מספר תעודת זהות

מבחן בחישוביות מכונות ושפות פורמליות
תשס״ה מועד א׳

פרופ׳ מ. בן-אור

מספר קורס: 67521

- זמן הבחינה שלוש שעות.
- כתבו את תשובותיכם על טופס המבחן.
- בשאלות בחירה **כן/לא** <u>רשמו את התשובה שבחרתם בכתב</u>.
- יש לענות על כל השאלות בחלק א׳ (1-5) ועל אחת מהשאלות בחלק ב׳ (6-7).
- אלא אם מצוין אחרת, עליכם להסביר בקצרה (בתוך התחום המוקצה לכל שאלה) את תשובותיכם. תשובות ללא הסבר לא ינוקדו. ההסבר צריך לכלול את הנקודות העיקריות בהוכחה מבלי להיכנס לפרטים טכניים, סימונים וכיוצא באלה. אם ההסבר שלכם ניתן ע״י דוגמא נגדית, עליכם לציין במפורש מהי הדוגמא ובקווים כלליים כיצד היא סותרת את הטענה. מותר להניח הנחות סבירות בלתי מוכחות כגון $P \neq NP$, אך אם הדוגמא הנגדית סותרת הנחה סבירה אך בלתי מוכחת (לדוגמא $P \neq NP$), ציינו מהי ההנחה וממה נובעת הסתירה.

הגדרות וסימונים: (זהים לחלוטין לאלה שהגדרנו בכיתה).

<u>סמונים כללים:</u>

- עבור שפה L, נסמן ב- L^c או ב- $\overline{L}$ את השפה $\Sigma^* \setminus L$ (כלומר המשלימה של L).

<u>מחלקות:</u>

- נסמן ב- REG את מחלקת השפות הרגולריות.
- נסמן ב- CFL את מחלקת השפות חסרות ההקשר.
- נסמן ב- R את מחלקת השפות ה**כריעות** על ידי מכונות טיורינג (מ״ט).
- נסמן ב- RE את מחלקת השפות ה**ניתנות לזיהוי** ע״י מ״ט.
- נסמן ב- P את מחלקת השפות המוכרעות ע״י מ״ט דטרמיניסטית בזמן פולינומי
- נסמן ב- NP את מחלקת השפות המוכרעות ע״י מ״ט לא דטרמיניסטית בזמן פולינומי.
- נסמן ב- EXP את מחלקת השפות המוכרעות ע״י מ״ט דטרמיניסטית בזמן אקספוננציאלי.
- נסמן ב- $PSPACE$ את מחלקת השפות המוכרעות ע״י מ״ט דטרמיניסטית בזכרון פולינומי.
- נסמן ב- L את מחלקת השפות המוכרעות ע״י מ״ט דטרמיניסטית בזכרון לוגריתמי
- נסמן ב- NL את מחלקת השפות המוכרעות ע״י מ״ט לא דטרמיניסטית בזכרון לוגריתמי.
- עבור מחלקה של שפות C, נסמן ב- coC את מחלקת השפות המשלימות לשפות ב- C. כלומר, $coC = \{L \mid \overline{L} \in C\}$.

<u>בעיות חישוב</u>:

$PATH=\{(G,s,t)\mid$ G גרף מכוון, ובו מסלול בין הקודקוד s לקודקוד t $\}$

- נאמר כי נוסחא מיוצגת בצורת $k-cnf$ אם היא מהצורה $\wedge_i\left(a_1^i\vee a_2^i\vee\ldots\vee a_k^i\right)$ כאשר כל a_j^i הוא משתנה או שלילתו.
- $k-SAT=\{\varphi\mid$ הנוסחא φ היא $k-cnf$ ויש לה השמה מספקת $\}$

<u>רדוקציות</u>:

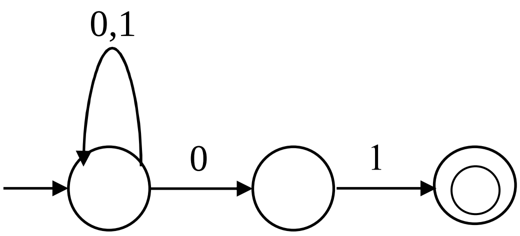
- עבור שפות L_1,L_2 נסמן ש- $L_1\le_P L_2$ אם יש רדוקצית מיפוי בזמן פולינומי מ- L_1 ל- L_2.
- עבור שפות L_1,L_2 נסמן ש- $L_1\le_L L_2$ אם יש רדוקצית מיפוי במקום לוגריתמי מ- L_1 ל- L_2.

