Pumppauslemmaopas

Juhana Laurinharju Jani Rahkola

19. lokakuuta 2015

Pumppauslemma

Jokaisella säännöllisellä kielellä on seuraava *pumppauslemmana* tunnettu ominaisuus.

Määritelmä (pumppauslemma)**.** Olkoon A säännöllinen kieli. Tällöin A:lla on jokin pumppauspituus $p \in \mathbb{N}$, $p > 0$. Nyt kaikki vähintään p:n mittaiset merkkijonot $s \in A$, $|s| \geq p$ voidaan jakaa kolmeen osaan $s = xyz$ siten, että seuraavat kolme ehtoa ovat voimassa.

1. $xy^iz \in A$ jokaisella $i \in \mathbb{N}$, erityisesti myös kun $i = 0$.
2. $|y| > 0$, eli toistettava osa y ei saa olla tyhjä merkkijono ε.
3. $|xy| \leq p$

Otetaan tästä konkreettinen esimerkki. Kieli

$$A = L(a^*b^*)$$

on säännöllinen ja sillä on täten jokin pumppauspituus p. Tämän kielen kohdalla eräs mahdollinen pumppauspituus on $p = 3$. Tämän pumppauspituuden saa esimerkiksi siitä, että kielen voi tunnistaa deterministisellä äärellisellä automaatilla jossa on kolme tilaa. Tarkastellaan jotain riittävän pitkää kielen A merkkijonoa. Valitaan $s = abbb$. Nyt pumppauslemman nojalla löytyy *jokin* jako $s = xyz$, jolle yllä olevat kolme ehtoa pätevät.

Tarkastellaan merkkijonon s mahdollisia jakoja.

- Kokeillaan ensin seuraavaa jakoa:

$$\begin{aligned} x &= \varepsilon, \\ y &= ab \\ \text{ja } z &= bb \end{aligned}$$

  Nyt ehdot 2 ja 3 ovat voimassa, mutta ensimmäinen ehto ei täyty, sillä esimerkiksi merkkijono $xyyz = ababbb$ ei kuulu kieleen A.

- Ensimmäinen jako ei siis täyttänyt kaikkia pumppauslemman ehtoja. Pumppauslemma ei kuitenkaan takaa, että nämä ehdot täyttyisivät jokaisella jaolla. Ainoa tae on se, että löytyy jokin ehdot täyttävä jako. Tällainen on esimerkiksi seuraava jako:

$$\begin{aligned} x &= a, \\ y &= b \\ \text{ja } z &= bb \end{aligned}$$

Nyt $|xy| = 2 \leq p$, $|y| = 1 > 0$. Entä miltä näyttää merkkijono xy^iz? Tarkastellaan tätä ensin i:n arvoilla 0, 1 ja 2:

$$\begin{aligned} xy^0z &= xz = abb \in A \\ xy^1z &= xyz = abbb \in A \\ xy^2z &= xyyz = abbbb \in A \end{aligned}$$

Ja yleisessä tapauksessa, kun $i \in \mathbb{N}$, niin

$$xy^iz = ab^ibb = ab^{i+2} \in L(a^*b^*) = A$$

Tehtävä 1. Olkoon

$$A = L((ab)^*)$$

säännöllinen kieli. Tällä kielellä on pumppauspituus $p = 2$. Valitaan kielestä merkkijono

$$s = ababab$$

joka on pidempi kuin pumppauspituus $p = 2$. Anna kaikki merkkijonon s jaot $s = xyz$ jotka täyttävät pumppauslemman ehdot 2 ja 3, eli

$$\begin{aligned} &|y| > 0 \text{ ja} \\ &|xy| \leq p. \end{aligned}$$

Tehtävä 2. Mitkä edellisen tehtävän jaoista $s = xyz$ toteuttavat pumppauslemman ensimmäisen ehdon? Ensimmäinen ehto on

$$xy^iz \in A \text{ kaikilla } i \in \mathbb{N}$$

eli keskikohtaa y voi toistaa.

Pumppauslemman käyttäminen

Pumppauslemma on hyödyllinen, koska sillä voidaan näyttää monia kieliä epäsäännöllisiksi. Mutta kuinka tämä onnistuu työkalulla, joka ei puhu mitään epäsäännöllisistä kielistä? Pumppauslemman jos-niin rakenteeseen piiloutuu kuitenkin myös väite epäsäännöllisyydestä. Seuraava lemma on nimittäin yhtäpitävä pumppauslemman kanssa.

