Poznámky k cvičeniu č. 1

Peter Kostolányi

27. septembra 2017

Označenia

Označenia nie sú v matematike úplne ustálené, čo môže v určitých situáciách viesť k nedorozumeniam. V nasledujúcom preto uvádzame niekoľko notačných dohôd, ktorých sa budeme držať.

- Symbolom $\mathbb{N}$ označujeme množinu všetkých prirodzených čísel *s nulou*, t.j. $\mathbb{N} = \{0, 1, 2, \ldots\}$.
- Symboly $\subseteq, \supseteq$ označujú *neostrú* inklúziu.
- Symboly $\subsetneq, \supsetneq$ označujú *ostrú* inklúziu.
- Symboly $\nsubseteq, \nsupseteq$ označujú *komplement neostrej inklúzie*. Pre množiny X, Y teda platí $X \nsubseteq Y$ práve vtedy, keď neplatí $X \subseteq Y$.
- Symboly $\subset, \supset$ môžu označovať ako ostré, tak aj neostré inklúzie – ide predovšetkým o otázku vkusu, ktorý sa môže od autora k autorovi líšiť. Tieto symboly *nebudeme používať*.

Slová

Abeceda je ľubovoľná neprázdna konečná množina. Jej prvky, ktorými môžu byť ľubovoľné objekty, nazývame *symboly* (*znaky*) alebo *písmená*. Symboly tak definujeme výhradne vo vzťahu k abecede – ak nie je daná abeceda, nemá zmysel ani pojem symbolu.

Slovo nad abecedou Σ je ľubovoľná konečná postupnosť symbolov z abecedy Σ. Pri zápise takýchto postupností nepoužívame čiarky; príkladom slova nad dvojprvkovou abecedou $\{a, b\}$ tak môže byť napríklad slovo $w = ababb$. *Prázdne slovo* nad abecedou Σ (špeciálny prípad slova) je prázdna postupnosť prvkov z abecedy Σ. V rámci tohto predmetu budeme prázdne slovo *vždy* označovať ε. V literatúre ale možno natrafiť aj na iné označenia, napríklad λ, 1, 1_{Σ^*} alebo e. *Dĺžka slova w*, označovaná symbolom $|w|$, je definovaná ako dĺžka zodpovedajúcej postupnosti znakov. Ak teda $w = a_1 a_2 \ldots a_n$, kde $a_1, a_2, \ldots, a_n$ sú symboly[1] z nejakej abecedy Σ, píšeme $|w| = n$.

Poznámka 1. Slová rovnakej dĺžky nad rôznymi abecedami, zhodujúce sa vo všetkých symboloch, stotožňujeme. To nám napríklad umožňuje hovoriť len o slove *abab* bez upresnenia, či ho chápeme ako slovo nad abecedou $\{a, b\}$ alebo, povedzme, ako slovo nad abecedou $\{a, b, c\}$. Podobne aj naše označenie prázdneho slova symbolom ε by bez uvedenej dohody nebolo korektné. Nejde však o nič neobvyklé, ale iba o špeciálny prípad zaužívaného zvyku stotožňovať funkcie líšiace sa len v koobore, ako napríklad $\sin\colon \mathbb{R} \to \mathbb{R}$ a $\sin\colon \mathbb{R} \to [-1, 1]$.

Stotožňovať budeme aj symboly abecedy Σ a jednopísmenové slová nad Σ, hoci z prísne formálneho hľadiska ide opäť o rôzne objekty. Nutnosť takéhoto dohovoru je v skutočnosti daná už naším spôsobom zápisu slov – napríklad symbol a a slovo a v ňom totiž nemožno odlíšiť.

Poznámka 2. Označenie prázdneho slova gréckym písmenom ε môže zvádzať k mylnému dojmu, že ε je prvkom abecedy. Treba si však uvedomiť, že ε slúži iba ako naše označenie pre prázdnu *postupnosť* symbolov z abecedy, ktorá symbolom nie je.[2] Čitateľ tento drobný rozpor medzi pojmom „symbol" v zmysle definovanom vyššie a pojmom „symbol" používaným v bežnej reči snáď bez väčších problémov strávi – v nasledujúcom sa mu totiž nevyhneme.