בהצלחה

חלק א

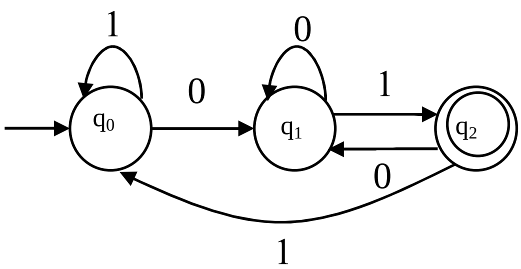
יש לענות על כל השאלות בחלק זה.

1. (15 נקודות) נתון האוטומט הלא דטרמיניסטי הבא:



א. האם המחרוזת 1101000 שייכת לשפה של האוטומט ? **לא**
אין צורך להסביר את תשובתכם.

ב. ציירו אוטומט סופי **דטרמיניסטי** עם מספר מינימלי של מצבים המקבל את אותה שפה. הסבירו בקצרה נכונות ומינימליות.



האוטומט המקורי מקבל את השפה של הביטוי הרגולרי $(0+1)^*01$. זאת ניתן לראות כיוון שכל מילה שמתקבלת, מסיימת במצב מקבל ולכן שתי האותיות האחרונות בה חייבות להיות 01. ומצד שני לכל מילה המסתיימת ב-01 יש ריצה מקבלת (להמתין במצב התחלתי עד שתי אותיות אחרונות ואז להגיע למצב מקבל).
מצויירDFA מינימלי לשפה זו.

נכונות : קל להוכיח באינדוקציה כי האוטומט נמצא במצב q_2 אם שתי האותיות האחרונות היו 01, במצב q_1 אם האות האחרונה היתה 0, ובמצב q_0 אחרת.

מינימליות: מחלקות השקילות של Myhill Nerode לשפה זו הן :

1. $(0+1)^*01$.
2. $(0+1)^*0$.
3. שאר המילים.

המחרוזת ε מפרידה בין מחלקה 1, ליתר המחלקות, ואילו המחרוזת "1" מפרידה בין מחלקה 2 למחלקה 3.

משום שיש (לפחות) 3 מחלקות שקילות, בכל DFA לשפה יש לפחות 3 מצבים ולכן האוטומט שצוייר הוא מינימלי.

2. (17 נקודות) לא"ב $\Sigma=\{0,1\}$ נגדיר פעולת משלים בצורה הבאה: $\overline{0}=1,\overline{1}=0$

ולמילים: $\overline{a_1a_2...a_n}=b_1b_2...b_n$ באשר $b_i=\overline{a_i}$ לכל i, לדוגמא $\overline{01101}=10010$.

עבור שפה L מעל $\Sigma=\{0,1\}$ נגדיר: $K(L)=\{w\overline{w} | w\in L\}$.

נקבע $L_0=0^*$.

עבור כל אחת מהשפות הבאות ציינו את המחלקה הקטנה ביותר שמכילה אותה מבין המחלקות הבאות: REG, CFL, P, או לא באף אחת מן המחלקות האלה.

א. $K(L_0)$

השפה $K(L_0)$ היא $\{0^n1^n \mid n\geq 0\}$. השפה היא חסרת הקשר ונגזרת מהדקדוק $S\rightarrow 0S1|\varepsilon$.

מצד שני, השפה אינה רגולרית (ואף ראינו זאת בכיתה). ההוכחה שהשפה אינה רגולרית, מתבססת על למת הניפוח לשפות רגולריות. נניח על דרך השלילה כי השפה רגולרית וכי קבוע הניפוח שלה הוא p, אז המילה 0^p1^p (השייכת לשפה) ניתנת להצגה כ- xyz כך ש-

$|xy|\leq p$, $|y|>0$ ולכל i המילה xy^iz גם היא בשפה.

כיון ש- $|xy|\leq p$, בהכרח y כולל רק אפסים ומשום כך xy^2z כולל יותר אפסים מאחדים בסתירה להגדרת השפה.

ב. $K(K(L_0))$

השפה $K(K(L_0))$ היא $\{0^n1^n1^n0^n \mid n\geq 0\}=\{0^n1^{2n}0^n \mid n\geq 0\}$. ברור כי שפה זו ניתנת לזיהוי בזמן פולינומי על ידי מ"ט דטרמיניסטית. השפה אינה חסרת הקשר על פי למת הניפוח לשפות חסרות הקשר. נניח על דרך השלילה כי השפה חסרת הקשר, וקבוע הניפוח שלה הוא p, אזי המילה $0^p1^{2p}0^p$ (השייכת לשפה) ניתנת להצגה כ- $uvxyz$ כך ש-

$|vxy|\leq p$, $|vy|>0$ ולכל i המילה uv^ixy^iz גם היא בשפה.

