אוטומט סופי דטרמיניסטי

אוטומט סופי דטרמיניסטי הוא חמישיה: $A = (Q, \Sigma, q_0, \delta, F)$, כאשר:
Q קבוצה סופית של מצבים, Σ א"ב הקלט, q_0 מצב התחלתי, δ פונקצית מעברים $\delta : Q \times \Sigma \to Q$
ו-F קבוצת מצבים מקבלים.

$\hat{\delta}$ הרחבה של δ למילים: $\hat{\delta}(q, \epsilon) = q, \quad \hat{\delta}(q, wa) = \delta(\hat{\delta}(q, w), a)$

השפה של אוטומט דטרמיניסטי: $L(A) = \left\{ w \in \Sigma^* \mid \hat{\delta}(q_0, w) \in F \right\}$

אוטומט סופי לא-דטרמיניסטי

שונה המודל הדטרמיניסטי רק בהגדרת רלצית המעברים, $\delta : Q \times \Sigma \to 2^Q$.

הרחבה של δ לקבוצת מצבים P: $\delta(P, \sigma) = \bigcup_{q \in P} \delta(q, \sigma)$

$\hat{\delta}$ הרחבה של δ למילים: $\hat{\delta}(q, \epsilon) = \{ q \}, \quad \hat{\delta}(q, wa) = \bigcup_{p \in \hat{\delta}(q,w)} \delta(p, a)$

השפה של אוטומט לא-דטרמיניסטי: $L(A) = \left\{ w \in \Sigma^* \mid \hat{\delta}(q_0, w) \cap F \neq \emptyset \right\}$

אוטומט סופי לא-דטרמיניסטי עם מסעי-ϵ

שונה המודל הלא-דטרמיניסטי רק בהגדרת רלצית המעברים, $\delta : Q \times \Sigma \cup \{\epsilon\} \to 2^Q$.
$CL^\epsilon(q)$ = אוסף כל המצבים שניתן להשיג מ-q ע"י סדרת מסעי-ϵ בלבד.

עבור קבוצת מצבים P: $CL^\epsilon(P) = \bigcup_{q \in P} CL^\epsilon(q)$

$\hat{\delta}$ הרחבה של δ למילים: $\hat{\delta}(q, \epsilon) = CL^\epsilon(q), \quad \hat{\delta}(q, wa) = \bigcup_{p \in \hat{\delta}(q,w)} CL^\epsilon(\delta(p, a))$

השפה של האוטומט: $L(A) = \left\{ w \in \Sigma^* \mid \hat{\delta}(q_0, w) \cap F \neq \emptyset \right\}$

ביטויים רגולרים

הגדרת ביטויים רגולרים

אוסף הביטויים הרגולרים מעל הא"ב Σ, שיסומן ע"י R, הוא הקבוצה המינימלית המקיימת:

1. $\emptyset \in R$, $\epsilon \in R$, $\sigma \in R$ לכל $\sigma \in \Sigma$.
2. אם $r_1, r_2 \in R$ אז $(r_1 + r_2) \in R$ וגם $(r_1 \cdot r_2) \in R$.
3. אם $r \in R$ אז $(r*) \in R$.

המשמעות של ביטויים רגולרים

עבור R, אוסף הביטויים הרגולרים מעל Σ, נגדיר העתקה L, $L : R \to 2^{\Sigma^*}$,
באינדוקצית מבנה על R כדלקמן:

1. $L[\emptyset] = \emptyset$, $L[\epsilon] = \{ \epsilon \}$, $L[\sigma] = \{ \sigma \}$ לכל $\sigma \in \Sigma$.
2. אם $r_1, r_2 \in R$ אז $L[(r_1 + r_2)] = L[r_1] \cup L[r_2]$, $L[(r_1 \cdot r_2)] = L[r_1] \cdot L[r_2]$.
3. אם $r \in R$ אז $L[(r^*)] = (L[r])^*$.