Ak w je slovo nad abecedou Σ a c je symbol abecedy Σ, označujeme ako $\#_c(w)$ – v literatúre sa možno stretnúť aj s notáciou $|w|_c$ – *počet výskytov symbolu c v slove w*. Formálne: ak $w = a_1 \ldots a_n$, kde $a_1, \ldots, a_n$ sú zo Σ, tak $\#_c(w) = |\{i \in \{1, \ldots, n\} \mid a_i = c\}|$.

[1] Takýto predpoklad znamená, že $a_1, \ldots, a_n$ sú premenné nadobúdajúce hodnoty v Σ. Ak napríklad $\Sigma = \{a, b\}$, tak pre $i = 1, \ldots, n$ môže platiť $a_i = a$ alebo $a_i = b$. Rozdiel medzi symbolom-premennou a symbolom-konštantou je väčšinou zrejmý z kontextu.

[2] Bezúčelné pokusy o abecedu obsahujúcu prázdne slovo nad sebou samou vedú k množinovým paradoxom.

Podslovo slova $w = a_1a_2\dots a_n$ je ľubovoľná *súvislá* podpostupnosť postupnosti $a_1a_2\dots a_n$. *Prefix* je podslovo začínajúce prvým písmenom, *sufix* je podslovo končiace posledným písmenom. Prázdne slovo ε je prefixom a zároveň sufixom každého slova.

Množinu všetkých slov[3] nad abecedou Σ označujeme Σ^*, množinu všetkých neprázdnych slov nad Σ – slovo je neprázdne, ak nie je prázdne – označujeme Σ^+ a množinu všetkých slov nad Σ dĺžky k označujeme Σ^k. Ako uvidíme neskôr, abeceda je špeciálnym prípadom jazyka a uvedené označenia súhlasia s označeniami pre zodpovedajúce operácie na jazykoch.

Notačné konvencie

Dôležitou súčasťou kultivovaného matematického prejavu je používanie notačných konvencií, ktorých cieľom je uľahčiť čitateľovi prípadne poslucháčovi jeho úlohu. Aj keď vo väčšine oblastí matematiky existuje viacero rôznych notačných konvencií a teória formálnych jazykov v tomto nie je výnimkou, je nanajvýš vhodné (v prípadoch, keď neexistuje žiaden dobrý dôvod pre opak) pridŕžať sa označení, ktoré možno považovať za „obvyklé“. Význam notačných konvencií možno ilustrovať na príklade definície limity postupnosti reálnych čísel $\{a_n\}_{n=0}^{\infty}$:

$$\lim_{n\to\infty} a_n = L \iff \forall\varepsilon > 0\ \exists n_0 \in \mathbb{N}\ \forall n \geq n_0 : |a_n - L| < \varepsilon.$$

To isté možno zapísať aj ako

$$\lim_{\clubsuit\to\infty} a_\clubsuit = \zeta'' \iff \forall U_{42} > 0\ \exists\varepsilon \in \mathbb{N}\ \forall\clubsuit \geq \varepsilon : |a_\clubsuit - \zeta''| < U_{42}.$$

Aj keď je aj druhý zápis po formálnej stránke absolútne korektný, istý rozdiel v čitateľnosti oboch zápisov je asi očividný.

V nasledujúcom preto zavedieme niekoľko notačných konvencií, ktorých sa budeme na týchto cvičeniach pridŕžať. Pôjde o dohody, ktorých platnosť je ďaleko od univerzálnej a je tiež potrebné mať na pamäti, že existujú aj situácie, kedy je vhodné a opodstatnené tieto konvencie opustiť.

- „Nejakú abecedu“ – čiže abecedu bez ďalšej špeciálnej úlohy (v gramatike a pod.) – zvyčajne označujeme veľkým gréckym písmenom Σ alebo Γ s prípadnými indexmi alebo inými „ozdobami“, napr. $\Sigma', \Sigma'', \Gamma_1, \Gamma_2, \dots$
- Symboly zvyčajne označujeme malými písmenami zo začiatku latinskej abecedy s prípadnými „ozdobami“, napr. $a, b, c, d, a_1, a_2, \dots, a_n, \dots$ To platí vo všeobecnosti – pre symboly niektorých špeciálnych abecied ale budeme neskôr používať úplne odlišné konvencie.
- Slová označujeme malými písmenami z konca latinskej abecedy s prípadnými „ozdobami“, napr. $w, u, v, x, y, z, w', w'', u_1, \dots, u_n, \dots$

