<u>מודלים חישוביים - תרגיל 6#</u>

אריאל סטולרמן
כלב אלפרנס
קבוצה 02

(1)

$L_1 = \left\{a^i b^j c^k \,\middle|\, i,j,k \geq 0, i+k=j\right\}$ (a)

להלן דקדוק חסר הקשר לשפה:

$G = (V, \Sigma, S, R)\ where:$

- $V = \{S, T, U\}$
- $\Sigma = \{a, b, c, \varepsilon\}$
- R: $S \to aTbbUc \mid aTb \mid bUc \mid \varepsilon$
  $T \to aTb \mid \varepsilon$
  $U \to bUc \mid \varepsilon$

$L_2 = \left\{a^i b^j \,\middle|\, i,j \geq 0, i \neq j\right\}$ (b)

להלן דקדוק חסר הקשר לשפה:

$G = (V, \Sigma, S, R)\ where:$

- $V = \{S, T, U\}$
- $\Sigma = \{a, b, \varepsilon\}$
- R: $S \to aT \mid Ub$
  $T \to aT \mid aTb \mid \varepsilon$
  $U \to Ub \mid aUb \mid \varepsilon$

$L_3 = \left\{a^i b^j c^k \,\middle|\, i,j,k \geq 0, i \neq j\ or\ j \neq k\right\}$ (c)

$G = (V, \Sigma, S, R)\ where:$

- $V = \{S, T, U, W, W', X, X'\}$
- $\Sigma = \{a, b, c, \varepsilon\}$
- R: $S \to aT \mid Uc$
  $T \to aT \mid bW \mid Xc \mid \varepsilon, U \to Uc \mid aW' \mid X'b \mid \varepsilon$
  $W \to bW \mid bWc \mid \varepsilon, W' \to aW' \mid aW'b \mid \varepsilon$
  $X \to Xc \mid bXc \mid \varepsilon, X' \to X'b \mid aX'b \mid \varepsilon$

(2)

$L_1 = \{a^n b^{2n} c^n \mid n \in \mathbb{N}\}$ (a)

טענה: L_1 אינה חסרת הקשר

נניח כי השפה חסרת הקשר עם קבוע הניפוח p. אזי המילה $w = a^p b^{2p} c^p \in L_1$, המקיימת $|w| \geq p$ ויש לה חלוקה $w = uvxyz$ כנדרש. לפיכך הרצף vxy המקיים $|vxy| \leq p$, $|vy| \geq 1$ יכול להכיל לכל היותר שתי אותיות שונות. אם נבחר $i = 0$ נקבל בכל מקרה אפשרי מילה שאינה מהצורה $a^k b^{2k} c^k$, בסתירה להנחה כי L_1 חסרת הקשר.

$L_2 = \{a^i b^j c^k \mid i,j,k \geq 0, k = min\,\{i,j\}\}$ (b)

טענה: L_2 אינה חסרת הקשר

נניח כי השפה חסרת הקשר עם קבוע הניפוח p. אזי המילה $w = a^p b^p c^p \in L_1$, המקיימת $|w| \geq p$ ויש לה חלוקה $w = uvxyz$ כנדרש. לפיכך הרצף vxy המקיים $|vxy| \leq p$, $|vy| \geq 1$ יכול להכיל לכל היותר שתי אותיות שונות. לשם מניעת בלבול, נסמן את חזקת הניפוח ל- v, y ב-i_0.

להלן אפשרויות הרצף vxy:

- מכיל רק $'a'$: נבחר $i_0 = 0$ ונקבל כי מספר ה-$'a'$ קטן ממש מ-p, ולכן $k \neq min\,\{i,j\}$.
- מכיל $'a','b'$: נבחר $i_0 = 0$ ונקבל שמספר ה-$'a','b'$ קטן מ-p, אך מספר ה-$'c'$ נשאר אותו דבר, ולכן גם כאן $k \neq min\,\{i,j\}$.
- מכיל רק $'b'$: נבחר $i_0 = 0$, הוכחה באופן דומה לסעיף הראשון.
- מכיל $'b','c'$: נבחר $i_0 = 2$ ונקבל כי מספר ה-$'b','c'$ גדול ממש מ-p, ולכן $i = p = min\,\{i,j\}$ ומספר ה-$'c'$ אינו i, לכן גם כאן $k \neq min\,\{i,j\}$.
- מכיל רק $'c'$: נבחר $i_0 = 2$ ונקבל כי מספר ה-$'c'$ גדול ממש מ-p ולכן גדול ממש מ-$min\{i,j\} = p$.

$\Leftarrow$ בכל מקרה שהוא מקבלים לאחר ניפוח/כיווץ מילה שאינה ב-L_1, בסתירה להנחה. לפיכך השפה אינה חסרת הקשר.

$L_3 = \{a^m b^n c^n d^m \mid m,n \in \mathbb{N}\}$ (c)

טענה: L_3 חסרת הקשר

להלן דקדוק חסר הקשר המתאים לשפה:

$G = (V, \Sigma, S, R)\ where:$

- $V = \{S, T\}$
- $\Sigma = \{a, b, c, d, \varepsilon\}$
- R: $\quad S \to aSd \mid T \mid \varepsilon$

  $\quad T \to bTc \mid \varepsilon$

(3)

$L^R \coloneqq Reverse(L) = \{w \mid w^R \in L\}$ (a)

טענה: L רגולרית $\Leftarrow$ L^R רגולרית:

נניח כי L רגולרית, אזי קיים אוטומט לא דטרמיניסטי $N_L = (Q, \Sigma, \delta, q_0, F)$ המקבל אותה. נגדיר NFA אותו נסמן N_L^R המקבל את L^R:

$N_L^R = (Q \uplus \{s\}, \Sigma, \delta', s, F')\ where:$

- $F' = \{q_0\}, q_0\ is\ the\ only\ accepting\ state\ in\ Q \uplus \{s\}$
- $s \notin Q$
- δ': $\quad \forall q \in Q, a \in \Sigma \cup \{\varepsilon\}: \quad \delta'(q, a) = \{q' \mid q \in \delta(q', a)\}$