מעבר מאוטומט סופי דטרמיניסטי לביטוי רגולרי

$$r_{ij}^k = r_{ij}^{k-1} + r_{ik}^{k-1} \cdot (r_{kk}^{k-1})^* \cdot r_{kj}^{k-1}$$

למת הניפוח לשפות רגולריות

לכל שפה רגולרית L קיים מספר טבעי n כך שכל מלה z ב-L, שאורכה לפחות n, ניתנת לפירוק בצורה $z = uvw$ באופן שמתקיימים התנאים הבאים:
1) $|uv| \leq n$ 2) $|v| \geq 1$ 3) $uv^iw \in L$ לכל $i \geq 0$

תכונות סגור של שפות רגולריות

השפות הרגולריות סגורות תחת הפעולות הבאות:

- משלים - $L^c = \{ w \in \Sigma^* \mid w \notin L \}$
- חיתוך - $L_1 \cap L_2 = \{ w \mid w \in L_1 \wedge w \in L_2 \}$
- איחוד - $L_1 \cup L_2 = \{ w \mid w \in L_1 \vee w \in L_2 \}$
- חיסור - $L_1 - L_2 = \{ w \mid w \in L_1 \wedge w \notin L_2 \}$
- שרשור - $L_1 \cdot L_2 = \{ w \mid w = uv,\ u \in L_1 \wedge v \in L_2 \}$
- איטרציה - $L^* = \bigcup_{i \geq 0} L^i$, כאשר $L^0 = \{ \epsilon \}$ $L^i = \{ w_1 ... w_i \mid \forall 1 \leq i \leq i,\ \ w_i \in L \}$

- הומומורפיזם -

$h : \Sigma \to \Delta^*$

$\hat{h}$ הרחבה של h למילים: $\hat{h}(\epsilon) = \epsilon$

$\hat{h}(w\sigma) = \hat{h}(w) \cdot h(\sigma)$

הרחבה לשפות: $\hat{h}(L) = \bigcup_{x \in L} \hat{h}(x)$

- הומומורפיזם הפוך.

$h^{-1}(L) = \{w |\ w \in \Sigma^*,\ h(w) \in L\}$

- חלוקה מימין - $L_1/L_2 = \{\ x \mid \exists y \in L_2,\ xy \in L_1\ \}$
- חלוקה משמאל - $L_2 \backslash L_1 = \{\ x \mid \exists y \in L_2,\ yx \in L_1\ \}$
- הצבה רגולית, פונקציה f המעתיקה אותיות ב-Σ לשפות רגולריות מעל Δ, $f : \Sigma \to 2^{\Delta^*}$ (הגדרות ההרחבה למילים ולשפות דומות לאלו של הומומורפיזם).

דקדוקים חסרי הקשר

דקדוק ח"ה הוא רביעיה: $G = (V, T, P, S)$, כאשר:
V קבוצה סופית של משתנים, T קבוצה סופית של טרמינלים, S משתנה התחלתי, P חוקי גזירה מהצורה $A \to \beta$ כאשר $A \in V$ ו-$\beta \in (V \cup T)^*$.

גזירה בצעד יחיד: נאמר כי מ-φ_1 ניתן לגזור את φ_2, ונסמן $\varphi_1 \Rightarrow \varphi_2$, אם: $\varphi_1 = \alpha A \beta$, $\varphi_2 = \alpha\gamma\beta$ וכן $A \to \gamma \in P$, כאשר $A \in V$, ו-$\alpha, \beta, \gamma \in (V \cup T)^*$.

גזירה במספר צעדים סופי: מסמנים $\varphi_1 \Rightarrow^* \varphi_2$, אם קיימת סדרה סופית של צעדי גזירה $\varphi_1 \Rightarrow \ldots \Rightarrow \ldots \Rightarrow \varphi_2$.