Neskôr ešte zavedieme niekoľko ďalších podobných konvencií, napríklad v súvislosti s jazykmi, gramatikami a automatmi. Treba ale upozorniť na fakt, že napríklad naše označenie prázdneho slova symbolom ε nemožno v žiadnom prípade považovať za notačnú konvenciu v zmysle opísanom vyššie. V tomto prípade ide o *definíciu* významu priradeného danému symbolu, ktorá je platná minimálne v rámci prednášky a cvičení k tomuto predmetu. Kým v prípade notačných konvencií nie je použitie „nekonvenčného“ symbolu chybou[4] a v ojedinelých prípadoch je dokonca žiadúce, v prípade prázdneho slova je použitie ľubovoľného označenia iného ako ε chybou v prípade, že sa toto označenie nevzťahuje na nejakú predom uvedenú alternatívnu definíciu.

Poznámka o platnosti definícií

Podobne ako označenia, ani definície matematických objektov často nebývajú úplne ustálené. Tak je tomu napríklad pri pojme abecedy, kde sa občas možno stretnúť aj s definíciami nevyžadujúcimi jej konečnosť alebo neprázdnosť. V literatúre – najmä u autorov hlásiacich sa k tzv. francúzskej

[3] Alebo v terminológii, ktorú zavedieme neskôr: jazyk všetkých slov.

[4] Význam symbolu totiž v takýchto situáciách nebýva definovaný predom.

škole – sa tiež možno stretnúť s pojmom „podslovo“ chápaným ako ľubovoľná (nie nutne súvislá) podpostupnosť, pričom pre súvislú podpostupnosť sa používa pojem „faktor“. Podobne rôzne druhy gramatík či automatov, s ktorými sa na tomto predmete budeme stretávať neskôr, bývajú často definované v rôznych (väčšinou v určitom zmysle ekvivalentných) obmenách, ktoré sú predovšetkým otázkou ich zamýšľaného použitia a v neposlednom rade aj vkusu autora.

Je však dobrým zvykom v každom odbornom texte, ako aj na každej prednáške či cvičení, predom sa dohodnúť na používaní určitých definícií a túto dohodu neskôr dodržiavať.[5] V rámci týchto cvičení sa preto všetky pojmy vzťahujú vždy iba na ich definície zavedené na prednáške, bez ohľadu na to, či sa v literatúre vyskytuje alebo nevyskytuje alternatívna definícia.

Operácie na slovách

Kľúčovou operáciou je v teórii formálnych jazykov *zreťazenie slov* – zreťazením slov $u = a_1 \dots a_n$ a $v = b_1 \dots b_m$ je, pomerne prirodzene, slovo $u \cdot v = a_1 \dots a_n b_1 \dots b_m$. Presnejšie:

Definícia 1. Nech Σ_1, Σ_2 sú abecedy, n, m sú z $\mathbb{N}$, $a_1, \dots, a_n \in \Sigma_1$ a $b_1, \dots, b_m \in \Sigma_2$ sú symboly, $u = a_1 \dots a_n \in \Sigma_1^*$ je slovo nad abecedou Σ_1 a $v = b_1 \dots b_m \in \Sigma_2^*$ je slovo nad abecedou Σ_2. *Zreťazenie slov* u a v je slovo $u \cdot v$ nad abecedou $\Sigma_1 \cup \Sigma_2$, definované ako $u \cdot v = a_1 \dots a_n b_1 \dots b_m$.

Zreťazenie stačí definovať pre dvojice slov nad rovnakou abecedou – za ňu možno vziať $\Sigma_1 \cup \Sigma_2$. V nasledujúcom budeme symbol $\cdot$ väčšinou vynechávať a zreťazenie slov u a v označovať uv.

Na základe zreťazenia možno definovať aj mocninu slova.

Definícia 2. Nech Σ je abeceda $w \in \Sigma^*$ je slovo. Potom $w^0 = \varepsilon$ a $w^{n+1} = w^n w$ pre všetky $n \in \mathbb{N}$.

Nultá mocnina slova je definovaná prirodzeným spôsobom ako neutrálny prvok vzhľadom na operáciu zreťazenia, ktorým je prázdne slovo ε – čitateľ ľahko overí, že pre všetky slová w skutočne platí $\varepsilon w = w \varepsilon = w$.

