Целта на този документ е с помощта на няколко задачи и примера да изясни основните идеи за построяването на минимален суфиксен автомат по дадена дума, [BBH+85].

Основните понятия, които използваме са:

Дефиниция 0.1 (Релация на Майхил – Нероуд). *За език $L \subseteq \Sigma^*$ и думи $u, v \in \Sigma^*$:*

$$u \equiv_L v \overset{def}{\iff} \forall \gamma \in \Sigma^* (u \circ \gamma \in L \leftrightarrow v \circ \gamma \in L).$$

Дефиниция 0.2. *За дума $w \in \Sigma^*$ с $Pref(w)$, $Inf(w)$ и $Suff(w)$ бележим съответно множествата от всички префикси, всички инфикси и всички суфикси на думата w.*

Дефиниция 0.3 ($end-pos_w$). *За дума $w = a_1 a_2 \ldots a_n$, $a_i \in \Sigma$, и дума $u \in \Sigma^*$:*

$$\text{end-pos}_w(u) = \{i \,|\, 0 \leq i \leq n \,\&\, u = a_{i-|u|+1} \ldots a_i\}.$$

Дефиниция 0.4 ((каноничен) представител). *Ако $w \in \Sigma^*$ и $v \in Inf(w)$, то $\overset{w}{\overleftarrow{v}}$ е най-дългата дума в класа на еквивалентност $[v]_{\equiv_{Suff(w)}}$.*

Дефиниция 0.5 ($slink$). *Ако $w \in \Sigma^*$ и $v \in \Sigma^+$, то $s = slink_w(v)$ е най-дългият суфикс на v, за който $v \not\equiv_{Suff(w)} s$.*

По дадена дума $w \in \Sigma^*$ искаме да намерим:

$$Q_w = \left\{ \overset{w}{\overleftarrow{v}} \,|\, v \in Inf(w) \right\}$$

за всяко $v \in Q_w$: $\quad slink_w(v)$

за всяко $v \in Q_w$: $\quad \delta_w(v, a) = \begin{cases} \overset{w}{\overleftarrow{va}}, & \text{ако } va \in Inf(w) \\ \neg!, & \text{иначе,} \end{cases}$

където $a \in \Sigma$.

1 Връзки между $\text{end-pos}_w(.)$ и $\equiv_{Suff(w)}$

Задача 1.1. *Нека $w = bbacbba$. Нека $L = Suff(w)$ е езикът от суфикси на думата w.*

1. *Определете $\text{end-pos}_w(a)$, $\text{end-pos}_w(ba)$ и $\text{end-pos}_w(bba)$.*
2. *Намерете елементите на класа на еквивалентност $[a]_{\equiv_L}$.*
3. *Намерете $\text{end-pos}_w(c)$, $\text{end-pos}_w(aa)$ и $\text{end-pos}_w(baa)$.*

Задача 1.2. *Докажете отново, че за всяка дума $w \in \Sigma^*$ и думи $u, v \in \Sigma^*$ следните са еквивалентни:*

1. $u \equiv_L v$,
2. $\text{end-pos}_w(u) = \text{end-pos}_w(v)$.

Задача 1.3. *Нека $w = bbacbba$ и $w' = w \circ a$.*

1. *Определете* end-pos$_{w'}(a)$, end-pos$_{w'}(ba)$, end-pos$_{w'}(bba)$, end-pos$_{w'}(aa)$, end-pos$_{w'}(baa)$, end-pos$_{w'}(w')$ *и сравнете получените резултати с тези от задача 1.*

2. *Сравнете* end-pos$_w(bb)$ *и* end-pos$_{w'}(bb)$.

Задача 1.4. *Нека* $w \in \Sigma^*$, $a \in \Sigma$ *и* $w' = w \circ a$ *са произволни. Докажете, че:*

1. $\mathrm{Inf}(w') \setminus \mathrm{Inf}(w) \subseteq \mathrm{Suff}(w')$.

2. *ако* $v \in \mathrm{Suff}(w')$, *то:* end-pos$_{w'}(v) =$ end-pos$_w(v) \cup \{|w'|\}$.