השפה של דקדוק ח"ה:

$$L(G) = \{\ w \in T^* \mid S \Rightarrow^* w\ \}$$

פישוט דקדוקים

1. זריקת משתנים מיותרים.
   (א) זריקת משתנים שאינם ניתנים לגזירה טרמינלית.
   (ב) זריקת משתנים שאינם ניתנים להשגה מ-S.
2. ביטול חוקי-ϵ.
3. (אופציונלי) זריקת משתנים מיותרים.
4. ביטול חוקי יחידה.
5. זריקת משתנים מיותרים.

משפט ההצבה

יהי $G = (V, T, P, S)$ דקדוק חסר הקשר. יהי $A \to \alpha_1 B \alpha_2$ כלל ב-P, ויהיו כל כללי-B $B \to \beta_1 \mid \beta_2 \mid \ldots \mid \beta_s$.
יהי $G' = (V, T, P', S)$ הדקדוק המתקבל מ-G בהשמטת הכלל $A \to \alpha_1 B \alpha_2$ והוספת הכללים האלה: $A \to \alpha_1\beta_1\alpha_2 \mid \alpha_1\beta_2\alpha_2 \mid \ldots \mid \alpha_1\beta_r\alpha_2$. אז: $L(G) = L(G')$.

הלמה לסילוק רקורסיה שמאלית

יהי $G = (V, T, P, S)$ דקדוק ח"ה ויהיו $A \to A\alpha_1 | A\alpha_2 | \ldots | A\alpha_r$ כל כללי A שאגף ימין שלהם מתחיל ב-A, וכן $A \to \beta_1 | \beta_2 | \ldots | \beta_s$ שאר כללי A. יהי $G' = (V \cup \hat{A}, T, P', S)$ הדקדוק המתקבל מ-G ע"י הוספת משתנה חדש $\hat{A}$ והחלפת כל כללי A ב-G באוסף הכללים:

$$A \to \beta_1 | \beta_2 | \ldots | \beta_s | \beta_1\hat{A} | \beta_2\hat{A} | \ldots | \beta_s\hat{A}$$
$$\hat{A} \to \alpha_1 | \alpha_2 | \ldots | \alpha_r | \alpha_1\hat{A} | \alpha_2\hat{A} | \ldots | \alpha_r\hat{A}$$

אזי $L(G') = L(G)$.

הצורה הנורמלית של חומסקי

דקדוק הינו בצורה הנורמלית של חומסקי אם כל כללי הגזירה שלו הם מהצורה: $A \to a$ או מהצורה $A \to BC$, $A, B, C \in V,\ a \in T$.

הצורה הנורמלית של גרייבך

דקדוק הינו בצורה הנורמלית של גרייבך אם כל כללי הגזירה שלו הם מהצורה: $A \to a\alpha$, $A \in V,\ a \in T,\ \alpha \in V^*$.

אלג׳ למעבר מדקדוק פשוט לדקדוק בצורה הנורמלית של גרייבך:

1. נשנה את הכללים כך שיתקיים אם קיים כלל גזירה $A_i \to A_j\alpha$ אז $i < j$.
   - נתחיל עם A_1 אם יש כלל $A_1 \to A_1\alpha$ נוסיף משתנה $\hat{A}_1$ ונשתמש במשפט לסילוק רקורסיה שמאלית.
   - נניח כי טיפלנו במשתנים $A_1, ..., A_{k-1}$ נשנה את כללי A_k כך:
     - עבור כלל $A_k \to A_j\alpha$ $j < k$ נשתמש בכלל ההצבה על A_j.
     - עבור כלל $A_k \to A_k\alpha$ נוסיף משתנה $\hat{A}_k$ ונשתמש בסילוק רקורסיה שמאלית.
2. נדאג לכך שכל כללי A_i יתחילו בטרמינל.
   - עבור A_n הדבר כבר כך (שכן לא קיים משתנה עם אינדקס גבוה יותר).
   - נניח כי טיפלנו בכללים $A_{i+1}, ..., A_n$ כעת נטפל בכללים של A_i ע"י משפט ההצבה.
3. נדאג לכך שכללי $\hat{A}_i$ יתחילו בטרמינל. כל חוק שאינו מתחיל בטרמינל מתחיל ב-A_j כלשהו, נשתמש בכלל ההצבה על A_j על מנת שגם חוקיהם יתחילו בטרמינל.
4. כעת כל הכללים הם מהצורה $A \to \sigma\alpha$ $\alpha \in (V \cup T)^*$ נרצה ש-α יכיל רק משתנים, ולכן נוסיף לכל σ משתנה X_σ וכלל $X_\sigma \to \sigma$, ונחליף כל מופע של σ ב-α ב-X_σ.