Pod *reverzom* alebo pod *zrkadlovým obrazom* slova w rozumieme „slovo w čítané sprava doľava“.

Definícia 3. Nech Σ je abeceda, n je z $\mathbb{N}$, $a_1, \dots, a_n \in \Sigma$ sú symboly a $w = a_1 \dots a_n \in \Sigma^*$ je slovo. *Reverz* alebo *zrkadlový obraz* slova w je slovo $w^R = a_n \dots a_1$.

Jazyky a množinové operácie na nich

Jazyk nad abecedou Σ je ľubovoľná množina slov nad abecedou Σ. V rámci notačnej konvencie väčšinou jazyky označujeme symbolom L s prípadnými „ozdobami“, napríklad $L, L', L_0, L_1, \dots$ *Prázdny jazyk* je prázdna množina a označujeme ho symbolom $\emptyset$. Je dôležité uvedomiť si, že jazyk obsahujúci iba prázdne slovo, čiže jazyk $\{\varepsilon\}$, je iný objekt ako jazyk $\emptyset$. Jazyk je *konečný*, ak ide o konečnú množinu. Uveďme zopár príkladov jazykov:

- $L_1 = \{\varepsilon, a, abb, baa\}$ je konečný jazyk nad abecedou $\{a, b\}$.
- $L_2 = \{a, b\}^*$ je jazyk všetkých slov nad abecedou $\{a, b\}$.
- $L_3 = \{w \in \{a, b\}^* \mid w = w^R\}$ je jazyk všetkých palindrómov[6] nad abecedou $\{a, b\}$.
- $L_4 = \{w \in \{a, b\}^* \mid \#_a(w) = \#_b(w)\}$ je jazyk tých slov nad abecedou $\{a, b\}$, ktoré obsahujú rovnaký počet výskytov oboch symbolov.
- $L_5 = \{a^n b^n \mid n \in \mathbb{N}\}$ je jazyk nad abecedou $\{a, b\}$.
- Každá abeceda je konečný jazyk nad sebou samým.

[5] V prípade, že je neskôr z nejakého dôvodu treba zmeniť definíciu, je nutné na to aspoň upozorniť. Najvhodnejším spôsobom upozornenia však vo väčšine prípadov býva nová definícia spojená so zmenou v terminológii.

[6] Slov, ktoré sa zľava doprava „čítajú rovnako“ ako sprava doľava.

Napríklad jazyk L_1 však súčasne možno chápať aj ako jazyk nad abecedou $\{a, b, c\}$ a podobne. Z tohto dôvodu pre každý jazyk L rôzny od $\emptyset$ a $\{\varepsilon\}$ definujeme abecedu Σ_L ako *najmenšiu abecedu* Σ takú, že L je jazykom nad Σ. Presnejšie: Σ_L je abeceda taká, že L je jazykom nad Σ_L a nie je jazykom nad žiadnou abecedou $\Sigma \subsetneq \Sigma_L$. Abecedy $\Sigma_\emptyset$ a $\Sigma_{\{\varepsilon\}}$ nemožno definovať, pretože abeceda musí byť z definície neprázdna.

Keďže je jazyk definovaný ako množina, má zmysel zaoberať sa množinovými operáciami aplikovanými na jazyky. V tomto duchu možno uvažovať napríklad:

- *Zjednotenie* $L_1 \cup L_2$ a *prienik* $L_1 \cap L_2$ jazykov L_1 a L_2.
- *Rozdiel* $L_1 - L_2$ jazykov L_1 a L_2. Túto notáciu uprednostňujeme pred v matematike zaužívaným označením $L_1 \setminus L_2$, ktoré sa v teórii jazykov používa pre tzv. ľavý kvocient.[7]
- *Komplement* L^C jazyka L. Podobne ako aj inde v matematike je komplement nutné uvažovať v rámci nejakého univerza. Ak nie je povedané inak, v teórii formálnych jazykov sa zväčša ako univerzum chápe jazyk Σ_L^* všetkých slov nad abecedou Σ_L. Platí teda $L^C = \Sigma_L^* - L$. Pri jazykoch $\emptyset$ a $\{\varepsilon\}$ musí byť univerzum explicitne stanovené ad hoc.

