מודלים חישוביים - עבודה 3

מתרגל אחראי - דולב נתאי

תאריך פרסום: 21/04/2020
תאריך הגשה: 10/05/2020

1 שאלה 1 (20 נק')

עבור א"ב Σ נגדיר סדר כלשהו על הא"ב, ונגדיר באמצעותו את הסדר הבא על Σ^*. הסדר יוגדר בדומה לסדר לקסיקוגרפי, פרט לכך שכאשר משווים שתי מילים בעלות אורך שונה, המילה הקצרה יותר תמיד תהיה קטנה יותר בסדר.[1] כעת, לכל $i \in \mathbb{N}$ נגדיר את $\#_i$ להיות המילה ה־i בסדר. לדוגמא עבור $\Sigma = \{a, b\}$ וסדר בו $a < b$ נקבל כי :

$$\#_0 = \varepsilon, \#_1 = a, \#_2 = b, \#_3 = aa, \#_4 = ab, ...$$

תהא L שפה כלשהי מעל Σ, נגדיר את המטריצה האינסופית $M(L)$ באופן הבא:

$$M(L)[i,j] = \begin{cases} 1 & if \quad \#_i \cdot \#_j \in L \\ 0 & if \quad \#_i \cdot \#_j \notin L \end{cases}$$

לדוגמא, עבור השפה $L = a^*$ מעל הא"ב $\Sigma = \{a, b\}$ המטריצה תראה כך:

$$M(L) = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & \dots \\ 1 & 1 & 0 & 1 & 0 & \ddots \\ 0 & 0 & 0 & 0 & 0 & \ddots \\ 1 & 1 & 0 & 1 & 0 & \ddots \\ 0 & 0 & 0 & 0 & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

הוכיחו כי L רגולרית אם ורק אם קיימות במטריצה מספר סופי של שורות ייחודיות.

2 שאלה 2 (15 נק')

תהא L שפה רגולרית כלשהי. יהיו $A = \{A_1, A_2, ..., A_n\}$ מחלקות השקילות תחת היחס $\sim_L$. תהא $A' \subseteq A$ קבוצה כלשהי של מחלקות שקילות מתוך A, ותהא L' השפה הבאה:

$$L' = \bigcup_{A_j \in A'} A_j$$

הוכיחו או הפריכו את הטענות הבאות:

1. L' רגולרית
2. $rank(L') \leqslant rank(L) = n$
3. $rank(L') \geqslant |A'|$
4. קיימת שפה L ובחירה של מחלקות שקילות A' כך ש $rank(L') < rank(L)$.

[1] סדר זה נקרא סדר SHORTLEX והוא לרוב הסדר בו נעשה שימוש בשפות פורמליות.

3 שאלה 3 (25 נק')

הוכח או הפרך כל אחת מהטענות הבאות

1. לכל שפה רגולרית L ולכל אות $\sigma \in \Sigma$ מתקיים $rank(L) = rank(f_\sigma(L))$ כאשר $f_\sigma(L)$ מוגדרת כמו בתרגול כשפה הבאה:
$$f_\sigma(L) = \{w \in \Sigma^* : \sigma \cdot w \in L\}$$

2. לכל שפה רגולרית מתקיים $rank(L) = rank(L^R)$
3. עבור שפה רגולרית לא ריקה L נסמן ב־k את אורך המילה הקצרה ביותר ב־L. אז $rank(L) > k$
4. קיימות שתי שפות רגולריות L_1 ו־L_2 כך ש $rank(L_1 \backslash L_2) > rank(L_1) \cdot rank(L_2)$

4 שאלה 4 (25 נק')

רצפי RNA הם מילים מעל הא"ב $\Sigma = \{A, U, G, C\}$ והם החומר התורשתי בנגיפים רבים, ביניהם נגיף הקורונה. חוקרים מאוניברסיטת בן גוריון הצליחו לזהות מספר שפות המגדירות רצפי RNA של נגיפי קורונה. הם רוצים לדעת מי מבין השפות האלו רגולריות, כדי שאולי יוכלו להשתמש במידע זה במאמציהם לפתח תרופה.

עבור כל אחת מהשפות הבאות, קבעו האם היא רגולרית או לא.

כאשר מראים רגולריות, אין צורך בהוכחה פורמלית, מספיקה הגדרה של האוטומט/ביטוי רגולרי והסבר קצר מדוע הוא מקבל את השפה. כאשר מראים אי־רגולריות, יש לספק הוכחה פורמלית באמצעות למת הניפוח או מחלקות שקילות.

1. $AA^*U^nA^n \cup U^*A^*$ עבור $n \in \mathbb{N}$
2. עבור מילה w ומחרוזת כלשהי $s \in \Sigma^*$ נסמן ב$|w|_s$ את מספר המופעים של s כתת־מחרוזת ב־w. למשל $|w|_{GC}$ הוא מספר המופעים של תת המחרוזת GC ב־w. נגדיר את L כשפה הבאה:
$$L = \{w : |w|_{GC} = |w|_{CG} \wedge |w|_{GU} = |w|_{GA} = |w|_{CU} = |w|_{CA} = 0\}$$

3. $L = \{w : |w| = n^2 + 2n, n \in \mathbb{N}\}$
4. $L = \{G^n \cdot U^m \cdot C^{n \cdot m}, n, m \geqslant 0\}$
5. $L = \{U^i \cdot G^j \cdot C^k : (i + j + k) \mod 3 = 0, i, j, k \geqslant 0\}$
6. $L = \{w : |w| = n + 2, n \in \mathbb{N}\}$

5 שאלה 5 (15 נק')

יהיו L_1, L_2 שפות רגולריות, ויהיו n_1, n_2 קבועי הניפוח המובטחים מלמת הניפוח עבור L_1 וL_2 בהתאמה. בסעיפים הבאים מוגדרת השפה L באמצעות L_1 ו־L_2. הוכיחו את הטענות הבאות עבור n, קבוע הניפוח של L.

1. אם $L = L_1 \cup L_2$ אז $n \leqslant \max(n_1, n_2)$. כמו־כן, קיימות שפות L_1, L_2 כך ש$n < max(n_1, n_2)$.
2. אם $L = L_1 \cap L_2$ אז קיימות שפות L_1, L_2 כך ש$n < min(n_1, n_2)$.
3. אם $L = L_1 \cdot L_2$ אז $n \leqslant n_1 + n_2$. כמו־כן, קיימות שפות L_1, L_2 כך ש$n < n_1 + n_2$.