Упражнение 7

1 Критерии за нерегулярност

Вече разгледахме два метода, чрез които показваме регулярността на даден език. Регулярните езици над дадена азука Σ обаче са изброимо много (защото регулярните изрази над Σ са изброимо много). От друга страна, множеството на всички езици над Σ, $\mathscr{P}(\Sigma^*)$, е неизброимо безкрайно като степенно множество на изброимо безкрайно множество, Σ^*. Тези съображения показват съществуването на нерегулярни езици над Σ. Ясно е, че каквато и да е Σ и каквито и да са тези езици, те най-малкото ще бъдат безкрайни. Но какво откроява безкрайните нерегулярни езици от безкрайните регулярни такива? Както знаем, всеки регулярен език се разпознава от ДКА. Ако L е безкраен регулярен език, а A е ДКА с n състояния, такъв че $L(A) = L$, то безкрайността на L влече съществуването на думи с дължина поне n. Нека w е една такава дума. Тогава пътят на w по A със сигурност минава през поне $n + 1$ състояния. По принципа на Дирихле, някое от състоянията по този път ще се повтаря; тоест пътят на w по A ще съдържа цикъл. Естествено движейки се по A ние можем да изберем да се "завъртим" по този цикъл произволен брой пъти, след което наличието на w в L ни гарантира, че ще можем да приключим този път в крайно състояние, ефективно вкарвайки безкрайно много думи (за всяко количество "завъртания"), зависими по някакъв начин от w, в L. Всичко това можем да изкажем чрез следната теорема.

Теорема 1(лема за разрастването/the pumping lemma). Нека L е регулярен език. Тогава съществува естествено число $p \geq 1$, такова че всяка дума $w \in L$ с $|w| \geq p$ може да се презапише във вида $w = xyz$, като
(1) $|y| \geq 1$, тоест $y \neq \epsilon$;
(2) $|xy| \leq p$;
(3) $xy^iz \in L$ за всяко $i \in \mathbb{N}$.

Първия (най-малкия) свидетел за съществуването на числото p е именно броят на състоянията на минималния автомат за L. Такъв автомат знаем че съществува, заради регулярността на L. Мислейки за минималния автомат A за L, всяка дума $w \in L$ с $|w| \geq p$ ще съдържа цикъл в пътя си по A. Представянето $w = xyz$ заедно със свойствата си (1), (2) и (3) точно съответства на този факт. Наистина, x е думата прочетена от началното състояние до началото на *първия* цикъл в пътя на w по A; y е думата прочетена по този цикъл (затова $|y| \geq 0$), а z е думата прочетена от края на цикъла до крайното състочние, в което w приключва пътя си по A. Свойството (2) следва от избора ни y да е думата прочетена по първия цикъл в пътя на w по A — първото повторение на състояние ще се случи след прочитането на y, тоест $|xy|$ няма как да бъде по-голяма от броя на състоянията на A. Свойството (3) е именно свойството, което дава и името на теоремата; щом можем да се въртим по така фиксирания първи цикъл в пътя на w по A произволен брой пъти, четейки по него думата y, то за всяко $i \in \mathbb{N}$, думата xy^iz ще да бъде в $L(A)$, а значи и в L.

Теорема 1 по същество е импликация от вида

$$L \text{ е регулярен} \implies P(L).$$

Разбира се в сила е и нейната контрапозиция

$$\neg P(L) \implies L \text{ не е регулярен.}$$

Това е първият критерии за нерегулярност, който ще разгледаме. Отрицанието на свойството P трябва да се разгледа по-подробно. Първо, да запишем свойството P формално.

$$P(L) \iff (\exists p \in \mathbb{N}^+)(\forall w \in L)[|w| \geq p \implies (\exists x \in \Sigma^*)(\exists y \in \Sigma^*)(\exists z \in \Sigma^*)[w = xyz \;\&\; |y| \geq 1 \;\&\; |xy| \leq p \;\&\; (\forall i \in \mathbb{N})[xy^iz \in L]]]$$

Сега можем да съобразим, че отрицанието на P е точно

$$\neg P(L) \iff (\forall p \in \mathbb{N}^+)(\exists w \in L)[|w| \geq p \;\&\; (\forall x \in \Sigma^*)(\forall y \in \Sigma^*)(\forall z \in \Sigma^*)[w = xyz \;\&\; |y| \geq 0 \;\&\; |xy| \leq p \implies (\exists i \in \mathbb{N})[xy^iz \notin L]]].$$

Ако успеем да покажем, че даден език притежава горното свойство, то контрапозицията на лемата за разрастването ни дава, че този език не е регулярен.