Zreťazenie jazykov

Zreťazenie $L_1 \cdot L_2$ jazykov L_1 a L_2 je definované ako jazyk zreťazení uv všetkých dvojíc slov u, v takých, že $u \in L_1$ a $v \in L_2$.

Definícia 4. Nech L_1, L_2 sú jazyky. *Zreťazenie jazykov* L_1 a L_2 je jazyk $L_1 \cdot L_2$ definovaný ako $L_1 \cdot L_2 = \{uv \mid u \in L_1; v \in L_2\}$.

Zreťazenie má vyššiu prioritu ako zjednotenie a prienik, napríklad $L_1 \cup L_2 \cdot L_3$ teda treba čítať ako $L_1 \cup (L_2 \cdot L_3)$.

Príklad 1. Uvažujme *konečné* jazyky $L_1 = \{\varepsilon, a, ab, abb\}$ a $L_2 = \{b, ab, bb\}$. Jazyk $L_1 \cdot L_2$ potom možno získať nasledujúcim postupom:

1. Zreťazenia zapíšeme do tabuľky.

	b	ab	bb
ε	b	ab	bb
a	ab	aab	abb
ab	abb	$abab$	$abbb$
abb	$abbb$	$abbab$	$abbbb$

2. Po odstránení duplikátov dostávame $L_1 \cdot L_2 = \{b, ab, bb, aab, abb, abab, abbb, abbab, abbbb\}$.

Príklad 2. Uvažujme abecedu $\Sigma = \{a, b\}$ a jazyk Σ^* všetkých slov nad abecedou Σ. Dokážeme, že $\Sigma \cdot \Sigma^* = \Sigma^+$, kde Σ^+ je jazyk všetkých *neprázdnych* slov nad Σ.

$\subseteq$: Nech $w \in \Sigma \cdot \Sigma^*$. Potom $w = uv$, kde $u \in \Sigma$ a $v \in \Sigma^*$. Zrejme $w = uv \in \Sigma^*$. Keďže $u \in \Sigma$, platí $|u| = 1$. Potom $|w| = |uv| \geq |u| = 1$, a teda $w = uv \in \Sigma^+$.

$\supseteq$: Nech $w \in \Sigma^+$. Potom existuje slovo $u \in \Sigma^*$ také, že buď $w = au$, alebo $w = bu$. V oboch prípadoch dostávame $w \in \Sigma \cdot \Sigma^*$.

Poznámka 3. Je užitočné ujasniť si úlohu, ktorú pri operácii zreťazenia zohrávajú význačné jazyky $\emptyset$ a $\{\varepsilon\}$. Jazyk $\{\varepsilon\}$ je neutrálnym prvkom vzhľadom na zreťazenie – pre všetky jazyky L platí $L \cdot \{\varepsilon\} = \{\varepsilon\} \cdot L = L$. Naopak jazyk $\emptyset$ je vzhľadom na zreťazenie agresívny – pre všetky jazyky L platí $L \cdot \emptyset = \emptyset \cdot L = \emptyset$. Tieto dva jazyky sú teda *fundamentálne* odlišné: kým $\{\varepsilon\}$ zohráva úlohu jednotky, $\emptyset$ zohráva úlohu nuly.

[7] V literatúre sa však možno stretnúť aj s opačným prístupom, pri ktorom sa používa v matematike bežná notácia $L_1 \setminus L_2$ pre rozdiel a alternatívne označenie pre ľavý kvocient.

Mocnina jazyka

Mocnina jazyka je, podobne ako mocnina slova, definovaná prirodzeným induktívnym predpisom, pričom nultá mocnina je definovaná ako neutrálny prvok vzhľadom na zreťazenie – čiže jazyk $\{\varepsilon\}$.

Definícia 5. Nech L je jazyk. Potom $L^0 = \{\varepsilon\}$ a $L^{n+1} = L^n \cdot L$ pre všetky $n \in \mathbb{N}$.

Dokážeme teraz, že jazyk L^n pozostáva zo zreťazení všetkých n-tíc slov z jazyka L.

Tvrdenie 1. *Nech L je jazyk a $n \in \mathbb{N}$. Potom $L^n = \{w_1 \dots w_n \mid w_1, \dots, w_n \in L\}$.*

Dôkaz. Indukciou vzhľadom na n.