  $\delta'(s, \varepsilon) = \{q_F \mid q_F \in F\}$

הסבר: נניח $w = a_1 a_2 \dots a_n \in L$, אזי קיים ב-$N_L$ רצף מצבים מקבל $q_0 \to q_1 \to \cdots \to q_F$, $q_F \in F$. אזי באוטומט החדש N_L^R קיים רצף מצבים $s \xrightarrow{\varepsilon} q_F \to \cdots \to q_1 \to q_0$ ($q_0 \in F'$) המקבל את המילה $w^R = a_n \dots a_2 a_1$, במעבר על

קשתות δ' ההפוכות לקשתות δ עליהן רצנו ב-N_L. באופן דומה ניתן להראות אוטומט $N_L^{R^R}$ המקבל את $w^{R^R} = w$, ולכן מתקיים $w \in L \Leftrightarrow w^R \in L(N_L^R)$, ולכן $L(N_L^R) = L^R$. לפיכך L^R היא שפה רגולרית.

טענה: L **חסרת הקשר** $\Leftarrow L^R$ **חסרת הקשר**:

נניח כי L חסרת הקשר, אזי קיים לה דקדוק G המזהה אותה. נבנה ל-L^R דקדוק G^R המזהה אותה באופן הבא:

$$G^R = (V, \Sigma, S, R')\ where: R' = \{A \to \alpha \mid A \in V, \alpha \in (V \cup \Sigma)^*, (A \to \alpha^R) \in R\}$$

נניח כי $w \in L$, אזי יש לה גזירה לפיה בעץ הגזירה: מ-S יש שני תתי עצים A,B. נניח באינדוקציה על גודל עץ הגזירה כי כל אחד מתתי העצים האלה ניתנים להיפוך. לפי הגדרת R', נהפוך את סדר תתי העצים של S, כלומר נקבל B,A, וקיבלנו את w^R. צעד הבסיס של האינדוקציה טרוויאלי. באופן דומה ניתן לבנות G^{R^R} המזהה את $w^{R^R} = w \in L$ ולכן נקבל שמתקיים $w \in L \Leftrightarrow w^R \in L(G^R)$, ולכן $L(G^R) = L^R$. לפיכך L^R היא שפה חסרת הקשר.

(b) $PP(L) = \{w \mid w \in L, \exists k \in \mathbb{N}\ s.t. w = w_1 w_2 \dots w_k, \forall 0 < i < k: w_1 \dots w_i \notin L\}$

טענה: L **רגולרית** $\Leftarrow PP(L)$ **רגולרית**:

תהי L שפה רגולרית, אזי קיים DFA: $A_L = (Q, \Sigma, \delta, q_0, F)$ המקבל אותה. נגדיר DFA אותו נסמן $A_{PP(L)}$ המקבל את $PP(L)$ באופן הבא:

$A_{PP(L)} = (Q \uplus \{q'\}, \Sigma, \delta', q_0, F), \qquad where:$

- $q' \notin Q, q'$ $is\ a\ non-accepting\ state$
- δ': $\quad \forall q \in Q \backslash F, a \in \Sigma: \quad \delta'(q, a) = \delta(q, a)$

  $\forall q_F \in F, a \in \Sigma: \quad \delta'(q_F, a) = q'$

  $\forall a \in \Sigma \quad \delta'(q', a) = q'$

האוטומט החדש $A_{PP(L)}$ מקבל את כל המחרוזות ה"ראשונות" המתקבלות ע"י A_L, כלומר: לכל $w = w_1 \dots w_n \in L$ מתקיים כי $w' \in PP(L)$ כאשר $i = min\{j \mid w_1 \dots w_j \in L\}$, $w' = w_1 \dots w_i$, $\forall i < k \le n: w_1 \dots w_k \notin PP(L)$.

נניח כי $w \in PP(L)$, אזי כל $prefix$ של w (כמוגדר) אינו ב-$PP(L)$. מבניית $A_{PP(L)}$ ברור כי אם הוא מקבל מילה כלשהי w אזי $w \in L$, כיוון שמקבל אותה לפי אותו מעבר מצבים כמו ב-A_L, וכמו כן כל $prefix$ של w אינו מתקבל על ידו. אם נניח בשלילה שקיים ל-w איזשהו $prefix$ שנסמנו w' המתקבל ע"י $A_{PP(L)}$, אזי כל הארכה של w' לא מתקבלת ע"י $A_{PP(L)}$, ובפרט w בסתירה להנחה. לפיכך $w \in L(A_{PP(L)})$.

נניח כי $w \in L(A_{PP(L)})$, אזי $w \in L$ ולא קיים ל-w איזשהו $prefix$ מ-L כמוגדר המתקבל ע"י $A_{PP(L)}$, מההסבר הנ"ל. לפיכך $w \in PP(L)$.

$\Leftarrow PP(L) = L(A_{PP(L)})$, לכן $PP(L)$ רגולרית ולפיכך שפות רגולריות סגורות תחת $PP(L)$.

טענה: שפות חסרות הקשר אינן סגורות תחת פעולת $PP(L)$:

תהי $L = \{a^i b^j c^k \mid i \le j\ or\ i \le k\}$. L היא שפה חסרת הקשר, להלן דקדוק חסר הקשר G המזהה אותה:

$G = (V, \Sigma, S, R), where\ V = \{S, T_1, T_2, X, Y\}\ and\ R$:

$S \to T_1 X \mid T_2 X$

$T_1 \to a T_1 b \mid Y, T_2 \to a T_2 c \mid Y$

$X \to cX \mid \varepsilon, Y \to bY \mid \varepsilon$

יהי p קבוע הניפוח של L. נתבונן במילה $w=a^p b^{p-1} c^p$. ברור שהמילה ב-L, וכמו כן ברור שהיא ב-$PP(L)$: תנאי השיוך ל-L מתקיים בגלל שמספר ה-$'c'$ הוא מינימלי (לא יכול להתקיים ע"י מס' ה-$'b'$ שקטן ממש ממספר ה-$'a'$). אם נוריד את מספר ה-$'c'$ המילה לא תהיה ב-L, כמו כן אם נוריד לחלוטין ונשאר רק עם ה-$'b'$ הקיימים, וכמובן אם נוריד וניוותר רק עם $'a'$ (לא נתייחס למילה הריקה). לפיכך כל תחילית של המילה הזו לא מקיימת את התנאי ולכן אינה ב-L. נחלק את $w=uvxyz$ לפי כללי למת הניפוח כך ש-$|vxy|\le p$, $|vy|\ge 1$, וברור כי $|w|\ge p$. נחלק למקרים:

- אם vxy מכיל רק $'a'$: נבחר $i=2$ ונקבל כי מספר ה-a גדול ממש ממספר ה-$'c','b'$, לכן המילה אינה ב-L ובפרט אינה ב-$PP(L)$.
- אם v מכיל $'a'$, y מכיל $'b'$: אם v מכיל יותר $'a'$ מאשר y מכיל $'b'$, נבחר $i=3$ כך שנקבל יותר $'a'$ מ-$'c','b'$, ולכן מילה שאינה ב-L, ובפרט אינה ב-$PP(L)$. אם המצב הפוך, גם נבחר $i=3$ ובכך נהפוך את מספר ה-$'a'$ גדול מ-p ונגדיל את מספר ה-$'b'$ כך שיהיה גדול ממספר ה-$'a','c'$ – כלומר תנאי השיוך ל-L נקבע על ידו. כעת, למילה שנקבל נוכל לקחת את התחילית המכילה רק $'a','b'$ והיא שייכת ל-L, לכן המילה המקורית לאחר ניפוח אינה ב-$PP(L)$.
- אם vxy מכיל רק $'b'$: נבחר $i=3$ ובכך נגדיל את מספר ה-$'b'$ בדומה לקודם. נקבל מילה שיש לה תחילית ב-L, ולכן המילה המקורית לאחר ניפוח אינה ב-$PP(L)$.
- אם v מכיל $'b'$, y מכיל $'c'$: נבחר $i=0$ וכך נקבל שמספר ה-$'a'$ גדול ממש ממספר ה-$'c','b'$ ולכן המילה אינה ב-L.
- אם vxy מכיל רק $'c'$: נבחר שוב $i=0$ ונקבל שמספר ה-$'a'$ גדול ממש ממספר ה-$'c','b'$.

$\Leftarrow$ לכל חלוקה העונה על תנאי למת הניפוח, ניתן לבחור i כך שלאחר הניפוח המילה אינה ב-$PP(L)$ $\Leftarrow$ שפות רגולריות אינן סגורות תחת הפעולה.

(c) $Half(L)=\{x\mid \exists y\in\Sigma^*\ s.t. |x|=|y|, xy\in L\}$

תשובה לגבי שפות רגולריות בסוף התרגיל.

טענה: שפות חסרות הקשר אינן סגורות תחת פעולת $Half$:

תהי L המוגדרת באופן הבא: $L=\{a^n b^n c^m d^{3m}\mid n,m\in\mathbb{N}\}$. ברור כי L חסרת הקשר, שכן היא שרשור של שתי שפות חסרות הקשר: $L=\{a^n b^n\mid n\in\mathbb{N}\}\cup\{c^m d^{3m}\mid m\in\mathbb{N}\}$. כל מילה בשפה L היא מאורך $2n+4m$, ומכאן כל מילה בשפה $Half(L)$ תהיה מאורך $n+2m$. נסווג את המילים ב-$Half(L)$ באופן הבא:

- עבור $n\ge 2m$: מילים מהצורה $a^n b^{2m}$.
- עבור $m\le n<2m$: מילים מהצורה $a^n b^n c^{2m-n}$.
- עבור $n<m$: מילים מהצורה $a^n b^n c^m d^{m-n}$.

נניח בשלילה כי $Half(L)$ חסרת הקשר בעלת קבוע ניפוח p. נתבונן במילה $w=a^p b^p c^p$, מצורה 2. מתקיים כי $m=n=p$, ולכן המילה עומדת בתנאי $m\le n<2m$. כעת נבחר למילה חלוקה $w=uvxyz$ העומדת בתנאי למת הניפוח כך ש-$|vxy|\le p$, $|vy|\ge 1$, וברור כי $|w|\ge p$. נחלק למקרים:

- vxy מכיל רק $'a'$: נבחר $i=2$ ונקבל שמספר ה-$'a'$ שונה ממספר ה-$'b'$ אך המילה אמורה להשתייך לצורה 2 – יש בה הופעות של $'c'$ ואין בה הופעות של $'d'$, ולכן המילה אינה ב-$Half(L)$.
- v מכיל $'a'$, y מכיל $'b'$: נבחר $i=0$, ונקטין גם את מספר ה-$'a'$ וגם את מספר ה-$'b'$. אם לא נקטין אותם במידה שווה נקבל שוב שמספר ה-$'a'$ שונה ממספר ה-$'b'$ בסתירה לצורה 2. אם כן הקטנו אותם במידה שווה, יתקיים כי לכל היותר מספר ה-$'a'$ ומספר ה-$'b'$ יהיה $p-1$, כלומר $w=a^{p-1}b^{p-1}c^p$. במקרה זה מתקיים כי

$n = p - 1$ ומכאן נחשב את m: $2m - n = p \Leftrightarrow 2m = 2p - 1 \Leftrightarrow m = p - 0.5$. במקרה זה יתקיים כי $n < m$, בסתירה לצורה 2 (הופעות של $'a','b','c'$ וללא הופעות $'d'$). לפיכך המילה שתתקבל אינה ב-$Half(L)$.

