אוטומטים 2 - הרצאה 4

אורן דנון

1 מבוא

בשיעור הקודם התחלנו לדבר על דקדוקים. הראינו מספר דוגמאות לדקדוקים וראינו כיצד משתמשים בהם(הצגנו מושגים בסיסים, כגון גזירה של מילה, אמרנו מה השפה של דקדוק, מה המרכיבים הבסיסים בהגדרת דקדוק: משתנים, א"ב, כללי גזירה משתנה התחלתי).
היום נתחיל מהגדרה פורמלית ומסודרת של דקדוק ושפה של דקדוק

1.1 הגדרות

1.1.1 דקדוק:

דקדוק G הוא רביעיה $G = (V, T, P, S)$ כאשר:

- V- קבוצה סופית לא ריקה של משתנים (בדר"כ נסמן משתני דקדוק באותיות אנגליות גדולות).
- T- קבוצה סופית לא ריקה של טרמינלים(א"ב). ומתקיים ש- $T \cap V = \emptyset$
- S- משתנה התחלתי כך ש- $S \in V$.
- P- קבוצה סופית של כללי גזירה. כל כלל גזירה היא מהצורה $\alpha \longrightarrow \beta$ כאשר $\alpha \in (U \cup T)^* V (U \cup T)^*,\ \beta \in (U \cup T)^*$

ההגדרה 1.1.1 דנה בסינטקס של דקדוק. עתה נגדיר את הסמנטיקה של דקדוק-שכתוב. כלומר כיצד מוגדרת השפה שהדקדוק "מחשב".

1.1.2 תבנית פסוקית

מילה $\alpha \in (T \cup V)^*$ נקראת תבנית פסוקית אם $S \Rightarrow^* \alpha$

1.1.3 גזירה

יהי $G = (V, T, P, S)$ דקדוק, נאמר ש- $\varphi_2 \in (V \cup T)^*$ נגזר ישירות מ- $\varphi_1 \in (V \cup T)^* V (V \cup T)^*$, ונסמן $\varphi_1 \Rightarrow \varphi_2$ אם $\varphi_1 = \beta\alpha\gamma$ כאשר $\alpha \in (V \cup T)^* V (V \cup T)^*$ ו- $\varphi_2 = \beta\delta\gamma$ וקיים ב- P כלל השכתוב $\alpha \to \delta$.
במילים אחרות - ניתן לעבור מ-φ_1 ל- φ_2 ע"י הפעלת כלל אחד מ- P.
נסמן ב- $\varphi_1 \Rightarrow^* \varphi_2$ את העובדה שניתן לגזור את φ_2 מ φ_1 באמצעות סדרה של אפס או יותר צעדי סדרה.
נסמן ב- $\varphi_1 \Rightarrow^n \varphi_2$ אם ניתן לעבור מ φ_1 ל- φ_2 ע"י בדיוק n צעדי גזירה.

1.1.4 שפה של דקדוק

יהי G דקדוק. אזי השפה של G היא קבוצת כל המילים הטרמינליות שניתן לגזור מ-S כלומר $L(G) := \{x \in T^* | S \Rightarrow^* x\}$

2 סוגים של דקדוקים

2.1 מבוא

נדבר קצת על קבוצות של דקדוקים עם מגבלות שונות, והכח של כל קבוצה. החלוקה שנדבר עליה נקראת ההירכיה של חומסקי(הומצאה ע"י נועם חומסקי בשנות ה־50) שמגדירה 4 קבוצות כאלו של דקדוקים: $G_3 \subseteq G_2 \subseteq G_1 \subseteq G_1$.

2.2 ההירכיה של חומסקי

- G_0 דקדוקים לא מוגבלים (מותרים כללי גזירה $\alpha \to \gamma$ כלשהם, בפרט מותר ש $|\alpha| < |\gamma|$) נרצה לדעת מהו הכח של קבוצת כל הדקדוקים האפשריים? מתברר באופן מפתיע קבוצת השפות שניתן לזהות ע"י דקדוק שכתוב היא $L_0 = \{L \subseteq \Sigma^* | \exists G \in G_0,\ L(G) = L\}$ $\Sigma = \{0, 1\}$, שווה לקבוצת השפות שניתן לחשב ע"י מחשב (ממודל ע"י מכונת טיורינג, אבל אפשר לחשוב עליו בתור מחשב עם python וזיכרון לא חסום).

- G_1 קבוצת הדקדוקים תלויי ההקשר. הם מוגדרים עם מגבלה על P המאפשרת לכלול כל גזירה מהצורה $\alpha \to \beta$ עם $|\alpha| \geq |\beta|$ (כללים לא מקצרי אורך), וכלל אחד מיוחד $S \to \varepsilon$ נגדיר $L_1 = \{L \subseteq \Sigma^* | \exists G \in G_1,\ L(G) = L\}$ גם כאן, G_1 מתאים למכונת טיורינג עם מגבלות מסוימות (יותר חלש ממכונת טיורינג)