1. Pre $n = 0$ je tvrdenie triviálne.
2. Nech tvrdenie platí pre $n = k$. Ukážeme, že platí aj pre $n = k + 1$. Skutočne,

$$L^{k+1} = L^k \cdot L = \{w_1 \dots w_k \mid w_1, \dots, w_k \in L\} \cdot L = \{w_1 \dots w_{k+1} \mid w_1, \dots, w_{k+1} \in L\},$$

kde prvá rovnosť je z definície 5, druhá z indukčného predpokladu a posledná z definície 4.

Tvrdenie je dokázané. □

Častou chybou u študentov býva zamieňanie si n-tej mocniny jazyka L s jazykom n-tých mocnín slov z jazyka L. Čitateľ určite ľahko dokáže, že vo všeobecnosti $L^n \neq \{w^n \mid w \in L\}$.

Aj keď je význam výrazov typu 0^0 v matematike pomerne diskutabilný, v teórii jazykov je zvykom definovať $\emptyset^0 = \{\varepsilon\}$, čo koniec koncov vyplýva aj z definície 5. Ide tu o určitú analógiu (dobre odôvodnenej) konvencie $0^0 = 1$, ktorá sa používa v niektorých oblastiach matematiky.

Iterácia a kladná iterácia

Definícia 6. Nech L je jazyk. *Iterácia* jazyka L je jazyk

$$L^* = \bigcup_{k=0}^{\infty} L^k$$

a *kladná iterácia* jazyka L je jazyk

$$L^+ = \bigcup_{k=1}^{\infty} L^k.$$

Operácia iterácie sa zvykne nazývať aj *uzáver* alebo *Kleeneho hviezdička* a operácia kladnej iterácie sa nazýva aj *kladný uzáver* alebo *Kleeneho plus*.

Vyššie zavedené označenia Σ^* a Σ^+ pre abecedu Σ súhlasia s definíciami iterácie a kladnej iterácie jazyka. Vyplýva to z nasledujúceho tvrdenia.

Tvrdenie 2. *Nech L je jazyk. Pre jeho iteráciu platí $L^* = \{w_1 \dots w_k \mid k \in \mathbb{N};\ w_1, \dots, w_k \in L\}$ a pre jeho kladnú iteráciu platí $L^+ = \{w_1 \dots w_k \mid k \in \mathbb{N};\ k \geq 1;\ w_1, \dots, w_k \in L\}$.*

Dôkaz. Ak rozpíšeme L^k podľa tvrdenia 1, dostávame pre iteráciu

$$L^* = \bigcup_{k=0}^{\infty} L^k = \bigcup_{k=0}^{\infty} \{w_1 \dots w_k \mid w_1, \dots, w_k \in L\} = \{w_1 \dots w_k \mid k \in \mathbb{N};\ w_1, \dots, w_k \in L\}$$

a pre kladnú iteráciu

$$L^+ = \bigcup_{k=1}^{\infty} L^k = \bigcup_{k=1}^{\infty} \{w_1 \dots w_k \mid w_1, \dots, w_k \in L\} = \{w_1 \dots w_k \mid k \in \mathbb{N};\ k \geq 1;\ w_1, \dots, w_k \in L\},$$

čo bolo treba dokázať. □

Reverz jazyka

Definícia 7. Nech L je jazyk. *Reverz* alebo *zrkadlový obraz* jazyka L je jazyk

$$L^R = \{w^R \mid w \in L\}.$$

Príklad 3. Ak $L_1 = \{\varepsilon, abb, bab\}$, tak $L_1^R = \{\varepsilon, bba, bab\}$. Pre jazyk $L_2 = \{a^n b^n \mid n \in \mathbb{N}\}$ platí $L_2^R = \{b^n a^n \mid n \in \mathbb{N}\}$. Napokon, ak $L_3 = \{w \in \{a,b\}^* \mid w = w^R\}$, tak $L_3^R = L_3$.

Ukážkové riešené úlohy

Úloha 1. Uvažujme dvojicu jazykov

$$L_1 = \{w \in \{a,b\}^* \mid \#_a(w) \leq \#_b(w)\}$$
$$L_2 = \{w \in \{a,b\}^* \mid \#_a(w) \geq \#_b(w)\}.$$

Nájdite jazyk $L_1 \cdot L_2$ a svoje tvrdenie dokážte.