- vxy מכיל רק $'b'$: בדומה למקרה הראשון, נבחר $i = 2$ ונגדיל את מספר ה-$'b'$ כך שיהיה שונה ממספר ה-$'a'$, ונקבל מילה שאינה ב-$Half(L)$.
- v מכיל $'b'$, y מכיל $'c'$: נבחר $i = 0$ ונקטין את מספר ה-$'b'$, ונקבל מספר $'b'$ שונה ממספר ה-$'a'$ (עדיין ישנן הופעות של $'c'$, כיוון שמלכתחילה היו לנו p הופעות של $'c'$ – לכל היותר $y = c^{p-1}$. לכן עדיין בצורה 2).
- vxy מכיל רק $'c'$: נבחר $i = 2$ ונקבל: $n = p$ ולכן $2m - n = 2p \Rightarrow 2m = 3p \Rightarrow m > p$. לפיכך $m > n$, בסתירה לצורה 2, לכן המילה אינה ב-$Half(L)$.

$\Leftarrow$ לכל חלוקה העונה על תנאי למת הניפוח, ניתן לבחור i כך שלאחר הניפוח המילה אינה ב-$Half(L)$ $\Leftarrow$ שפות חסרות הקשר אינן סגורות תחת הפעולה.

(4)

יהי G דקדוק חסר הקשר ב-CNF. להלן הוכחה כי לכל $w \in L(G)$, $|w| = n$ $(n > 0)$, w נגזרת ב-$2n - 1$ צעדים בדיוק:

הוכחה באינדוקציה:

עבור $n = 1$: הגזירה חייבת להתבצע ב-$2 \cdot 1 - 1 = 1$ פעולות בדיוק. W חייבת להגזר מהצעד הבודד $S \rightarrow a$, עבור $a \in \Sigma, a \neq \varepsilon$ כלשהו. כל גזירה אחרת תהיה מהצורה $S \rightarrow AB$ עבור $A, B \in V \backslash \{S\}$ כלשהם, וזה יבטיח כי גודל w יהיה לכל הפחות 2, בסתירה לקביעת גודלה להיות 1.

נניח נכונות עבור $n = k$, נוכיח נכונות עבור $n = k + 1$: תהי $w \in L(G)$ המקיימת $|w| = k + 1$, נראה שניתן לבנות אותה ב-$2(k + 1) - 1 = 2k + 1$ צעדים. נסתכל על $w' = w_1 \dots w_k$. נטען כי ניתן להגיע לתחילית זו של w באמצעות $2k - 1$ צעדים בדיוק: נסתכל על המצב בו $w = w_1 \dots w_{k-1}AB$ $(w_i \in \Sigma \backslash \{\varepsilon\}, A, B \in V \backslash \{S\})$, זהו המצב ממנו נגזר $w = w_1 \dots w_k B$ ולאחר מכן $w = w_1 \dots w_{k+1}$, היא המילה הסופית עליה מסתכלים. נגדיר דקדוק חסר הקשר G' באופן הבא: G' זהה ל-G פרט לכך שלכל $(T \rightarrow AB) \in R$ ב-G, $T \in V$ ו-A,B כמוגדרים מעלה, יהיו קיימים בנוסף ב-R' גם החוקים מהצורה $T \rightarrow w_k$, כאשר w_k היא כמוגדרת לעיל. כיוון שהוספנו ל-G' חוקים מצורה מותרת ב-CNF, גם G' הוא מצורת CNF. כיוון שב-G יכולנו לגזור את המילה w, ובפרט להגיע למצב $w_1 \dots w_{k-1}AB$, ב-G' אנו יכולים להחליף את הגזירה האחרונה בלבד ולהגיע ל-$w' = w_1 \dots w_k$. כיוון ש-G' היא מצורת CNF ו-$|w'| = k$, ע"פ הנחת האינדוקציה w' נגזרת ב-$2k - 1$ צעדים בדיוק ב-G'. כיוון שכל שעשינו כדי להגיע אליה ב-G' היה להחליף את הגזירה האחרונה, נובע כי ע"י G ניתן להגיע ל-$w_1 \dots w_{k-1}AB$ גם כן ב-$2k - 1$ צעדים בדיוק. כפי שהראינו, ישנו צורך בעוד 2 צעדים בדיוק כדי להגיע ל-$w = w_1 \dots w_{k+1}$, ולכן מספר הצעדים הכולל הוא בדיוק $2k + 1$, כנדרש.

(5)

(a)

יהי A אוטומט סופי דטרמיניסטי. טענה: $L(A)\ is\ infinite \Leftrightarrow \exists e \in L(A) s.t. |Q| < |w| \leq 2 \cdot |Q|$

נניח תחילה כי $L(A)$ היא אינסופית. כיוון שהיא אינסופית, חייב להיות איזשהו מעגל על Q, כלומר מעגל העובר דרך קבוצת מצבים מסויימת $Q' \subset Q$, $1 \leq |Q'| \leq |Q|$, ובנוסף מעגל זה חייב להשתתף במסלול מקבל (כלומר לפחות מצב אחד בו שייך למסלול מקבל). אם לא היה ב-A לפחות מעגל אחד כזה, $L(A)$ היתה שפה סופית, כיוון שהיה מספר סופי של מסלולים מקבלים ב-A.

נבחר אם כן מעגל אחד (ואולי יחיד) כזה ב-A ונסמן את גודלו $k = |Q'|$ (כלומר מכיל k מצבים). נבחר מסלול מקבל כלשהו החופף עם המעגל. כעת, לבניית w נתחיל ממילה כלשהי w' (שאינה ε) המתקבלת ע"י אותו מסלול נטול מעגלים מקבל שבחרנו קודם. נסמן את אורך המילה ב-m, כאשר $1 \leq m \leq |Q| - 1$. כדי להגדיל את w' עד לקבלת w כנדרש, נרוץ במסלול המקבל את w' עד הגעה למצב הראשון במסלול החופף עם המעגל, נסמן מצב זה q_c (יתכן שזהו q_0, כלומר ישר נכנסנו למעגל). נרוץ במעגל j פעמים (מיד נקבע את j) עד שנגיע ל-j הנדרש, ולאחר מכן נמשיך מ-q_c על המסלול המקבל עד סיום. סה"כ הצעדים = אורך w יהיה:

$$\underbrace{c}_{\text{מספר צעדים עד להגעה לתחילת המעגל}} + \underbrace{j \cdot k}_{\text{מספר צעדים בריצה במעגל}} + \underbrace{m - c}_{\text{מספר הצעדים שנותר עד קבלה}} = m + j \cdot k$$