כיון ש- $|vxy|\leq p$, הניפוח יכול להשפיע רק על תת מילה של הרישא 0^p1^{2p} או על תת מילה של הסיפא $1^{2p}0^p$ ובכל מקרה uv^2xy^2z אינה בשפה, בסתירה להנחה שהשפה חסרת הקשר.

3. (17 נקודות) עבור כל אחת מהשפות הבאות ציינו את המחלקה הקטנה ביותר המכילה את השפה, מבין המחלקות הבאות: REG, CFL, P, R, RE, coRE או לא באף אחת מהמחלקות האלה.

א. $L_1 = \{\langle A,P\rangle \mid$ A הוא אוטומט סופי דטרמיניסטי, P אוטומט מחסנית ו $L(A) \subseteq L(P)\}$

השפה ב-coRE\RE. ראשית נראה כי השפה ב-coRE. קלט x **אינו** בשפה אם הוא לא מהצורה $\langle A,P\rangle$, או אם קיימת מילה y כך ש- $y \in L(A)$ אבל $y \notin L(P)$. קל לזהות אם x לא מהצורה $\langle A,P\rangle$. אם x בצורה הנכונה, ניתן לרוץ על כל המילים ב- Σ^* בסדר כלשהו בזו אחר זו, ולכל אחת לבדוק אם $y \in L(A)$ וכן $y \notin L(P)$. כיון שכל בדיקה כזו היא סופית (ולמעשה אפילו פולינומית), אם יש y כזאת, נגלה אותה לאחר זמן סופי. לכן $\overline{L_1}$ ב-RE. על מנת לראות שהשפה אינה ב-RE, נראה רדוקציה מ-ALL_{CFG} (שראינו בכיתה שאינה ב-RE) ל-L_1.

בהינתן ל-ALL_{CFG} הקלט $\langle G\rangle$, הרדוקציה תהפוך את הדקדוק G לאוטומט מחסנית P ששפתו היא השפה הנגזרת מ- G (ראינו איך עושים זאת בזמן פולינומי), ותחזיר כקלט ל-L_1 את $\langle U,P\rangle$, כאשר U הוא DFA שמקבל את Σ^*.

ברור ש $\langle U,P\rangle$ ב-L_1 אםם G ב- ALL_{CFG} (ושהרדוקציה חשיבה).

ב. $L_2 = \{\langle A,P\rangle \mid$ A הוא אוטומט סופי דטרמיניסטי, P אוטומט מחסנית ו $L(P) \subseteq L(A)\}$

השפה ב-P. על מנת לראות שהשפה ב-P, נשים לב ש- $L(P) \subseteq L(A)$ אםם $L(P) \setminus L(A) = \phi$ כלומר אםם $L(P) \cap (\Sigma^* \setminus L(A)) = \phi$. משום ש- A DFA, קל לבנות בזמן פולינומי DFA A' לשפה $\Sigma^* \setminus L(A)$. כמו כן, P אוטומט מחסנית, לכן קל לבנות בזמן פולינומי (כפי שראינו ב-ex4) אוטומט מחסנית P' לחיתוך השפות $L(P) \cap L(A')$. כמו כן, ניתן לבנות בזמן פולינומי דקדוק חסר הקשר G כך ש- $L(G) = L(P')$. לבסוף, למדנו בכיתה אלגוריתם פולינומי להכריע האם שפה של דקדוק חסר הקשר ריקה.

<u>הערה:</u> בשאלה זו ניתן ניקוד מלא גם לתלמידים שלא נימקו מדוע השפה אינה ב-CFG. השפה אינה ב-CFG משום שכלל לא ניתן להכריע ב-CFG האם מחרוזת היא קידוד של אוטומט מסוג כלשהו, משום שהדבר דורש השוואה של מחרוזות רבות (לדוגמא, לבדוק שבפונקציות מעברים מופיעים רק מחרוזות שמייצגות מצבים). על מנת לבנות הוכחה פורמלית יש להסכים על קידוד, ולהשתמש בנימוקי ניפוח.

4. (17 נקודות) עבור שפה רגולרית כלשהי L מעל א"ב Σ ושפה K נגדיר את השפה:

$$C^L(K) = \{x \in \Sigma^* \mid \exists y \in L, |y| \le |x| \wedge xy \in K\}$$

כאשר $|x|$ הוא אורך המילה x, ו-xy היא שרשור המילים x ו-y.

א. אם נתון ש $K \in \mathrm{P}$ האם גם $C^L(K) \in \mathrm{P}$?