Първо трябва да удовлетворим кванторите в началото на формулата $\boxed{(\forall p \in \mathbb{N}^+)}$ и $\boxed{(\exists w \in L)}$, както и първия конюнкт в областта им на действие $\boxed{|w| \geq p}$. За целта избираме в ролята на w подходяща дума с дължина по-голяма от и зависеща от p в L. Зависимостта на дължината на w от p е това, което удовлетворява първия квантор.

След това удовлетворяваме втория конюнкт в областта на действие на $\boxed{(\exists w \in L)}$. За целта разглеждаме произволно представяне на така избраната w във вида $w = xyz$ със свойствата (1) и (2) и показваме такова естествено число i, че $xy^iz \notin L$.

Пример 1. Ще докажем посредством **Теорема 1**, че езикът $L = \{a^nb^n \mid n \in \mathbb{N}\}$ не е регулярен, като покажем, че L не притежава свойството P (тоест, че L притежава свойството $\neg P$). Нека $p \in \mathbb{N}^+$. Да разгледаме думата $w = a^pb^p \in L$. Очевидно $|w| \geq p$. Сега да разгледаме някое представяне $w = xyz$ на w със свойствата (1) и (2). Щом $|xy| \leq p$ и $|y| \geq 1$, то $y = a^k$ за някое k, такова че $1 \leq k \leq p$. Тогава за $i = 2$ имаме, че $xy^2z = a^{p+k}b^p \notin L$.

Сега да си припомним, че **теоремата на Майхил-Нероуд** *също* беше импликация от вида

$$L \text{ е регулярен} \implies N(L),$$

където $N(L)$ е следното свойство:

$$N(L) \iff \text{съществува ДКА за } L \text{ с толкова на брой състояния, колкото са класовете на еквивалентност на } \approx_L.$$

Това свойство влече, че броят на класовете на еквивалентност на $\approx_L$ трябва да бъде краен. Тоест имаме импликацията

$$N(L) \implies \approx_L \text{ има краен брой класове на еквивалентност.}$$

Двете импликации дотук ни дават общо

$$L \text{ е регулярен} \implies \approx_L \text{ има краен брой класове на еквивалентност.}$$

Значи в сила е и контрапозитивното

$$\approx_L \text{ има безкраен брой класове на еквивалентност} \implies L \text{ не е регулярен.}$$

Това е вторият критерии за нерегулярност, който ще разгледаме. Но как можем да покажем, че даден език има безкраен брой класове на еквивалентност по релацията си на Нероуд? Идеята е да намерим безкрайна редица от думи $w_1, w_2, ...$, такава че всеки две думи в нея са в различен клас на еквивалентност по $\approx_L$. Най-простият начин е да създадем подходяща зависимост между дължината на думата w_i и индекса ѝ — числото i.

Пример 2. Ще покажем отново, че езикът $L = \{a^n b^n \mid n \in \mathbb{N}\}$ не е регулярен; този път използвайки новия критерии. Да разгледаме редицата $w_1, w_2, ...$, за която

$$w_i = a^i, \text{ за всяко } i \in \mathbb{N}.$$

Нека $i, j \in \mathbb{N}$ са такива, че $i \neq j$. Ще покажем, че $[w_i]_L \neq [w_j]_L$. За целта е достатъчно да покажем, че $w_i \notin [w_j]_L$, тъй като $w_i \in [w_i]_L$. Еквивалентно на това, трябва да покажем, че $w_i \not\approx_L w_j$. Да разгледаме думата $b^i \in \Sigma^*$. От една страна $a^i b^i \in L$. От друга страна $a^j b^i \notin L$, тъй като $i \neq j$. Значи действително $w_i \not\approx_L w_j$. Така показахме, че $\approx_L$ има безкраен брой класове на еквивалентност. Следователно L не е регулярен.

Пример 3. Ще покажем и трети начин, по който можем да показваме, че даден език е нерегулярен. Този начин е да се възползваме от регулярните операции и езици, за които вече сме показали, че са нерегулярни. Например да разгледаме езика

$$L = \{w \in \{a, b\}^* \mid w \text{ има равен брой } a\text{-та и } b\text{-та}\}.$$

L не е регулярен, защото ако беше, то такъв би бил и $L \cap a^*b^*$—по затвореност на регулярните езици относно сечение. Обаче, $L \cap a^*b^* = \{a^nb^n \mid n \in \mathbb{N}\}$, за който вече два пъти показахме, че не е регулярен.

2 Задачи

Задача 1. Използвайте лемата за разрастването или теоремата на Майхил-Нероуд, за да покажете, че следните езици не са регулярни.
(а) $\{ww \mid w \in \{a,b\}^*\}$
(б) $\{ww^{rev} \mid w \in \{a,b\}^*\}$
(в) $\{w\overline{w} \mid w \in \{a,b\}^*\}$, където $\overline{w}$ е думата получена от w, като сме заменили всяко срещане на a с b и обратно
(г) $\{w^{|w|} \mid w \in \{a,b\}^*\}$.