3. *ако* $v \notin \mathrm{Suff}(w')$, *то:* end-pos$_{w'}(v) =$ end-pos$_w(v)$.

2 Свойства на $slink_w$

Задача 2.1. *Нека* $w = bbacbba$ *и* $w' = w \circ a$. *Намерете:*

1. $slink_w(w)$, $slink_w(bba)$, $slink_{w'}(a)$, $slink_w(bbac)$, $slink_w(bb)$, $slink_w(b)$.

2. $slink_{w'}(w)$, $slink_{w'}(w')$, $slink_{w'}(bba)$, $slink_{w'}(a)$.

Задача 2.2. *Нека* $w \in \Sigma^*$, $a \in \Sigma$ *и* $w' = w \circ a$. *Докажете, че ако* $u, v \in \mathrm{Inf}(w')$ *са такива, че:*

$$u \equiv_{\mathrm{Suff}(w)} v \,\&\, u \not\equiv_{\mathrm{Suff}(w')} v,$$

то поне една от двете думи u *и* v *е еквивалентна на* w' *или* $slink_{w'}(w')$ *относно* $\equiv_{\mathrm{Suff}(w')}$. *Заключете, че:*

$$ind(\equiv_{\mathrm{Suff}(w)}) + 1 \leq ind(\equiv_{\mathrm{Suff}(w')} \leq ind(\equiv_{\mathrm{Suff}(w)}) + 2.$$

Задача 2.3. *За думите* $w = bbacbba$ *и* $w' = w \circ a$ *намерете* Q_w *и* $Q_{w'}$.

Задача 2.4. *Докажете, че за всеки* $w \in \Sigma^*$, $a \in \Sigma$ *и* $w' = w \circ a$ *е в сила:*

$$Q_{w'} = Q_w \cup \{w', slink_{w'}(w')\}.$$

3 Пресмятане на $slink_{wa}(wa)$

Задача 3.1. *Нека* $w = bbacbba$ *и* $w' = w \circ a$ *намерете:*

1. *всички* $v \in Q_w$, *за които* $v \circ a \in \mathrm{Inf}(w)$.

2. *всички* $v \in \mathrm{Suff}(w)$, *за които* $v \circ a \in \mathrm{Inf}(w)$.

3. *всички* $v \in \mathrm{Suff}(w) \in Q_w$, *за които* $v \circ a \in \mathrm{Inf}(w)$.

`FindStem`$(Q, slink, \delta, w, w', a)$
@1 $u \leftarrow w$
@2 `while` $u \neq NULL$ `and` $\neg!\delta(u, a)$
@3 $\quad \delta(u, a) \leftarrow w'$
@4 $\quad u \leftarrow slink(u)$
@5 `done`
@6 `return` u

`NewSlink`$(Q, slink, \delta, u, w', a)$
@1 $v \leftarrow \delta(u, a)$, $s \leftarrow u \circ a$
@2 $Q \leftarrow Q \cup \{s\}$
@3 $slink(s) \leftarrow slink(v)$
@4 $slink(v) \leftarrow s$
@5 $slink(w') \leftarrow s$
@6 `for` $b \in \Sigma :\ \delta(v, b) \neq NULL$ `do`
@7 $\quad \delta(s, b) \leftarrow \delta(v, b)$
@8 `return` s

Задача 3.2. *Нека $w \in \Sigma^*$, $a \in \Sigma$ и $w' = w \circ a$. Да разгледаме процедурата* ***FindStem***[1]*, в която предполагаме, че $Q = Q_w \cup \{w'\}$, $slink = slink_w$ и $\delta = \delta_w$. Нека u е резултатът от изпълнението на* ***FindStem****.*

Докажете, че:

1. $slink_{w'}(w') = \begin{cases} \varepsilon & \text{ако } u = NULL \\ u \circ a & \text{иначе.} \end{cases}$
2. *ако $u \neq NULL$ и $u \circ a \neq \delta_w(u, a)$, то $|\delta_w(u, a)| > |u| + 1$.*
3. *при предположенията на предишната подточка докажете, че след изпълнението на* ***NewSlink***[2] *функцията $slink = slink_{w'}$.*