בהנחה ש-$P \neq NP$ השפה אינה ב-P. על מנת לראות זאת נקבע $L = \Sigma^*$ ו-

$\{\varphi$ נוסחה בוליאנית ו-x השמה מספקת ל-$\varphi\}$ $K = \{\langle \varphi, x \rangle \mid$. ברור ש-$K$ ניתנת

לזיהוי בזמן פולינומי. כמו כן ברור שיש רדוקציה פשוטה מSAT ל $C^L(K)$.

הרדוקציה ממפה את 'φ' ל-'$\langle \varphi,$', ומשום שאורך ההשמה חסום באורך הנוסחה

נקבל ש-'$\langle \varphi,$' ב-$C^L(K)$ אםם φ ספיקה. (ברור שהרדוקציה ניתנת לחישוב בזמן פולינומי).

ב. אם נתון ש $K \in \mathrm{NP}$ האם גם $C^L(K) \in \mathrm{NP}$?

כן, $C^L(K) \in \mathrm{NP}$.

נתאר מכונה לא דטרמיניסטית פולינומית ששפתה $C^L(K)$. בהנתן x המכונה מנחשת

מחרוזת y באורך קטן או שווה |x|. המכונה בודקת אם $y \in L$ (דבר שניתן לביצוע

בזמן ליניארי ב|y| פשוט ע"י הרצת האוטומט של L על y). אם $y \notin L$ המכונה דוחה.

אחרת, המכונה מריצה את המכונה הלא דטרמיניסטית הפולינומית ששפתה K, על

המחרוזת xy.

ג. אילו היינו מסירים את מגבלת האורך על הסיפא y כלומר היינו מגדירים:

$$C^L(K)=\left\{x\in\Sigma^* \mid \exists y\in L,\ \ xy\in K\ \right\}$$

האם היית משנה את תשובתך לסעיפים א׳ וב׳? (נמק!)

התשובה לסעיף א׳ נותרת כשהיתה, הרדוקציה שניתנה אינה תלויה באורך y.

לגבי סעיף ב׳, יש לשנות את התשובה שכן בתנאים החדשים לא מובטח אפילו שהשפה ב- R (קל וחומר ב-NP).

על מנת לראות זאת נקבע $L=\Sigma^*$ וכן

$K=\{\left\langle M,w,1^n\right\rangle|$ המכונה M מקבלת את הקלט w בתוך n צעדים $\}$

שפה זו פולינומית (למעשה ניתנת לחישוב בזמן $O(n)$) משום שניתן לבדוק אם המחרוזת מהמבנה הנכון, ואם כך לסמלץ את M על w למשך n צעדים.

מצד שני, קל לראות ש- $A_{TM}\le C^L(K)$ פשוט ע״י הרדוקציה המעבירה את $\left\langle M,w\right\rangle$ למחרוזת $\left\langle M,w,\right\rangle$ ׳ לה יש המשך ב $C^L(K)$ אםם $\left\langle M,w\right\rangle\in A_{TM}$ (ברור שרדוקציה חשיבה).

5. (17 נקודות) בהנחה (שאינה ידועה כיום) ש- $3SAT \le_L PATH$.
עבור כל אחת מהטענות הבאות ציין האם היא נכונה או אינה נכונה ונמק (בזהירות, השאלה אינה קלה).

א. $NP = coNP$ **כן**

אם $3SAT \le_L PATH$ אזי, כיון שכל שפה ב-NP ניתנת לרדוקציה ב- log space ל- $3SAT$, מתקיים $NP \subseteq NL$, ויחד עם $NL \subseteq P$ מקבלים $NP \subseteq P$, כלומר P= NP.

כיון ש-P סגורה להשלמה, coNP = NP , כנדרש.

ב. $NP = PSPACE$ **לא**

אם $3SAT \le_L PATH$ אז $NP \subseteq NL$. ממשפט סאביץ' ידוע ש- $NL \subseteq SPACE(\log^2(n))$

וממשפט ההירכיה נובע כי $SPACE(\log^2(n))$ מוכל ממש ב-PSPACE.

לכן, אם $3SAT \le_L PATH$ אזי NP מוכל ממש ב-PSPACE.

חלק ב

יש לענות על <u>אחת</u> מתוך שתי השאלות הבאות.

6. (17 נקודות) עבור קבוצת גרפים $S=\{G_1,G_2,...,G_n\}$ נאמר שבקבוצה בדיוק l גרפים לא איזומורפיים אם קיימת תת קבוצה בגודל l של S, בה כל זוג גרפים אינם איזומורפיים, וכן בכל תת קבוצה בגודל $l+1$ של S ישנו זוג גרפים איזומורפיים.