Задача 2. Докажете, че езикът $\{a^nba^mba^{n+m} \mid n,m \geq 1\}$ не е регулярен.

Задача 3. Една дума w над $\{a,b\}$ наричаме *балансирана*, ако следните две неща са изпълненини:
(1) във всеки префикс на w, броят на a-тата е не по-малък от броя на b-тата;
(2) броят на a-тата в w е равен на броя на b-тата.
Докажете, че езикът $\{w \in \{a,b\}^* \mid w \text{ е балансирана}\}$ не е регулярен.

Задача 4. Кои от следните твърдения са верни? Аргументирайте се.
(а) Всяко подмножество на регулярен език е регулярен език.
(б) Всеки регулярен език си има собствено подмножество, което е регулярен език.
(в) Ако L е регулярен, то регулярен е и $\{xy \mid x \in L \ \& \ y \notin L\}$.
(г) $\{w \mid w = w^{rev}\}$ е регулярен.
(д) Ако L е регулярен език, то такъв е и $\{w \mid w \in L \ \& \ w^{rev} \in L\}$.
(е) Ако R е множество от регулярни езици, то $\bigcup R$ е регулярен език.
(ж) $\{xyx^{rev} \mid x,y \in \Sigma^*\}$ е регулярен.

3 Решения

Задача 1. Ако използвате PL едни възможни избори са:
(а) $w = a^pba^pb$, $i = 2$;
(б) $w = a^pbba^p$, $i = 2$;
(в) $w = a^pb^p$, $i = 2$;
(г) $w = (a^pb)^{p+1}, i = 2$.

Ако използвате Майхил-Нероуд:
(а) $w_i = a^i b$, за $i \in \mathbb{N}$; конкатенирате $a^i b$ към w_i и w_j за $i \neq j$;
(б) $w_i = a^i b$, за $i \in \mathbb{N}$; конкатенирате ba^i към w_i и w_j за $i \neq j$;
(в) $w_i = a^i$, за $i \in \mathbb{N}$; конкатенирате b^i към w_i и w_j за $i \neq j$;
(г) $w_i = (a^i b)^i$, за $i \in \mathbb{N}$; конкатенирате $a^i b$ към w_i и w_j за $i \neq j$;

Алтернативен подход за (в) е да пресечем езика с a^*b^*, получавайки $\{a^n b^n \mid n \in \mathbb{N}\}$.

Задача 2. С PL: $w = a^p b a^p b a^{2p}$, $i = 2$. С Майхил-Нероуд: $w_i = a^{i+1} b a^{i+1} b$, за $i \in \mathbb{N}$ и конкатенирате $a^{2(i+1)}$ към w_i и w_j за $i \neq j$.

Задача 3. Пресичаме езика с a^*b^* и получаваме $\{a^n b^n \mid n \in \mathbb{N}\}$.

Задача 4. (а) Не е вярно. Най-просто, Σ^* е регулярен език и всеки език над Σ е подмножество на Σ^*. Но както показахме в това упражнение, над Σ съществуват и нерегулярни езици.
(б) Не е вярно. Тривиално, $\varnothing$ е регулярен език, който няма собствени подмножества.
(в) Вярно е. Имаме

$$\begin{gathered}\{xy \mid x \in L \ \&\ y \notin L\} = \\ \{xy \mid x \in L \ \&\ y \in \overline{L}\} = \\ \{x \mid x \in L\} \circ \{y \mid y \in \overline{L}\} = L\overline{L}.\end{gathered}$$

Но ние вече показахме, че регулярните езици са затворени относно операциите допълнение и конкатенация.
(г) Не е вярно. Ако пресечем този език с езика на думите с четна дължина, $(aa + ab + ba + bb)^*$, получаваме езика от (б) на **Задача 1**.
(д) Вярно е. Имаме

$$\begin{gathered}\{w \mid w \in L \ \&\ w^{rev} \in L\} = \\ \{w \mid w \in L \ \&\ w \in L^{rev}\} = L \cap L^{rev}.\end{gathered}$$

Но ние вече показахме, че регулярните езици са затворени относно операциите обръщане и сечение.
(е) Не е вярно. Въпреки че обединението, приложено краен брой пъти, запазва регулярността, ако го приложим върху подходящо безкрайно множество от регулярни езици, можем да получим нерегулярен език. Конкретно, да разгледаме в ролята на R множеството от крайни и следователно регулярни езици $\{\{\epsilon\}, \{ab\}, \{aabb\}, \{aaabbb\}, \ldots\}$. Очевидно $\bigcup R = \{a^n b^n \mid n \in \mathbb{N}\}$.
(ж) Вярно е. Този език е точно Σ^*.