דקדוקים רגולריים

דקדוק G ייקרא לינארי ימני אם כל כללי הגזירה שלו הם מהצורה: $A \to a$ או מהצורה $A \to aB$ כאשר $A, B \in V,\ a \in T$, וכן מותר כלל מהצורה $S \to \epsilon$.

באופן דומה מגדירים דקדוק לינארי שמאלי, כאשר הכללים הם מהצורה: $A \to Ba$, $A \to a$, $S \to \epsilon$.

דקדוק ייקרא רגולרי אם הוא לינארי ימני או לינארי שמאלי.

אוטומט מחסנית

אוטומט מחסנית הוא שביעיה: $M = (Q, \Sigma, \Gamma, \delta, q_0, \dashv, F)$ כאשר:

Q קבוצה סופית של מצבים, Σ א"ב הקלט, Γ א"ב המחסנית, δ רלצית מעברים $\delta : Q \times (\Sigma \cup \{ \epsilon \}) \times \Gamma \to 2^{Q \times \Gamma^*}$, q_0 מצב התחלתי, $\dashv$ האות ההתחלתית במחסנית, F קבוצת מצבים מקבלים.

תאור רגעי: (q, w, α) כאשר: q הינו המצב הנוכחי, $w \in \Sigma^*$ הקלט שטרם נקרא, $\alpha \in \Gamma^*$ תכולת המחסנית.

צעד באוטומט: מסמנים $(q, aw, A\alpha) \vdash (p, w, \beta\alpha)$ אם"ם $(p, \beta) \in \delta(q, a, A)$, כאשר $a \in \Sigma \cup \{ \epsilon \}$. סדרה סופית של צעדים תסומן ע"י $\vdash^*$.

שני אופני הקבלה של אוטומט מחסנית:

1. <u>קבלה ע"י מצב מקבל:</u> מלה w מתקבלת אם"ם קיימת סדרת מעברים $(q_0, w, \dashv) \vdash^* (q_f, \epsilon, \gamma)$ כאשר $q_f \in F$ ו-$\gamma \in \Gamma^*$.
2. <u>קבלה ע"י ריקון מחסנית:</u> מלה w מתקבלת אם"ם קיימת סדרת מעברים $(q_0, w, \dashv) \vdash^* (p, \epsilon, \epsilon)$ כאשר $p \in Q$.

אוטומט מחסנית דטרמיניסטי

אוטומט מחסנית $M = (Q, \Sigma, \Gamma, \delta, q_0, \dashv, F)$ ייקרא דטרמיניסטי אם מתקיימים שני התנאים הבאים:

1. לכל $q \in Q$, $a \in \Sigma \cup \{ \epsilon \}$ ו-$Z \in \Gamma$: $|\delta(q, a, Z)| \le 1$.
2. לכל $q \in Q$ ו-$Z \in \Gamma$ אם $\delta(q, \epsilon, Z) \neq \emptyset$ אזי לכל $\sigma \in \Sigma$, $\delta(q, \sigma, Z) = \emptyset$.

שקילות דקדוק ח"ה ואוטומט מחסנית

מעבר מדקדוק ח"ה לאוטומט מחסנית

יהי $G = (V, T, P, S)$ דקדוק ח"ה בצורה הנורמלית של גרייבך כך ש-$L = L(G)$.

