30/12

17

נושא 5 - ביטויים רגולריים

א) הגדרה

ביטוי רגולרי הוא דרך נוספת לתאר שפה רגולרית. בעזרת ביטוי שהתחביר שלו דומה להגדרת פסוק בתחשיב הפסוקים רק שהם מעל א"ב כלשהו Σ. כל ביטוי רגולרי מורכב משני חלקים:

(1) אטומים - סימנים שהם מותר להשתמש: ϕ (קבוצה ריקה), ε (מילה ריקה) ואותיות הא"ב ב-Σ.
כל אחד מהם לבדו הוא גם כן ביטוי רגולרי.

ב' רגולרי מוגדר בצורה רקורסיבית.

(2) פעולות יצירה - כמו קשרים המאפשרים לחבר בין אטומים וביטויים רגולריים ולקבל ביטוי רגולרי. פעולות היצירה מאוד דומים במשמעותן לפעולות על שפות. יש 4 פעולות יצירה:

① + איחוד שני ביטויים רגולריים ③ * איטרציה על ביטוי רגולרי (0 עד ∞).

② • שרשור שני ביטויים רגולריים. ④ + כמו איטרציה רק שמ-1 עד ∞.

* אוסף הביטויים הרגולריים מעל א"ב Σ יסומן ב-R_Σ או בקיצור R.
* נסמן את השפה של ביטוי רגולרי r ב-L[r].

דוגמאות:

(1) $L[\phi]=\phi,\ L[\varepsilon]=\{\varepsilon\},\ L[r^*]=(L[r])^*,\ L[r_1+r_2]=L[r_1]\cup L[r_2],\ L[\sigma]=\{\sigma\}$

(2) $L[(\varepsilon+b)\cdot(b^*)]=L[\varepsilon+b]\cdot L[b^*]=(L[\varepsilon]\cup L[b])\cdot(L[b])^*=\{\varepsilon\cup b\}\cdot\{b^*\}=\{b\}^*$

ב) קיצורי כתיבה של ביטויים רגולריים

נשתמש במספר כללים כדי לקצר את הכתיבה של ביטויים רגולריים.

* נגדיר סדר על פעולות היצירה כדי להימנע מהרבה סוגריים (כמו בסדר פעולות חשבון). הסדר הוא *,•,+. במידה ואנו רוצים לשנות את סדר זה נשתמש בסוגריים.
* בדר"כ נשמיט אופרטור שרשור ופשוט נרשום את שני הביטויים הרגולריים צמודים.
* הביטוי הרגולרי r ישמש לציון לשפה שמייצג. לא בהוכחות רק בדוגמאות כן.

דוגמאות נוספות:

(1) $a^*b^*c^*$ - שפת כל המילים שיש בהם קודם a, אח"כ b ולבסוף c.

(2) $(a+b)^*$ - שפת כל המילים מעל $\Sigma=\{a,b\}$. שונה מ-a^*b^* המכיל כל המילים שיש בהם a ואח"כ b

16/12

18

בס"ד

ג) השפה $L^k_{i,j}$

שפה זו תעזור לנו בהוכחת המשפט המרכזי בפרק זה בסעיף הבא.

לכל i,j,k ואס"ד A נגדיר את השפה $L^k_{i,j}$ להיות שפת כל המילים שמובילות את האוטומט A ממצב q_i אל מצב q_j בלי לעבור בדרך במצבים שמספרם גדול מ-k. כלומר, אם בדרך בהרצת w מ-q_i אל q_j עוברים במצב q_l כאשר $l>k$ אזי $w\notin L^k_{i,j}$, אם אך ורק גם כל שאר המצבים בדרך אזי $w\in L^k_{i,j}$. לא צריך להתקיים $k\geq i$ ו- $k\geq j$.

$$L^k_{i,j}=\{w \mid \hat{\delta}(q_i,w)=q_j \text{ AND } \forall u v\in w \ \ u,v\neq w \ \ \hat{\delta}(q_i,u)=q_l \Rightarrow l\leq k\}$$

3) שקילות ביטויים רגולריים לאוטומטים

נוכיח את השקילות באמצעות שני משפטים המראים את שני כיווני ההוכחה.

כיוון 1 **משפט:** לכל ביטוי רגולרי $r\in R_\Sigma$ מעל Σ, השפה $L[r]$ היא שפה רגולרית.

הוכחה: באינדוקציה מבנית על r, כלומר על כל ביטוי רגולרי שניתן לקבל ע"י הוספת פעולות יחידה.