הראו מערכת הוכחה אינטרקטיבית לשפה הבאה

$L=\{\langle G_1,G_2,...,G_n,l\rangle \mid$ ב- $G_1,...,G_n$ בדיוק l גרפים לא איזומורפיים $\}$

נמקו בקצרה מדוע מערכת ההוכחה עומדת בדרישות.

הפרוטוקול: בשלב ראשון **המוכיח** שולח למוודא קבוצה של l גרפים מתוך הגרפים הנתונים, ולכל גרף שאינו בקבוצה, המוכיח שולח פרמוטציה שעל ידה הוא מתקבל מאחד הגרפים בקבוצה (וכמובן את מספר הגרף שממנו הוא מתקבל).

המוודא בודק שאכן הגרפים שאינם בקבוצה מתקבלים על ידי הפרמוטציות הנתונות, ולאחר מכן, לכל זוג גרפים בקבוצה, המוכיח מוכיח למוודא כי זוג הגרפים אינם איזומורפיים באמצעות הפרוטוקול שראינו בכיתה. המוודא משתכנע רק עם השתכנע בכל השלבים.

קל לראות כי הפרוטוקול פולינומי.

שלמות : ברור כי אם קבוצת הגרפים כוללת בדיוק l גרפים לא איזומורפיים אז ישנם l גרפים לא איזומורפיים וכל גרף אחר איזומורפי לאחד מהם.

לכן, המוכיח יכול לשכנע את המוודא בהסתברות 1.

נאותות : אם בקבוצה יש $l+1$ גרפים לא איזומורפיים או יותר, אז לפחות אחד הגרפים מחוץ לקבוצת l הגרפים אינו איזומורפי לאף אחד מהגרפים בתוכה, ולכן במקרה זה המוכיח יכשל בשכנוע המוודא בהסתברות 1.

אם בקבוצה יש פחות מ-l גרפים לא איזומורפיים, אז בקבוצת l הגרפים ישנם לפחות זוג אחד של גרפים איזומורפיים. לכן, כאשר ינסה המוכיח לשכנע את המוודא כי הם אינם איזומורפיים, יכשל בהסתברות חצי, וכיון שההסתברות שהמוודא ישתכנע שכל l הגרפים אינם איזומורפיים חסומה על ידי ההסתברות שהמוודא ישתכנע עבור אותו זוג, הסתברות זו היא לכל היותר חצי.

7. (17 נקודות) הוכיחו שהשפה הבאה שלמה ב NL :

$K = \{ \langle G,v \rangle \mid$ G גרף מכוון ובו מעגל העובר דרך הקודקוד v $\}$

ראשית שייכות ל-NL : הרעיון הוא לנחש את המעגל (כאשר זוכרים בכל שלב רק שני קודקודים סמוכים). נשמור שלושה משתנים: קודקוד "מקור" קודקוד "מטרה" ומונה (שיחזיק מספרים מ-1 עד n, כאשר n הוא גודל הגרף). נאתחל את קודקוד המקור ל-v ונאתחל את המונה ל-1. כל עוד המונה קטן או שווה ל-$|G|$, ננחש קודקוד "מטרה" שאליו יש קשת מקודקוד ה"מקור", נבדוק אם הוא הקודקוד v, אם כן - נקבל, אחרת, נוסיף אחד למונה, נציב בקודקוד המקור את קודקוד המטרה ונחזור על הלולאה. אם יצאנו מהלולאה (כלומר המונה עלה על $|G|$ - נדחה).

ברור כי האלגוריתם משתמש רק במקום לוגריתמי, וכן כי האלגוריתם מקבל אםם יש ב-G מעגל העובר דרך v (אם יש מעגל כזה, בהכרח יש מעגל פשוט שגודלו לכל היותר $|G|$).

נראה עתה קשיות ב-NL ע"י רדוקציה מ-PATH.

בהנתן קלט $\langle G,s,t \rangle$ ל-PATH, נבנה גרף G' בו אותם הקודקודים כב-G, ובנוסף קודקוד נוסף שנסמנו v. הקשתות ב-G' יהיו כל הקשתות ב-G ובנוסף קשת מ-t ל-v וכן קשת מ-v ל-s.

הרדוקציה מעבירה את $\langle G,s,t \rangle$ ל-$\langle G',v \rangle$.

ברור כי ניתן לחשב את הרדוקציה במקום לוגריתמי.

כמו כן, ברור כי ב-G' יש מעגל העובר דרך v אםם ב-G יש מסלול מ-s ל-t.