נגדיר: $M = (\{ q_0 \}, T, V, \delta, q_0, S, \emptyset)$. את δ נגדיר בהתאמה לכללי P, כלומר לכל $a \in T$ ו-$A \in V$: $\delta(q_0, a, A) = \{ (q_0, \alpha) \mid A \to a\alpha \in P,\ \alpha \in V^* \}$.

מעבר מאוטומט מחסנית לדקדוק ח"ה

יהי $M = (Q, \Sigma, \Gamma, \delta, q_0, \dashv, \emptyset)$ אוטומט מחסנית המקבל ע"י ריקון.
נגדיר $G = ((Q \times \Gamma \times Q) \cup \{ S \}, \Sigma, P, S)$. הכללים של P מוגדרים כך:

1. לכל $q \in Q$ יש כלל $S \to [q_0, \dashv, q]$.
2. לכל מעבר באוטומט M, $(q_1, B_1B_2...B_m) \in \delta(q, a, A)$, שעבורו $m > 0$, וכל בחירה של מצבים $q_2, ..., q_{m+1}$ (לוא דוקא שונים זה מזה) יש כלל:
$[q, A, q_{m+1}] \to a[q_1, B_1, q_2][q_2, B_2, q_3]...[q_m, B_m, q_{m+1}]$, כאן $a \in \Sigma \cup \{ \epsilon \}$.
3. לכל מעבר באוטומט M, $(q_1, \epsilon) \in \delta(q, a, A)$, יש כלל: $[q, A, q_1] \to a$, גם כאן $a \in \Sigma \cup \{ \epsilon \}$.

למת הניפוח לשפות חסרות הקשר

לכל שפה חסרת הקשר L קיים מספר טבעי n כך שכל מלה z ב-L, שאורכה לפחות n, ניתנת לפירוק בצורה $z = uvwxy$ באופן שמתקיימים התנאים הבאים:
1) $|vwx| \leq n$ 2) $|vx| \geq 1$ 3) $uv^iwx^iy \in L$ לכל $i \geq 0$.

תכונות סגור של שפות חסרות הקשר

- השפות החסרות ההקשר סגורות תחת הפעולות הבאות: איחוד, שרשור, איטרציה, הומומורפיזם, הומומורפיזם הפוך, הצבה ח"ה, חלוקה מימין (ומשמאל) בשפה רגולרית, חיתוך עם שפה רגולרית.
- השפות חסרות ההקשר אינן סגורות תחת הפעולות: חיתוך, משלים, חיסור.

שפות מוכרות

- שפות ח"ה ולא רגולריות: $\{ a^ib^j \mid iRj, \quad R \in \{ <, \leq, >, \geq, = \} \}$, $\{ ww^R \mid w \in \Sigma^*, \ |\Sigma| \geq 2 \}$, $\{ w \mid \#_a(w)R\#_b(w), \quad R \in \{ <, \leq, >, \geq, = \} \}$,
- שפות לא ח"ה: $\{ a^p \mid p \ is \ prime \}$, $\left\{ a^{i^2} \mid i \geq 0 \right\}$, $\{ a^{i!} \mid i \geq 0 \}$, $\{ a^nb^nc^n \mid n \geq 0 \}$, $\{ ww \mid w \in \Sigma^* \}, \ |\Sigma| \geq 2$.

אלגוריתם CYK

for i=:1 to n do
  $V_{i,1} := \{A \mid$ הוא כלל ב-P $A \to a$ והאות ה-i של w היא $a\}$
for j=:2 to n do
  for i=:1 to n+1-j do
    begin
      $V_{i,j} = \emptyset$
      for t=:1 to j-1 do
        $V_{i,j} = V_{i,j} \cup \{ A \mid$ הוא כלל ב-P $A \to BC$,$B \in V_{i,t}$,$C \in V_{i+t,j-t} \}$
    end;