מכאן ניתן לראות כי לכל $1 \leq m \leq |Q| - 1$, $1 \leq k \leq |Q|$ ניתן לבחור $j \in \mathbb{N}$ כך ש-$|Q| < |w| \leq 2 \cdot |Q|$. נראה כמה מקרי הקצה:

- עבור $k = 1, m = 1$: נבחר $j = |Q|$ ונקבל כי $|w| = 1 + |Q| \cdot 1 = |Q| + 1$, מקיים את התנאי. במקרה בו Q היא הקטנה ביותר שניתן, כלומר שגודלה 1, נקבל $|w| = 2 \cdot |Q|$, גם כנדרש.
- עבור $k = |Q|, m = |Q| - 1$ נבחר $j = 1$ ונקבל כי $|w| = |Q| - 1 + 1 \cdot |Q| = 2 \cdot |Q| - 1$, מקיים את התנאי.

לסיכום, הוכחנו כי עבור $L(A)$ אינסופית ניתן למצוא $w \in L(A)$ כך ש-$|Q| < |w| \leq 2 \cdot |Q|$.

נניח כעת כי ב-$L(A)$ קיימת מילה w כך ש-$|Q| < |w| \leq 2 \cdot |Q|$. כיוון שקיימת כזו מילה, אזי מגיעים אליה לפחות ב-$|Q|$ צעדים, אך אם $L(A)$ היתה סופית, המילה הארוכה ביותר שניתן היה לבנות (פוטנציאלית) היא בגודל $|Q| - 1$ (ללא חזרות על מצבים, אחרת ישנו מעגל $\Leftarrow$ השפה אינסופית, כפי שמוכח בסוף). לפיכך, כדי לבנות את w חייבים לעבור לפחות במצב אחד מתוך Q יותר מפעם אחת, ומכאן שקיים מעגל ב-A. לכן, ניתן לעבור במעגל $j \in \mathbb{N}$ פעמים בדרך למצב המקבל, ולכן לקבל $\aleph_0$ מילים מ-$L(A) \Leftarrow L(A)$ שפה אינסופית, כנדרש.

(b)

לאור סעיף (a), ניתן לבנות כעת אלגוריתם הבודק האם אוטומט סופי דטרמיניסטי A מקבל שפה אינסופית:

$isInfinit(A) \coloneqq$

$\quad for\ i \coloneqq |Q| + 1\ to\ 2 \cdot |Q|\ do:$

$\quad\quad for\ each\ w \in \{w \mid |w| = i, w \in \Sigma^*\}\ do: \quad //for\ each\ i\ there\ are\ |\Sigma|^i\ different\ words$

$\quad\quad\quad if\ A\ accepts\ w\ return\ T$

$\quad return\ F$

נכונות:

$isInfinite(A)$ פונקציה חשיבה, כיוון שמבצעת מספר סופי של פעולות: לכל A שהוא DFA, קבוצת המצבים Q היא סופית, כמו גם Σ, לפיכך מספר המילים שנבדוק יהיה סופי: כל המילים האפשריות מגודל $|Q| + 1$ ועד גודל $2 \cdot |Q|$ מעל השפה Σ (סה"כ: $\sum_{i=|Q|+1}^{2 \cdot |Q|} |\Sigma|^i$ מילים שונות – מספר סופי). כל מילה שבודקים היא סופית ולכן ישנה תשובה האם A מקבל אותה או לא.

כעת, לפי סעיף (a), אם קיימת מילה מאורך כזה, אזי $L(A)$ אינסופית, ולכן כאשר נתקל במילה ש-A מקבל נחזיר T. אחרת, נחזיר F.

(c)

יהי G דקדוק חסר הקשר מצורת CNF. טענה: $L(G) is\ infinte \Leftrightarrow \exists w \in L(G) s.t.\ 2^{|V|} < |w| \leq 2^{|V|+1}$

נניח כי $L(G)$ היא סופית, ונראה כי לכל $w \in L(G)$ מתקיים כי $|w| \leq 2^{|V|}$. אם $L(G)$ היא סופית, אזי לא קיים ב-R מסלול $A \Rightarrow^* AB$ או $A \Rightarrow^* BA$ עבור $A, B \in V$ – כלומר: לא קיימת דרך לחזור אל משתנה שכבר גזרנו דרך אותו מסלול גזירה. לו היה ניתן לחזור אל אותו משתנה דרך אותו מסלול גזירה, ניתן היה לחזור על אותו קטע מעגלי שוב ושוב j פעמים עבור כל $j \in \mathbb{N}$ - ובכך לקבל כי $|L(G)| \geq \aleph_0$.

אם כן, ניתן לסדר את V בסדר $\{S, A_1, A_2, \dots, A_k\}$ כך שלכל i מתקיים: $A_i \Rightarrow^* A_l A_m$ כאשר $m, l > i$ הוא מסלול גזירה אפשרי (יתכן וקיימים חוקי גזירה ב-R המאפשרים גזירה זו), וכמו כן מתקיים: $A_i \Rightarrow^* A_l A_m$ כאשר $m, l \leq i$ הוא מסלול גזירה לא אפשרי (לא קיימים ב-R חוקי גזירה המאפשרים גזירה זו). כלומר: נסדר את V בצורה כזו שנוכל לגזור רק "קדימה" לפי סדר זה. קיים לפנינו "גרף חסר מעגלים" (כל משתנה A_i הוא צומת בגרף והמעברים בין המשתנים הם קשתות הגרף), ולכן קיים בו מיון טופולוגי כמתואר.