4 Конструирането на преходите $\delta_{w'}$

Задача 4.1. *Нека $w \in \Sigma^*$, $a \in \Sigma$ и $w' = w \circ a$. Нека u е дефинирано както в задача 3.2 , а $v \in Q_w \cup \{slink_{w'}(w')\}$ е произволно. Докажете, че ако $a \neq b \in \Sigma$, то:*

1. *$\delta_{w'}(w', b)$ не е дефинирано.*
2. *$\delta_{w'}(v, b) = \delta_w(v, b)$ за $v \in Q_w$.*
3. *$\delta_{w'}(v, b) = \delta_w(\delta_w(u, a), b)$ ако $v = slink_{w'}(w') \notin Q_w$.*

Задача 4.2. *Нека $w \in \Sigma^*$, $a \in \Sigma$ и $w' = w \circ a$. Нека u е дефинирано както в задача 3.2, а $v \in Q_w \cup \{slink_{w'}(w')\}$ е произволно. Докажете, че:*

1. *$\delta_{w'}(v, a) = \delta_w(v, a)$ ако $v \notin Suff(w)$.*
2. *$\delta_{w'}(v, a) = w'$, ако $v \in Suff(w) \cap Q_w$, $|v| > |u|$,*
3. *$\delta_{w'}(v, a) = \delta_{w'}(\delta_w(u, a), a)$, ако $v = slink_{w'}(w') \notin Q_w$.*

Задача 4.3. *Разгледайте процедурата* ***InsertNextChar****, която се изпълнява с $Q = Q_w$, $slink = slink_w$ и $\delta = \delta_w$. Докажете, че ако се изпълни ред @11, то непосредствено след това:*

1. *$Q = Q_{w'}$, където $w' = w \circ a$,*
2. *$slink = slink_{w'}$,*

[1] Може да се абстрахирате засега от ред @3.
[2] Може да се абстрахирате засега от редове @6 и @7.

3. *$\delta(v,b) = \delta_{w'}(v,b)$ за всички състояния $v \in Q$ и $b \in \Sigma$ с изключение може би на $b = a$ и $v \in \mathit{Suff}(w)$ с $|v| \le |u|$.*

Задача 4.4. *Нека $w \in \Sigma^*$, $a \in \Sigma$ и $w' = w \circ a$. Нека u е дефинирано както в задача 3.2, а $v \in \mathit{Suff}(w)$ с $|v| \le |u|$ е произволно. Тогава* $\delta_{w'}(v,a) = \begin{cases} \mathit{slink}_{w'}(w') & \text{ако } \delta_w(v,a) = \delta_w(u,a) \\ \delta_w(v,a), & \text{иначе.} \end{cases}$

Задача 4.5. *Докажете, че ако $Q = Q_w$, $\mathit{slink} = \mathit{slink}_w$ и $\delta = \delta_w$ за някоя дума $w \in \Sigma^*$, то след изпълнението на `InsertNextChar` $Q = Q_{w\circ a}$, $\mathit{slink} = \mathit{slink}_{w\circ a}$ и $\delta = \delta_{w\circ a}$ като резултатът е $w \circ a$.*

`Redirect`$(Q, \mathit{slink}, \delta, u, s, a)$

- $v \leftarrow \delta(u,a)$
- `while` u `is defined and` $\delta(u,a) = v$ `do`
  - $\delta(u,a) \leftarrow s$
  - $u \leftarrow \mathit{slink}(u)$
- `done`

`InsertNextChar`$(Q, \mathit{slink}, \delta, w, a)$

- $w' \leftarrow w \circ a$
- $Q \leftarrow Q \cup \{w'\}$
- $u \leftarrow \mathit{FindStem}(Q, \mathit{slink}, \delta, w, w', a)$
- `if` $u = \mathit{NULL}$ `then`
  - $\mathit{slink}(w') \leftarrow \varepsilon$
  - `return` w'
- $v \leftarrow \delta(u,a)$
- `if` $|v| = |u| + 1$ `then`
  - $\mathit{slink}(w') \leftarrow v$
  - `return` w'
- $s \leftarrow \mathit{NewSlink}(Q, \mathit{slink}, \delta, u, w', a)$
- $\mathit{Redirect}(Q, \mathit{slink}, \delta, u, s, a)$
- `return` w'

Литература

[BBH+85] A. Blumer, J. Blumer, D. Haussler, A. Ehrenfeucht, M. T. Chen, and J. Seiferas. The smallest automation recognizing the subwords of a text. *Theoretical Computer Science*, 40:31 – 55, 1985. Eleventh International Colloquium on Automata, Languages and Programming.