Lemma. Olkoon A jokin kieli. Oletetaan lisäksi, että kaikilla luonnollisilla luvuilla $p > 0$ jotain kielen A merkkijonoa s, $|s| \geq p$ ei voida jakaa *millään tapaa* kolmeen osaan $s = xyz$ siten, että seuraavat ehdot olisivat voimassa:

1. $xy^iz \in A$ jokaisella $i \in \mathbb{N}$, erityisesti myös kun $i = 0$.
2. $|y| > 0$, eli toistettava osa y ei saa olla tyhjä merkkijono ε.
3. $|xy| \leq p$

Nyt kieli A on epäsäännöllinen.

Todistus. Olkoon A kieli jolla ei ole pumppauspituutta. Oletetaan vastoin väitettä, että A on säännöllinen. Nyt pumppauslemman nojalla kielellä A kuitenkin tulisi olla pumppauspituus. Siis A ei voi olla säännöllinen. □

Pumppauslemman avulla voidaan siis todistaa epäsäännöllisiksi vain sellaiset kielet, joilla ei ole pumppauspituutta. Tulee kuitenkin muistaa, että joillain epäsäännöllisilläkin kielillä on olemassa pumppauspituus.

Käydään läpi todistus erään jo tutun kielen epäsäännöllisyydelle. Olkoon kieli $A = \{0^n1^n \mid n \in \mathbb{N}\}$. Osoitetaan tämän kielen epäsäännöllisyys yllä todistetun lemman avulla.

Väite. *Kieli* $A = \{0^n1^n \mid n \in \mathbb{N}\}$ *on epäsäännöllinen.*

Todistus. Olkoon p positiivinen luonnollinen luku. Valitsemme merkkijonon $s \in A$, jolla $|s| \geq p$. Käymmä läpi kaikki merkkijonon s jaot osiin $s = xyz$ joilla pumppauslemman ehdot $|xy| \leq p$ ja $y \neq \varepsilon$ ovat voimassa. Tavoittena on valita s niin, että toistettava osa y sisältää aina kielen A epäsäännöllisen ominaisuuden. Tämän ominaisuuden ei tulisi säilyä kun osaa y toistaa monta kertaa tai sen poistaa.

Valitaan merkkijono

$$s = 0^p1^p \in A$$

joka kuuluu kieleen A ja sisältää vähintään p merkkiä. Merkkijonoon s valittiin p kappaletta nollia ja ykkösiä, jotta kaikki ensimmäiset p merkkiä olisivat samoja, tässä tapauksessa nollia. Koska y on jokin ensimmäisen p:n merkin alimerkkijono, y on muotoa 0^k ja kaikki 1-merkit jäävät loppuosaan z. Nyt jos osaa y toistaa, eli muodostaa esimerkiksi merkkijonon $xyyz$, niin 0-merkkien lukumäärä muuttuu. Kaikki 1-merkit ovat kuitenkin loppuosassa z, joten niiden lukumäärä säilyy ennallaan. Joten merkkijono $xyyz$ ei enää toteuta kielen A ehtoa.

Merkkijonon s ensimmäiset p merkkiä ovat nollia, joten kaikki mahdolliset ehtojen 2 ja 3 mukaiset jaot ovat olennaisesti samaa muotoa. Ne voidaan kuvata seuraavasti

$$\begin{aligned} x &= 0^n \\ y &= 0^m,\ m > 0 \\ \text{ja } z &= 0^{p-(m+n)}1^p \end{aligned}$$

missä $n, m \in \mathbb{N}$. Jotta y ei olisi tyhjä merkkijono, täytyy m:n olla positiivinen. Jotta merkkijonon xy pituus olisi korkeintaan p, tulee summan $n + m$ olla korkeintaan p.

Tarkastellaan merkkijonoa, jossa osa y esiintyy kaksi kertaa

$$\begin{aligned} xy^2z &= xyyz \\ &= 0^n0^{2m}0^{p-(m+n)}1^p \\ &= 0^{n+2m+(p-(m+n))}1^p \\ &= 0^{n+2m+p-m-n}1^p \\ &= 0^{p+m}1^p \end{aligned}$$

Koska m on aidosti positiivinen luku, niin merkkijonossa xy^2z on enemmän nollia kuin ykkösiä. Se ei siis kuulu kieleen A. Ehto 1 ei voi olla voimassa, joten lemman nojalla A on epäsäännöllinen. □

Tehtävä 3. Osoita kieli

$$A = \{0^n1^m \mid n, m \in \mathbb{N}, n < m\}$$

epäsäännölliseksi lemman avulla.

Epäsäännöllisyyksien huomaaminen

Kun kielen haluaa näyttää epäsäännölliseksi, pitää siitä huomata jokin ominaisuus joka ei ole säännöllinen. Siis ominaisuus jota ei voi tunnistaa äärellisellä automaatilla. Automaatissa on vain rajoitettu määrä tiloja eikä automaatti voi muistaa mitään muuta, kuin nykyisen tilansa. Erityisesti automaatin muisti on siis ennalta rajattu.

