 a 1, ktoré spĺňajú aj nespĺňajú podmienky.

Ako ale zistíme počet slov? Na počet vyhovujúcich stačí použiť obe pravidlá vyššie a dostaneme: $2|S_{n-1}| + |S_{n-2}^{'}|$.

Potrebujeme ešte vedieť počítať počet nevyhovujúcich. To je pomerne jednoduché – je ich počet všetkých mínus počet vyhovujúcich. Počet všetkých slov dĺžky n nad touto abecedou spočítame ako 2^n. Čiže vieme povedať, že $|S_n^{'}| = 2^n - |S_n|$

Použitím týchto vzorcov viete ľahko dopočítať, koľko je takých slov $|w| \leq 10$.

1. príklad

(a) Majme abecedu $\Sigma = \{0, 1\}$. Vymyslite aspoň tri príklady jazykov nad touto abecedou. Sú vaše jazyky konečné alebo nekonečné?

(b) Majme abecedu $\Sigma = \{a, b, c, d\}$. Vymyslite dva konečné jazyky L_1, L_2 nad abecedou Σ.

  (i) $L_1 \cup L_2 =$?
  (ii) $L_1 \cap L_2 =$?
  (iii) $L_1.L_2 =$?

(iv) $L_2.L_1 = ?$

(c) Čo z bežného (informatického) života môže byť jazykom? Nad akou abecedou?

(d) Majme abecedu $\Sigma = \{a, b, c, d, e\}$ a jazyk $L = \cup_{n=1}^{6} \Sigma^n$. Koľko slov z jazyka má vlastný prefix ab?

(e) Máme abecedu Σ_{bool} a jazyky $A, B, C \subseteq \Sigma^*$ také, že:

- $A = \{0, 1, 00, 11, 000, 111, 0000, 1111\}$
- $B = \{w \in \Sigma_{bool}^* | |w| \geq 2\}$
- $C = \{w \in \Sigma_{bool}^* | |w| \leq 2\}$

Určite jazyky:

(i) $A \cup B$

(ii) $A \cap B$

(iii) $A - B$

(iv) $A \cap C$

(v) $B \cup C$

(vi) $B \cap C$

Riešenie 1. príkladu

Nad riešením častí (a) až (d) sa zamyslite samostatne.

(e) Najprv si slovne popíšme, aké slová obsahujú jazyky B a C. V jazyku B sú slová nad Σ_{bool} dĺžky aspoň 2. Do jazyka C patria všetky slová Σ_{bool} dĺžky maximálne 2, čiže $\{\lambda, 0, 1, 00, 01, 10, 11\}$.

(i) $A \cup B = \{w \in \Sigma_{bool}^* | |w| \geq 1\} = \Sigma_{bool}^+$

(ii) $A \cap B = \{00, 11, 000, 111, 0000, 1111\}$

(iii) $A - B = \{0, 1\}$

(iv) $A \cap C = \{0, 1, 00, 11\}$

(v) $B \cup C = \Sigma_{bool}^*$

(vi) $B \cap C = \{w \in \Sigma_{bool}^* | |w| = 2\} = \Sigma_{bool}^2$

2. príklad

Majme abecedu Σ, nech $A, B, C \subseteq \Sigma^*$. Dokážte, že:

(a) $A\{\lambda\} = \{\lambda\}A = A$

(b) $(AB)C = A(BC)$

(c) $A(B \cup C) = AB \cup AC$

(d) $(B \cup C)A = BA \cup CA$

(e) $A(B \cap C) \subseteq AB \cap AC$

(f) $(B \cap C)A \subseteq BA \cap CA$

(g) Ak $A \subseteq B$, potom pre každé $n \in Z^+$ platí, že $A^n \subseteq B^n$.

(h)* Platí nasledovná rovnosť $A(B \cap C) = AB \cap AC$, pre jazyky nad abecedou $\Sigma = \{0\}$?

(i)* Nech $A \subseteq \Sigma_1^*$ a $B, C \subset \Sigma_2^*$ pre abecedy Σ_1, Σ_2 také, že $\Sigma_1 \cap \Sigma_2 = \emptyset$. Platí $A(B \cap C) = AB \cap AC$?

Riešenie 2. príkladu

Rovnosť v (a) vyplýva priamo z definície zreťazenia a prázdneho slova. Rovnosť v (b) zas z definície zreťazenia. Príklady (c) a (e) boli na prednáške.

(g) Túto implikáciu dokážeme pomocou matematickej indukcie.

- Báza: Ukážeme, že tvrdenie platí pre triviálny prípad, čiže v našom prípade pre $n = 1$. Ak $A \subseteq B$, potom $A^1 \subseteq B^1$. Prvá mocnina jazyka je jazyk samotný, preto to samozrejme platí.