כיוון ש-G הוא מצורת CNF, הפוטנציאל המקסימלי למילה באורך הגדול ביותר יהיה אם יהיו ב-R חוקי הגזירה: $S \to A_1 A_1$, $\forall 1 \leq i < k: A_i \to A_{i+1} A_{i+1}$, $A_k \to a\ (a \in \Sigma \backslash \{\varepsilon\})$. זאת כיוון שבהינתן חוקי הגזירה האלה נוכל להגיע לגזירה מקסימלית ללא חזרה על משתנה שכבר גזרנו באותו מסלול גזירה, כלומר ללא מעגלים. גזירה מקסימלית תביא בהכרח לאורך מילה מקסימלי, שכן כל גזירת משתנה מ-V לצמד משתנים מ-V מגדילה את גודל המילה הנגזרת.

נסתכל כעת על גודל המילה שקיבלנו:

$$\underbrace{S}_{2^0\ times} \to \underbrace{A_1 A_1}_{2^1\ times} \to \underbrace{A_2 A_2 A_2 A_2}_{2^2\ times} \to \cdots \to \underbrace{A_k \dots A_k}_{2^{|V|}\ times} \to w, where{:}\ |w| = 2^{|V|}$$

לפיכך גודל המילה הפוטנציאלי המקסימלי עבור $L(G)$ סופית יהיה $2^{|V|}$. לכן: $L(G)$ סופית $\Leftarrow$ לא קיימת $w \in L(G)$ כך ש-$2^{|V|} < |w| \leq 2^{|V|+1}$, ומכאן: אם קיימת כזו $w \in L(G)$, אזי $L(G)$ היא שפה אינסופית.

כעת נניח כי $L(G)$ היא אינסופית, ונראה שקיימת $w \in L(G)$ כנדרש. אם $L(G)$ אינסופית, אזי כאמור קיים לפחות משתנה אחד $A \in V$ עבורו קיים מסלול גזירה: $A \Rightarrow^* AB$ או $A \Rightarrow^* BA$ עבור $B \in V$. כמו כן, ניתן בהכרח להגיע מ-A בגזירה כלשהי לטרמינלים (A "נמצאת על מסלול מקבל"). כפי שהוכח במקרה בו $L(G)$ סופית, המילה הגדולה ביותר שהיתה יכולה לקבל ללא חזרה על אותו משתנה יותר מפעם אחת באותו מסלול גזירה היתה מגודל לכל היותר $|w| = 2^{|V|}$. לפיכך, נוכל לבחור לעבור בעת הגזירה במעגל העובר דרך A פעם אחת, ופעם זו תתרום לכל היותר גם היא $2^{|V|}$ לאורך המילה שתתקבל. לכן נקבל: $|w| \leq 2 \cdot 2^{|V|} = 2^{|V|+1}$. בכל מקרה בו המעגל קטן יותר (תורם פחות טרמינלים למילה), נוכל לבחור לעבור דרך A מספר פעמים שנדרש ע"מ לקבל $|w| > 2^{|V|}$, כנדרש.

(d)

לאור סעיף (c), ניתן לבנות כעת אלגוריתם הבודק האם דקדוק חסר הקשר G מקבל שפה אינסופית:

$isInfinite(G) \coloneqq$

  $for\ i \coloneqq 2^{|V|} + 1\ to\ 2^{|V|+1}\ do{:}$

    $for\ each\ w \in \{w \mid |w| = i, w \in \Sigma^*\}\ do{:}$ $//for\ each\ i\ there\ are\ |\Sigma|^i\ different\ words$

      $if\ G\ recognizes\ w\ return\ T$

  $return\ F$

הצעה למימוש $recognizes$: נקח את G שהוא מצורת CNF (ועבור G שאינו מצורת CNF נקח G' שקול לו מצורת CNF), וידוע שכל מילה שאורכה n ניתנת לגזירה ב-$2n-1$ צעדים בדיוק ע"י G מצורת CNF. לפיכך, עבור w מילה כלשהי מאורך i, ניתן לבדוק את כל המילים באורך זה הנגזרות ע"י G ולבדוק האם w היא אחת מהן. אם כן, נחזיר T, אחרת נחזיר F.

נכונות:

בדומה לסעיף (b), $isInfinite(G)$ פונקציה חשיבה, כיוון שמבצעת מספר סופי של פעולות: לכל G שהוא CFG בצורת CNF, קבוצת המשתנים V היא סופית, כמו גם Σ, לפיכך מספר המילים שנבדוק יהיה סופי: כל המילים האפשריות מגודל $2^{|V|}+1$ ועד גודל $2^{|V|+1}$ מעל השפה Σ (סה"כ: $\sum_{i=2^{|V|}+1}^{2^{|V|+1}}|\Sigma|^i$ מילים שונות – מספר סופי). כל מילה שבודקים היא סופית ולכן ישנה תשובה האם G מזהה אותה או לא (לפי האלגוריתם המוצע לזיהוי מילה ע"י G).

כעת, לפי סעיף (c), אם קיימת מילה מאורך כזה, אזי $L(G)$ אינסופית, ולכן כאשר נתקל במילה ש-G מזהה נחזיר T. אחרת, נחזיר F.

(e)

להלן אלגוריתם אשר בהינתן A אוטומט סופי דטרמיניסטי, קובע האם $|L(A)|=2009$:

$is2009(A) :=$

$let\ F: N \rightarrow \Sigma^*\ be\ an\ enumerator\ for\ all\ words\ in\ \Sigma^*, alphabetically\ ordered$

$if\ isInfinite(A)\ or\ (\sum_{i=0}^{|Q|-1}|\Sigma|^i < 2009)\ then\ return\ F$

$counter \leftarrow 0$

$for\ i := 0\ to\ \sum_{i=0}^{|Q|-1}|\Sigma|^i\ do:$

$if\ A\ accepts\ F(i)\ then\ counter \leftarrow counter + 1$

$if\ counter > 2009\ then\ return\ F$

$if\ counter < 2009\ then\ return\ F$

$else\ return\ T$

נכונות:

$is2009(A)$ חשיבה: קיים אנומרטור ל-Σ^* שכן זהו א"ב סופי, וניתן למנותו בסדר אלפאביתי. כמו כן, הראנו כי $isInfinite(A)$ היא פונקציה חשיבה, וכי $\sum_{i=0}^{|Q|-1}|\Sigma|^i$ הוא מספר סופי ולכן חשיב. לפיכך, גם הלולאה הינה סופית, וכל שאר הפעולות ברורות וחשיבות.