Automaatti ei voi muistaa kaikkia näkemiään merkkejä, tai käsitellyn syötteen pituutta, tai kaikkien nähtyjen a-merkkien lukumäärää.

Äärellisellä automaatilla on aina vakiomäärä tiloja. Esimerkiksi k kappaletta. Nyt jos syötteen pituus on enemmän kuin k, niin automaatin täytyy syötettä läpi käydessä kulkea jonkin syklin läpi. Kun automaatti on tämän syklin ensimmäisessä tilassa, se ei voi muistaa kuinka monta kertaa sykli on kuljettu, tai edes sitä onko syklissä käyty ollenkaan. Syötemerkkijonoa voidaan siis muokata niin, että automaatti kulkee tämän syklin läpi useamman kerran tai ei yhtään kertaa ja automaatti päätyy edelleen samaan lopputilaan.

Yllä todistettiin kieli $\{0^n 1^n \mid n \in \mathbb{N}\}$ epäsäännölliseksi. Tämän kielen tunnistavan automaatin pitäisi muistaa, kuinka monta 0-merkkiä ollaan nähty, jotta se voisi 1-merkkien kohdalla varmistaa niitä olevan yhtä paljon.

Kun tästä kielestä valittiin alkio $s = 0^p 1^p$, käytettiin hyväksi tätä tunnistettua epäsäännöllisyyttä. Kielen ehto vaatii, että nollia ja ykkösiä on aina yhtä monta. Pumppauslemma toisaalta takaa, että säännöllisillä kielillä ensimmäisestä p merkistä löytyy toistettava kohta. Jotain alkuosaa toistamalla saadaan kuitenkin aina merkkijono jossa 0-merkkejä on enemmän kuin 1-merkkejä, eikä kielen ominaisuus siis säily toiston yhteydessä riippumatta siitä, mitä kohtaa ensimmäisestä p merkistä toistettiin.

Tehtävä 4. Osoita kieli

$$A = \{a^n b^n c^n \mid n \in \mathbb{N}\}$$

epäsäännölliseksi lemman avulla.

Vastaoletuksella todistaminen

Edellisen kappaleen lemma sanoo suoraan sen, mitä pumppauslemma kertoo epäsäännöllisistä kielistä. Useimmiten todistukset jonkin kielen epäsäännöllisyydelle käyttävät kuitenkin niin kutsuttua ristiriitatodistusta. Siinä todistus alkaa vastaoletuksella, jossa nimensä mukaisesti oletetaan todistettava väitteen vääräksi. Tästä oletuksesta pyritään osoittamaan jokin ristiriita jo todeksi tiedetyn kanssa.

Jos haluamme osoittaa jonkin kielen epäsäännölliseksi, vastaoletuksemme on, että kieli onkin säännöllinen. Nyt pumppauslemman tulisi päteä kielelle. Käydään läpi esimerkki.

Väite. *Kieli $A = \{ww \mid w \in \Sigma^*\}$ on epäsäännöllinen, missä $\Sigma = \{a, b\}$*

Todistus. Aloitamme todistuksen siis olettamalla vastoin väitettä, että kieli A on säännöllinen. Nyt pumppauslemma sanoo, että kielellä tulee olla jokin pumppauspituus $p \in \mathbb{N}, p > 0$. Emme tiedä mitään p:n arvosta, mutta tiedämme sen olevan pumppauspituus. Siispä merkkijonon $s = a^p b a^p b$ jollekkin jaolle $s = xyz$ tulee päteä kaikki pumppauslemman kolme ehtoa.

Noudattaen ehtoja 2 ja 3 merkkijono s voidaan jakaa seuraavilla tavoilla:

$$\begin{aligned} x &= a^n \\ y &= a^m,\ m > 0 \\ z &= a^{p-(n+m)} b a^p b \end{aligned}$$

Koska oletimme kielen A olevan säännöllinen, tulisi pumppauslemman ensimmäisenkin ehdon olla voimassa. Siis merkkijonon xy^0z tulisi kuulua kieleen. Kuitenkin merkkijono $xy^0z = a^n a^{p-(n+m)} b a^p b = a^{p+m} b a^p b$ ei kuulu kieleen A. Päädyimme siis ristiriitaan pumppauslemman kanssa. Uskomme kaikkien esittämiemme argumenttien pitävän paikkansa. Ainoaksi vaihtoehdoksi siis jää, että tekemämme oletus kielen A säännöllisyydestä oli väärä. Siis kielen A täytyy olla epäsäännöllinen. □

Tehtävä 5. Osoita vastaoletuksella, että kieli

$$A = \left\{ ww^{\mathcal{R}} \mid w \in \Sigma^* \right\}$$

on epäsäännöllinen, missä $\Sigma = \{0, 1\}$.