- G_2 דקדוקים חסרי הקשר. כאן מרשים רק כללי גזירה מהצורה $A \to \alpha$, כאשר A הוא משתנה, ובפרט $|\alpha| \geq 1$ אלא אם $A = S$ ואז מותר $\alpha = \varepsilon$)ההגבלה על $|\alpha|$ לא חיונית כאן, אבל מכניסים אותה כדי לשמור על $G_2 \subseteq G_1$) קבוצת השפות $L_2 = \{L \subseteq \Sigma^* | \exists G \in G_2,\ L(G) = L\}$ נקראת שפות חסרות הקשר. מסתבר ש־ L_2 מתאימה לשפות שניתן לקבל ע"י אוטומט מחסנית אי־דטרמיניסטי (נוכיח בהמשך הקורס). רוב שפות התכנות ה"שפויות" הקיימות מוגדרות ע"י דקדוק חסר הקשר. כלומר, הדקדוק מקבל רק את קבוצת התכניות החוקיות בשפה. באוטומט מחסנית לעומת דקדוק משתמשים בדרך כלל בקומפיילרים כדי לפרסר תוכנה ולהכין אותה להרצה (לא רק לבדוק חוקיות) בקורס הזה נדבר בעיקר על שפות/דקדוקים חסרי הקשר.

- דקדוק לינארי ימני. מאפשריפ רק כללי גזירה מהצורה $A \to aB$ עבור $a \in T,\ A, B \in V$ או $A \to a,\ a \in T,\ A \in V$ או $S \to \varepsilon$. קבוצת השפות $L_3 = \{L \subseteq \Sigma^* | \exists G \in G_3,\ L(G) = L\}$ המתאימה, היא בדיוק קבוצת השפות הניתנות לחישוב באמצעות DFA־ השפות הרגולריות. באופן דומה ניתן להגדיר דקדוק לינארי שמאלי שבו המשתנה מופיע משמאל לטרמינל כלומר $A \to Ba$. קבוצת הדקדוקים הלינאריים השמאליים מגדירה גם כן את קבוצת השפות הלינאריות.

2.2.1 הוכחה שL_3 תאימה בדיוק לקבוצת השפות הרגולריות:

הוכחה. כיוון ראשון: תהי L שפה רגולרית נוכיח שקיים עבורה דקדוק לינארי ימני. כיוון ש־ L רגולרית קיים עבורה DFA, $A = (Q, \Sigma, q_0, F, \delta)$ רעיון הבניה הוא לבנות דקדוק של G ש"יסמלץ" את ריצת A. נקודד את Q ע"י מצבי משתני G. כלומר נקבע $V = Q$. כדי שהסימולציה תעבוד נרצה שיתקיים התנאי הבא:

$$\forall_x \in \Sigma^*,\ \hat{\delta}(q_0, x) = P \Leftrightarrow q_0 \Rightarrow_G^* xp$$

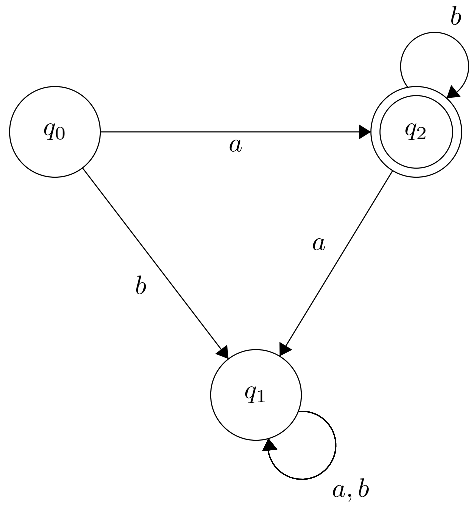
נבנה את הדקדוק כך ש־

$$\forall_x \in \Sigma^*, \forall_q,\ \hat{\delta}(q, x) = P \Leftrightarrow q \Rightarrow_G^* xp$$

לשם כך נוסיף ב G את כללי הגזירה הבאים: אם $\delta(p, \sigma) = q$ אזי נוסיף את הכלל $p \to \sigma q$ כדי לטפל בקבלת מילה, בנוסף אם $q \in F$ נוסיף גם את הכלל $p \to \sigma$ נגדיר גם $S = q_0,\ T = \Sigma$ □

2.2.2 דוגמה:

נתבונן בשפה המתקבלת ע"י האוטומט הבא:



נבנה את G כך: $G=(V=\{q_0,q_1,q_2\},S=q_0,T=\{a,b\},P)$

$$P=\{q_0\to aq_2, q_0\to a, q_0\to bq_1, q_1\to aq_1, q_1\to bq_1, q_2\to aq_1, q_2\to bq_2, q_2\to b\}$$

כאשר $q_0\to a$ נמצאת כיוון ש q_2 מצב מקבל, וקיימת השורה $\delta(q_0,a)=q_2$
נראה כיצד מילה המתקבלת ב A במסלול חישוב מסויים נגזרת ב־ G באמצעות סדרת גזירה העוקבת אחרי המסלול. נתבונן ב־$x=abb$ היא מתקבלת במסלול $q_0 \xrightarrow{a} q2 \xrightarrow{b} q2 \xrightarrow{b} q2$ הסדרה המתאימה:

$$q_0\Rightarrow aq_2\Rightarrow abq_2\Rightarrow abb$$