תחילה נבדוק האם השפה אינסופית, או האם מספר המילים הפוטנציאלי שהשפה מקבלת (מקסימום המילים שיתכן ותקבל בהינתן Σ, Q לקלט A אם השפה סופית) קטן מ-2009. בשני מקרים אלה ניתן לקבוע מיד כי $|L(A)| \neq 2009$. כעת, נרוץ על כל המילים האפשריות שגודלן הוא מ-0 ועד $|Q|-1$ (כל מילה מגודל $|Q|$ ומעלה בודאות לא תתקבל ע"י A), ונמנה את מספר המילים ש-A מקבל מתוכן. אם עברנו בדרך את ה-2009 מילים, ניתן לקבוע שוב כי $|L(A)| \neq 2009$. בתום הלולאה נבדוק רק שאין לנו פחות ממש מ-2009, שכן אז נחזיר F. אם אין לנו פחות מ-2009, והרי אין לנו יותר מ-2009, אזי מתקיים כי $|L(A)| = 2009$ ויוחזר T, כנדרש.

(f)

באופן דומה לסעיף (e), להלן אלגוריתם אשר בהינתן G דקדוק חסר הקשר מצורת CNF, קובע האם $|L(G)| = 2009$:

$is2009(G) :=$

$let\ F: N \to \Sigma^*\ be\ an\ enumerator\ for\ all\ words\ in\ \Sigma^*, alphabetically\ ordered$

$if\ isInfinite(G)\ or\ (\sum_{i=0}^{2^{|V|}} |\Sigma|^i < 2009)\ then\ return\ F$

$counter \leftarrow 0$

$for\ i := 0\ to\ \sum_{i=0}^{2^{|V|}} |\Sigma|^i\ \ do:$

$if\ G\ recognizes\ F(i)\ then\ counter \leftarrow counter + 1$

$if\ counter > 2009\ then\ return\ F$

$if\ counter < 2009\ then\ return\ F$

$else\ return\ T$

נכונות:

$is2009(G)$ חשיבה: קיים אנומרטור ל-Σ^* שכן זהו א״ב סופי, וניתן למנותו בסדר אלפאביתי. כמו כן, הראנו כי $isInfinite(G)$ היא פונקציה חשיבה, וכי $\sum_{i=0}^{2^{|V|}} |\Sigma|^i$ הוא מספר סופי ולכן חשיב. לפיכך, גם הלולאה הינה סופית, וכל שאר הפעולות ברורות וחשיבות.

תחילה נבדוק האם השפה אינסופית, או האם מספר המילים הפוטנציאלי שהשפה מקבלת (מקסימום המילים שיתכן ותזהה בהינתן Σ, V לקלט G <u>אם השפה סופית</u>) קטן מ-2009. בשני מקרים אלה ניתן לקבוע מיד כי $|L(G)| \neq 2009$. כעת, נרוץ על כל המילים האפשריות שגודלן הוא מ-0 ועד $2^{|V|}$ (כל מילה מגודל $2^{|V|} + 1$ ומעלה בודאות לא תזוהה ע״י G), ונמנה את מספר המילים ש-G מזהה מתוכן. אם עברנו בדרך את ה-2009 מילים, ניתן לקבוע שוב כי $|L(G)| \neq 2009$. בתום הלולאה נבדוק רק שאין לנו פחות ממש מ-2009, שכן אז נחזיר F. אם אין לנו פחות מ-2009, והרי אין לנו יותר מ-2009, אזי מתקיים כי $|L(G)| = 2009$ ויוחזר T, כנדרש.

(6)

(a) תהי שפה: $L = \{3^n 0^i 1^j 2^k \mid if\ n = 1\ then\ i = j = k\}$

להלן הוכחה שלא ניתן להוכיח באמצעות למת הניפוח ש-L אינה חסרת הקשר:

תחילה נחלק את השפה לשתי קבוצות:

$L = L_1 \cup L_2 := \{30^k 1^k 2^k \mid k \in \mathbb{N}\} \cup \{3^n 0^i 1^j 2^k \mid n, i, j, k \in \mathbb{N}, n \neq 1\}$

נניח כי L היא שפה חסרת הקשר בעלת קבוע ניפוח $p > 0$. כיוון ש-$L = L_1 \cup L_2$ אזי p הוא גם קבוע הניפוח של L_1, L_2.

נראה כי כל מילה בשפות אלו שומרת על משפט הניפוח. נסתכל תחילה על L_1:

נסתכל על כל המילים w ב-L_1 המקיימות $|w| \geq p$: $|w| = |30^k 1^k 2^k| = 1 + 3k \geq p \Leftrightarrow k \geq \frac{p-1}{3}$. נבחר ל-$w$ את החלוקה הבאה: $w = uvxyz$ כאשר: $u = \varepsilon, v = 3$ ו-xyz מתחלק על פני $0^k 1^k 2^k$ באופן כלשהו המקיים $vxy \leq p$.

חלוקה זו אפשרית כיוון ש: במקרה בו k מינימלי (כלומר שווה ל-$\frac{p-1}{3}$), ניתן לבחור $xyz = 0^k 1^k 2^k$ ומתקיים:

$|vxy| \leq |vxyz| = 1 + 3 \cdot \frac{p-1}{3} = p \leq p$. במקרה בו k גדול ניתן פשוט לבחור איזשהו צירוף מיתרת w עד גודל p

שיהווה את vxy. כעת אם ננפח את המילה לכל i שהוא נקבל: עבור $i = 1$ נשאר עם אותה מילה (טרוויאלי); עבור כל i אחר נקבל מילה מהצורה $3^i0^h1^l2^m$ כאשר $i \neq 1$, h, l, m כלשהם. מילה זו שייכת ל-L_2 ולכן שייכת ל-L.