Riešenie. Dokážeme, že $L_1 \cdot L_2 = \{a,b\}^*$.

$\subseteq$: Nech $w \in L_1 \cdot L_2$. Potom $w = uv$, kde $u \in L_1 \subseteq \{a,b\}^*$ a $v \in L_2 \subseteq \{a,b\}^*$. Preto aj $w = uv \in \{a,b\}^*$.

$\supseteq$: Nech $w \in \{a,b\}^*$. Potom buď $\#_a(w) \leq \#_b(w)$, alebo $\#_a(w) \geq \#_b(w)$. Ak $\#_a(w) \leq \#_b(w)$, tak $w = w\varepsilon$, pričom $w \in L_1$ a $\varepsilon \in L_2$. Ak $\#_a(w) \geq \#_b(w)$, tak $w = \varepsilon w$, pričom $\varepsilon \in L_1$ a $w \in L_2$. V oboch prípadoch tak dostávame $w \in L_1 \cdot L_2$. $\square$

Úloha 2. Uvažujme jazyk $L = \{w \in \{a,b\}^* \mid |w| \geq 4\}$. Nájdite jazyk L^* a svoje tvrdenie dokážte.

Riešenie. Dokážeme, že $L^* = L \cup \{\varepsilon\}$.

$\subseteq$: Nech $w \in L^*$. Potom existuje $k \in \mathbb{N}$ a slová $w_1, \dots, w_k \in L$ tak, že $w = w_1 \dots w_k$. Ak $k = 0$, tak $w = \varepsilon \in L \cup \{\varepsilon\}$. Ak $k > 0$, tak $|w| \geq |w_1| \geq 4$, keďže $w_1 \in L$. Preto $w \in L \subseteq L \cup \{\varepsilon\}$.

$\supseteq$: Nech $w \in L \cup \{\varepsilon\}$. Potom $w \in L = L^1 \subseteq L^*$, alebo $w \in \{\varepsilon\} = L^0 \subseteq L^*$. $\square$

Úloha 3. Nech L_1, L_2 sú jazyky. Porovnajte[8] jazyky $(L_1 \cup L_2)^*$ a $L_1^* \cup L_2^*$. Svoje tvrdenie dokážte.

Riešenie. Dokážeme, že $(L_1 \cup L_2)^* \supseteq L_1^* \cup L_2^*$, pričom opačná inklúzia vo všeobecnosti neplatí.

$\not\subseteq$: Nech $L_1 = \{a\}$ a $L_2 = \{b\}$. Potom $ab \in (L_1 \cup L_2)^*$, ale $ab \notin L_1^* \cup L_2^*$.

$\supseteq$: Nech $w \in L_1^* \cup L_2^*$. Potom buď $w \in L_1^*$, alebo $w \in L_2^*$.

Ak $w \in L_1^*$, podľa tvrdenia 2 existuje $k \in \mathbb{N}$ a slová $w_1, \dots, w_k \in L_1$ tak, že $w = w_1 \dots w_k$. Slová $w_1, \dots, w_k$ sú súčasne aj v $L_1 \cup L_2$; z tvrdenia 2 preto $w = w_1 \dots w_k \in (L_1 \cup L_2)^*$.

Podobne ak $w \in L_2^*$, tak $w = w_1 \dots w_k$, kde $k \in \mathbb{N}$ a $w_1, \dots, w_k \in L_2$. Slová $w_1, \dots, w_k$ sú zároveň aj v $L_1 \cup L_2$, a teda $w = w_1 \dots w_k \in (L_1 \cup L_2)^*$.

Tvrdenie je dokázané. $\square$

Úloha 4. Nech L_1, L_2 sú jazyky. Porovnajte jazyky $(L_1 \cup L_2)^*$ a $L_1^* \cdot (L_2 \cdot L_1^*)^*$. Svoje tvrdenie dokážte.