בסיס - ביטוי רגולרי ללא פעולות יחידה הוא אטום ואכן כל אטום הוא שפה רגולרית: $\emptyset$, ε, $\sigma\in\Sigma$.

הנחה - נניח שהטענה נכונה עבור r_1 ו- r_2, כלומר $L[r_1]$ ו- $L[r_2]$ שפות רגולריות.

צעד - לכן שכל פעולת יחידה על r_1 ו- r_2 תיצור שפה רגולרית. טענה זו נובעת מסגירות שפות רגולריות לאיחוד, שרשור, ואיטרציה שראינו בפרק קודם.

הוכחה זו מייצרת אלגוריתם ליצור אס"ד שקול לביטוי רגולרי r נתון, כלומר שמקבל את אותה השפה. נבנה את האס"ד בשלבים לפי הוכחת סגירות לאיחוד, שרשור ואיטרציה מפרק קודם. במידה ובונים את האס"ד בדרך זו אין צורך להוכיח נכונות.

כיוון 2 **משפט:** לכל שפה רגולרית $L\subseteq\Sigma^*$ קיים ביטוי רגולרי r כך ש- $L[r]=L$.

מצב התחלתי q_1, יש סה"כ m מצבים

הוכחה: כיוון ש-L רגולרית יש לה אס"ד A המקבל אותה. כמו שהגדרנו בפרקים קודמים, מתקיים $L(A)=\bigcup_{q_j\in F} L(q_j)$. נשים לב שהגדרה זו שקולה להגדרה הבאה $L(A)=\bigcup_{q_j\in F} L^m_{1,q_j}$, שכן מתחילים ממצב התחלתי q_1 ומסיימים במצב מקבל q_j וניתן לעבור בכל מצב בדרך, שהרי m הוא האינדקס הכי גדול.

אם נוכיח שקיים ביטוי רגולרי לכל שפה $L^k_{i,j}$ סיימנו שכן איחוד ביטויים רגולריים הוא ביטוי רגולרי עם פעולת יחידה '+' ביניהם. נגדיר ביטוי רגולרי ל- $L^k_{i,j}$ באמצעות נוסחת רקורסיה. נחלק את המילים ב- $L^k_{i,j}$ ל-3 מקרים:

(1) עבור k=0 המעבר בין q_i ל-q_j מיידי ללא מצבים באמצע שכן אין אינדקס קטן או שווה מ-0 ב-A.

לכן הביטוי הרגולרי הוא ε אם i=j או חיבור כל האותיות ב-Σ המובילות מ-q_i ל-q_j.

(2) אם k>0 ולא עוברים דרך q_k אזי $L^k_{i,j}=L^{k-1}_{i,j}$ ולכן הביטוי הרגולרי יוגדר ברקורסיה $r^k_{i,j}=r^{k-1}_{i,j}$.

23/12

(3) אם כן ועוברים דרך q_k, אזי נוכל לחלק כל מילה w כזו לשלושה חלקים:

1. מ-q_i עד ל-q_k
2. מ-q_k בלולאה חזרה ל-q_k
3. מ-q_k ל-q_j.

בכל אחד מהחלקים לא עוברים באמצע ב-q_k ולכן מתקיים: $L_{i,j}^{k} = L_{ik}^{k-1}\cdot\left(L_{k,k}^{k-1}\right)^{*}\cdot L_{k,j}^{k-1}$.

נוסחת הרקורסיה תהיה $r_{i,j}^{k} = r_{i,k}^{k-1}\cdot\left(r_{k,k}^{k-1}\right)^{*}\cdot r_{kj}^{k-1}$.

לסיכום, הביטוי הרגולרי המוגדר ברקורסיה לכל $L_{i,j}^{k}$ הוא: $r_{i,j}^{k} = r_{ik}^{k-1}\cdot\left(r_{k,k}^{k-1}\right)^{*}\cdot r_{k,j}^{k-1} + r_{i,j}^{k-1}$.

כאשר תנאי העצירה הוא $k=0$. נמצא כי כל $L_{i,j}^{k}$ ניתן לייצג בביטוי רגולרי ובפרט L_{1,q_j}^{m}. מ.ש.ל.

הוכחה זו מספקת אלגוריתם למציאת ביטוי רגולרי לאוטומט ע"י כך שמחשבים את הביטוי הרגולרי לשפה L_{1,q_j}^{m} כאשר q_j מצב מקבל. בשקיפיות 27 מצגת 4 יש דוגמה כזו. נוח מאוד לחשב ביטוי כזה באמצעות טבלה כי פעמים רבות דרוש לחשב את הביטוי הרגולרי כמה פעמים.