נסתכל על כל המילים w ב-L_2 המקיימות $|w| \geq p$: אם $n = 0$ אז $w = 0^i1^j2^k$, עבור $i + j + k \geq p$ כלשהם. במקרה זה נקח את החלוקה $w = uvxyz$ כך ש-$u = \varepsilon$, v יהיה התו הראשון במחרוזת, ושארית החלוקה תקיים $|vxy| \leq p$ (אפשרי כיוון ש-$|uvxzy| = 1 + 1 + |xyz| = |w| \geq p \Rightarrow |xyz| \geq p - 2$, וניתן לבחור מתוך מינימום $p - 2$ התווים הנותרים מספר כלשהו של תווים שיקיים את הנדרש). עבור החלוקה המוגדרת נקבל כי לכל i נקבל מילה מהצורה $0^h1^l2^m$, עבור h, l, m כלשהם, ומילה זו שייכת ל-L_2 שכן $n = 0$ (3 לא מופיע במילה), ולכן אין הגבלה על h, l, m. במקרה בו $n \geq 2$ נבחר את $u = \varepsilon$, $v = 3$ וחלוקת שארית w תקיים $|vxy| \leq p$. כאן נקבל מילה מהצורה: $3^q0^h1^l2^m$, כאשר $q \geq 2$ או $q = 0$ (אם נבחר $i = 0$ בניפוח). כך או כך, אין לנו הגבלה על היחס בין h, l, m והמילה שייכת ל-L_2 ולפיכך ל-L.

$\Leftarrow$ לא ניתן להוכיח באמצעות למת הניפוח (בלבד) שהשפה L המוגדרת לעיל אינה חסרת הקשר.

(b)

להלן הוכחה ש-L כמוגדרת בסעיף (a) אינה חסרת הקשר:

נגדיר את השפה הרגולרית הבאה: $L' = 3 \cdot (0 + 1 + 2)^*$. נניח בשלילה כי L חסרת הקשר, אזי מתכונות סגירות של שפות חסרות הקשר יתקיים כי $L \cap L' = \{30^n1^n2^n \mid n \in \mathbb{N}\}$ גם כן חסרת הקשר. נניח כי קבוע הניפוח של השפה שקיבלנו הוא p, נתבונן במילה $w = 30^p1^p2^p$. נמצא חלוקה $w = uvxyz$ המקיימת את תנאי למת הניפוח, וברור כי $|w| \geq p$. אם $v = 3$, נבחר $i = 2$ ונקבל ש-3 מופיע במילה יותר מפעם אחת, ולכן המילה אינה בשפה. לכל מקרה אחר vxy יכלול לכל היותר 2 תווים שונים מבין $'a','b','c'$, וע"י בחירת $i = 2$ נשנה את המילה כך שמספר הופעות ה-$'a','b','c'$ במילה לא יהיה שווה, ולכן המילה שתתקבל לא תהיה בשפה. לכן, החיתוך שקיבלנו אינו שפה חסרת הקשר, ומכאן ש-L אינה חסרת הקשר.

טענה: L רגולרית $\Leftarrow$ $Half(L)$ רגולרית:

תהי L שפה רגולרית, אזי השפה $L_{even} = \{w \in L \mid |w| \in \mathbb{N}_{even}\}$ היא גם שפה רגולרית: יהי A אוטומט DFA מקבל לשפה L, אזי ניתן לבנות אוטומט DFA A_{even} המורכב משני אוטומטי A (נסמנם A_1, A_2) כך שמתחילים ב-$q_0 \in A_1$, וכל מעבר δ ב-A יקח אותנו מ-A_1 ל-A_2 או הפוך. כמו כן, כל מצב מקבל יהיה מקבל אך ורק ב-A_1 (כלומר לא יהיו מצבים מקבלים ב-A_2). כך, כל מילה שאורכה זוגי ב-L תתקבל ע"י האוטומט A_{even}, שכן תגיע לאחר מספר זוגי של מצבים למצב מקבל ב-A_1, וכל מילה שאינה זוגית מ-L או כל מילה שאינה ב-L, לא תתקבל. לפיכך L_{even} היא שפה רגולרית.

ברור כי $Half(L) = Half(L_{even})$. להלן הסבר (לא פורמלי) מדוע $Half(L_{even})$ היא שפה רגולרית:

נבנה אוטומט DFA $A_{Half(L_{even})}$ באופן הבא:

נעתיק בדיוק את A_{even}, רק ללא מצבים מקבלים כלל. נתחיל לבחון מסלולים מקבלים ב-A_{even}. נניח וקיים לנו מסלול מקבל $q_o \to \cdots \to q_{\frac{k}{2}} \to \cdots q_k$, אזי ב-$A_{Half(L_{even})}$ נסמן את q_k כמצב מקבל. טענה אותה לא נוכיח: לכל מסלול שבסופו מצב מקבל, ישנו מספר סופי של מצבים שנצטרך לסמן כמקבלים ב-$A_{Half(L_{even})}$ כדי לקבל אוטומט שאכן מקבל את $Half(L_{even})$. הסיבה לכך היא שבכל מעגל הנמצא על מסלול מקבל ניתן לבחור לעבור כמה פעמים שנרצה (החל מ-0), וישנו מספר סופי של קומבינציות מעברים בכל המעגלים על המסלול, תלוי מספר המעגלים במסלול, אורכם, ואורך המסלול, אותן נצטרך לבדוק עד שנוכל לקבוע כי "מיצינו" את המסלול המקבל הזה. לפיכך נוכל לכל אוטומט A_{even} לסמן את כל המצבים הנדרשים להיות מקבלים ב-$A_{Half(L_{even})}$ - תוך מספר סופי של צעדים. מבניית האוטומט ברור כי מתקיים $A_{Half(L_{even})}\ accepts\ w \Leftrightarrow w \in Half(L_{even}) = Half(L)$, ולפיכך $Half(L)$ היא שפה רגולרית, כנדרש.