Tehtävä 6. Osoita vastaoletuksella, että kieli

$$A = \{0^n 1^m 0^n \mid n, m \in \mathbb{N}\}$$

on epäsäännöllinen.

Epäsäännöllisyyden säilyttävät operaatiot

Vaikka säännölliset kielet ovat suljettu hyvin monien operaatioiden suhteen, niin epäsäännöllisyys ei kuitenkaan ole suljettua kovinkaan monien operaatioiden suhteen. Epäsäännöllisetkin kielet ovat kuitenkin suljettuja kahden hyödyllisen operaation, komplementin $\overline{A}$ ja käänteiskielen $A^{\mathcal{R}}$, suhteen.

Välillä on helpompaa näyttää kielen A komplementti $\overline{A}$ tai käänteiskieli $A^{\mathcal{R}}$ epäsäännölliseksi kuin kieli A itse.

Väite. *Jos $A^{\mathcal{R}}$ on epäsäännöllinen, niin myös A on epäsäännöllinen.*

Todistus. Tehdään vastaoletus. Oletetaan vastoin väitettä, että A onkin säännöllinen. Aikaisemmin kurssilla on osoitettu, että tällöin myös $A^{\mathcal{R}}$ on säännöllinen. Tämä on ristiriidassa oletuksen $A^{\mathcal{R}}$ on epäsäännöllinen kanssa.

Siis vastaoletuksen täytyy olla väärä ja kieli A on epäsäännöllinen. □

Tehtävä 7. Näytä, että jos kieli $\overline{A}$ on epäsäännöllinen, niin myös A on epäsäännöllinen.

Väite. *Kieli $A = \{a^m b^n c^n \mid n, m \geq 1\}$ ei ole säännöllinen.*

Todistus. Edellä osoitetun nojalla riittää siis näyttää kieli $A^{\mathcal{R}}$ epäsäännölliseksi, missä

$$A^{\mathcal{R}} = \{c^n b^n a^m \mid n, m \geq 1\}.$$

Tehdään vastaoletus. Oletetaan, että kieli $A^{\mathcal{R}}$ on säännöllinen. Tällöin sillä on jokin pumppauspituus p. Nyt voidaan valita merkkijono

$$s = c^p b^p a \in A^{\mathcal{R}},$$

jolla pätee $|s| \geq p$. Koska pumppauslemman kolmannen ehdon nojalla $|xy| \leq p$, niin valitsimme merkkijonon s näin. Nyt toisettavassa osassa y on pelkästään merkkiä c. Osaa y toistamalla saadaan siis aikaan merkkijono, jossa c-kirjaimia on eri määrä kuin b-kirjaimia. Tämä rikkoo kielen ehtoa, joka vaatii saman määrän c-merkkejä ja b-merkkejä.

Säännöllisten kielten pumppauslemman nojalla s voidaan nyt jakaa kolmeen osaan seuraavasti:

$$\begin{aligned} s &= xyz, \\ |xy| &\leq p \text{ ja} \\ |y| &> 0 \end{aligned}$$

joten

$$\begin{aligned} xy &= c^k \text{ jollain } k \leq p, \\ z &= c^{p-k} b^p a \end{aligned}$$

ja erityisesti

$$y = c^n \text{ jollain } n > 0.$$

Nyt pumppauslemman nojalla myös merkkijonon

$$xz = c^{k-n} c^{p-k} b^p a = c^{p-n} b^p a$$

tulisi kuulua kieleen $A^{\mathcal{R}}$. Nyt kuitenkin $n > 0$, joten

$$xz = c^{p-n} b^p a \notin A.$$

Tämä on ristiriidassa säännöllisten kielten pumppauslemman kanssa, joten kielellä $A^{\mathcal{R}}$ ei ole pumppausominaisuutta. Kieli $A^{\mathcal{R}}$ ei siis voi olla säännöllinen. Täten kieli A ei ole säännöllinen. $\square$

Tehtävä 8. Näytä kieli

$$A = \{abca^n b^n c^n \mid n \in \mathbb{N}\}$$

epäsäännölliseksi.

Komplementtia voi käyttää hyödyksi kielen epäsäännöllisyyden näyttämisessä samalla tavalla.

Tehtävä 9. Kieli A sisältää kaikki merkkijonot aakkostosta $\{0, 1\}$, joissa merkkejä 0 ja 1 on eri määrä. Näytä kieli A epäsäännölliseksi.