Riešenie. Dokážeme, že $(L_1 \cup L_2)^* = L_1^* \cdot (L_2 \cdot L_1^*)^*$. Postupne teda overíme obidve inklúzie:

[8]Zistite, či *vo všeobecnosti* – teda pre všetky dvojice jazykov L_1, L_2 – platia inklúzie $(L_1 \cup L_2)^* \subseteq L_1^* \cup L_2^*$ resp. $(L_1 \cup L_2)^* \supseteq L_1^* \cup L_2^*$. Na dôkaz niektorej z inklúzií je potrebné uvažovať *všetky* dvojice jazykov L_1, L_2. Na vyvrátenie inklúzie naopak stačí nájsť konkrétnu dvojicu jazykov L_1, L_2, pre ktorú inklúzia neplatí. Pri riešení úloh tohto typu je vždy potrebné overiť obidve inklúzie – nestačí sa teda uspokojiť s konštatovaním, že daná dvojica jazykov sa vo všeobecnosti nerovná.

$\subseteq$: Nech $w \in (L_1 \cup L_2)^*$. Potom podľa tvrdenia 2 existuje $k \in \mathbb{N}$ a slová $w_1, \ldots, w_k \in L_1 \cup L_2$ tak, že $w = w_1 \ldots w_k$. Nech $1 \leq i_1 < \ldots < i_s \leq k$ sú indexy také, že $w_{i_1}, \ldots, w_{i_s}$ sú práve všetky slová spomedzi $w_1, \ldots, w_k$, ktoré sú v L_2. Zvyšné zo slov $w_1, \ldots, w_k$ preto musia byť v L_1, a teda všetky slová

$$u_0 = w_1 w_2 \ldots w_{i_1 - 1}, \quad u_1 = w_{i_1+1} w_{i_1+2} \ldots w_{i_2-1}, \quad \ldots, \quad u_s = w_{i_s+1} w_{i_s+2} \ldots w_k$$

musia byť podľa tvrdenia 2 v L_1^* (niektoré z týchto slov môžu byť aj prázdne). Pre $j = 1, \ldots, s$ teda platí $w_{i_j} u_j \in L_2 \cdot L_1^*$, z čoho podľa tvrdenia 2 vyplýva $w_{i_1} u_1 \ldots w_{i_s} u_s \in (L_2 \cdot L_1^*)^*$. Zrejme platí

$$w = u_0 w_{i_1} u_1 \ldots w_{i_s} u_s,$$

pričom vieme, že $u_0 \in L_1^*$ a $w_{i_1} u_1 \ldots w_{i_s} u_s \in (L_2 \cdot L_1^*)^*$. Z definície zreťazenia jazykov teda vyplýva $w \in L_1^* \cdot (L_2 \cdot L_1^*)^*$, čo bolo treba dokázať.

$\supseteq$: Nech $w \in L_1^* \cdot (L_2 \cdot L_1^*)^*$. Z definície zreťazenia jazykov a tvrdenia 2 vyplýva $w = u_0 v_1 \ldots v_k$, kde $u_0 \in L_1^*$ a $v_1, \ldots, v_k \in L_2 \cdot L_1^*$. Keďže $u_0 \in L_1^*$, z tvrdenia 2 vyplýva $u_0 = x_{0,1} \ldots x_{0,j_0}$ pre nejaké $j_0 \in \mathbb{N}$ a nejaké slová $x_{0,1}, \ldots, x_{0,j_0} \in L_1 \subseteq L_1 \cup L_2$. Keďže pre $i = 1, \ldots, k$ platí $v_i \in L_2 \cdot L_1^*$, z definície zreťazenia jazykov a tvrdenia 2 vyplýva, že $v_i = y_i x_{i,1} \ldots x_{i,j_i}$ pre nejaké $j_i \in \mathbb{N}$ a nejaké slová $y_i \in L_2 \subseteq L_1 \cup L_2$ a $x_{i,1}, \ldots, x_{i,j_i} \in L_1 \subseteq L_1 \cup L_2$. Platí teda

$$w = x_{0,1} \ldots x_{0,j_0} y_1 x_{1,1} \ldots x_{1,j_1} \ldots y_k x_{k,1} \ldots x_{k,j_k},$$

kde všetky slová $x_{i,j}$, $i = 0, \ldots, k$, $j = 1, \ldots, j_i$ a y_i, $i = 1, \ldots, k$ sú v $L_1 \cup L_2$. Z tvrdenia 2 teda vyplýva $w \in (L_1 \cup L_2)^*$, čo bolo treba dokázať.

Tvrdenie je dokázané. $\square$