- Indukčný predpoklad: Predpokladáme, že tvrdenie platí pre $n = k$. Čiže Ak $A \subseteq B$, potom $A^k \subseteq B^k$, pre $k \in Z^+$.

- Indukčný krok: Ideme dokázať, že ak je splnený IP, tvrdenie platí aj pre $n = k + 1$. Nech $x \in A^{k+1}$ potom (z def. mocniny jazyka) $x \in \{x_1x_2 \mid x_1 \in A, x_2 \in A^k\}$. Vychádzame z predpokladu, ak $A \subseteq B$ a $x_1 \in A$, potom aj $x_1 \in B$. Tiež z IP, ak $A^k \subseteq B^k$ a $x_2 \in A^k$, potom aj $x_2 \in B^k$. Čiže, $x \in \{x_1x_2 \mid x_1 \in B, x_2 \in B^k\} = B^{k+1}$.

3. príklad

Majme abecedu $\Sigma = \{0, 1\}$.

(a) Nech $A = \Sigma^2$. Čo je potom A^*?

(b) Nech A je rovnaké ako v časti (a), a $B = \{0, 1\}$. Aké slová obsahuje jazyk BA^*?

(c) Aký je rozdiel medzi jazykmi $\{0\}\{0, 1\}^*$ a $\{0\}\{0, 1\}^+$? Čo obsahujú?

(d) Ako zapíšete jazyk všetkých slov nad abecedou Σ, ktoré obsahujú sufix 11?

(e) Ako zapíšete jazyk všetkých slov nad abecedou Σ, ktoré obsahujú podslovo 010?

(f) Čo budeme musieť zmeniť v riešení predchádzajúcich dvoch úloh, aby jazyky obsahovali slová s vlastným sufixom resp. vlastným podslovom?

(g) Do ktorých z nasledovných jazykov nad Σ^* patrí slovo 00010?

  (i) $\{0, 1\}^*$
  (ii) $\{00\}\{0\}^*\{10\}$
  (iii) $\{000, 101\}\{10, 11\}$
  (iv) $\{000\}^*\{1\}^*\{0\}$
  (v) $\{00\}^*\{10\}^*$
  (vi) $\{0\}^*\{1\}^*\{0\}^*$

Riešenie 3. príkladu

(a) $A = \Sigma^2 = \Sigma.\Sigma = \{00, 01, 10, 11\}$. Čiže A obsahuje slová dĺžky 2 nad abecedou Σ. Ak spravíme Kleeneho uzáver tohto jazyka, dostaneme zjednotenie všetkých mocnín jazyka A. Pozrime sa aspoň na prvé štyri: $A^0 = \{\lambda\}$, $A^1 = A$, $A^2 = AA = \{0000, 0001, 0010, 0011...\}$, $A^3 = AAA = \{000000, 000001, 000010...\}$. Vidíme, že dĺžka slov v každej ďalšej mocnine bude o dva dlhšia. A^* tak obsahuje všetky slová nad Σ párnej dĺžky.

(b) Vieme, že A^* sú slová párnej dĺžky, keď ich prireťazíme k slovám z B, ktoré sú dĺžky 1, dostaneme všetky slová nepárnej dĺžky.

(c) Jazyk $\{0,1\}^*$ obsahuje aj λ, kým $\{0,1\}^+$ nie. Čiže jazyk $\{0\}\{0,1\}^*$ bude obsahovať všetky slová nad Σ s prefixom 0, a jazyk $\{0\}\{0,1\}^+$ všetky slová s vlastným prefixom 0 nad Σ (tj. dĺžky aspoň 2).

(d) Podobne ako v predchádzajúcom príklade: $\{0,1\}^*\{11\}$.

(e) $\{0,1\}^*\{010\}\{0,1\}^*$

(f) Vo výsledku z (d) stačí namiesto $*$ dať $+$. Ale stačí to aj v (e)? Aké najkratšie slová by patrili do jazyka $\{0,1\}^+\{010\}\{0,1\}^+$? Aspoň dĺžky 5, pretože aj na začiatku, aj na konci slova by musel byť aspoň jeden znak. Ale napr. aj slovo 1010 by malo patriť do tohto jazyka, keďže má vlastné poslovo 010.

Preto to skúsme zapísať aj zjednotenie dvoch jazykov: $\{0,1\}^*\{010\}\{0,1\}^+ \cup \{0,1\}^+\{010\}\{0,1\}^*$. Tak máme zaručené, že buď na konci alebo na začiatku slova bude ešte aspoň jeden znak okrem poslova 010.

(g) Slovo 00010 patrí do jazykov (i), (ii), (iii), (iv), (vi). Nepatrí iba do jazyka (v), keďže zápis $\{00\}^*$ znamená, že počet núl na začiatku slova musí byť párny.