Упътвания към някои от задачите

Упътване 4.1 (Упътване за задача 2.2). Използвайте задачи 1.2 и 1.4.

Тъй като $u \equiv_{\mathit{Suff}(w)} v$, но $u \not\equiv_{\mathit{Suff}(w')} v$, то една от думите u и v е суфикс на w', а другата – не, защото техните $\mathit{end\text{-}pos}_{w'}(.)$ множества могат да се различават единствено по елемента $|w'|$. Ако $|w'| \in \mathit{end\text{-}pos}_{w'}(v) \setminus \mathit{end\text{-}pos}_{w'}(u)$, разгледайте два случая:

1. $\mathit{end\text{-}pos}_{w'}(v) = \{|w'|\}$. Тогава $v \equiv_{\mathit{Suff}(w')} w'$ като обърнете внимание, че $\mathit{end\text{-}pos}_{w'}(u) = \emptyset$. Тази ситуация може да интерпретираме като $u \notin \mathit{Inf}(w')$ и v се среща един единствен път в w' при това като суфикс.

2. $|w'| \subsetneq \mathit{end\text{-}pos}_{w'}(v)$. Тогава има позиция $k \neq |w'|$, за която $k \in \mathit{end\text{-}pos}_{w'}(v)$. Оттук $k < |w'|$ и следователно $v \not\equiv_{w'} w'$ и $v \in \mathit{Inf}(w)$. Нека $\overleftarrow{v} \equiv_{\mathit{Suff}(w')} v$ е най-дългият представител в класа на v относно $\equiv_{\mathit{Suff}(w')}$. Тогава $\overleftarrow{v}$ завършва на позиция $|w'|$ в w',

следователно е суфикс на w'. Аналогично, тъй като $\overleftarrow{v}$ завършва и на позиция k в w. Тъй като u също завършва на позиция k в w, но не и на позиция $|w'|$ в w', то:

$$u = \alpha \circ b \circ \overleftarrow{v} \text{ и } w' = \beta \circ c \circ \overleftarrow{v}.$$

В w имаме, че $\mathit{end\text{-}pos}_w(\overleftarrow{v}) = \mathit{end\text{-}pos}_w(v) = \mathit{end\text{-}pos}_w(u)$. Оттук получаваме, че $\mathit{end\text{-}pos}_w(\overleftarrow{v}) = \mathit{end\text{-}pos}_w(b \circ \overleftarrow{v})$, но тогава $c \circ \overleftarrow{v} \notin \mathit{Inf}(w)$, откъдето наистина $\overleftarrow{v} = \mathit{slink}_{w'}(w')$.

Упътване 4.2 (Упътване за задача 2.4). Използвайте задача 2.2 и разсъжденията от нея.

Упътване 4.3 (за задача 4.1). Забележете, че ако $b \neq a$ то $\mathit{end\text{-}pos}_w(v \circ b) = \mathit{end\text{-}pos}_{w'}(v \circ b)$. Сега използвайте задача 1.4

Упътване 4.4 (за задача 4.2). Отново използвайте задача 1.4.

Упътване 4.5 (за задача 4.4). Забележете, че ако v е както в условието и $\delta_w(v, a) \neq \delta_w(u, a)$, то позицията $|w| + 1 = |w'|$ се добавя към всички елементи от класа на еквивалентност $[v \circ a]_{\mathit{Suff}(w)}$. Тогава той остава непроменен.

Относно тези v, за които $\delta_w(v, a) = \delta_w(u, a)$. За тях имаме, че $v \circ a \equiv_{\mathit{Suff}(w)} u \circ a$ и тъй като $v \circ a$ е суфикс и на w' както и $u \circ a = \mathit{slink}_{w'}(w')$, то $v \circ a \equiv_{\mathit{Suff}(w')} u \circ a$.