Juan Carlos Llamas Núñez 3º DG Mat-Inf

Ejercicio 1.- Utilizando la definición y suponiendo que el alfabeto es $\Sigma = \{a, b\}$, calcular la función Π para el patrón $P = ababbabbababb$.

Recordamos que Π es una función de $\{1 \cdots m\}$ en $\{0 \cdots m-1\}$.
En este caso $m = 13$. La definición de Π es:

$\Pi(q) = \text{máx}\{k \mid k<q \;\; P_k \sqsupset P_q\}$ para $q \in \{1 \cdots m\}$, es decir,

calcula la longitud del prefijo más largo de P que es sufijo de P_q (sin contar a P_q).

a	b	a	b	b	a	b	b	a	b	a	b	b
1	2	3	4	5	6	7	8	9	10	11	12	13

$\Pi(1) = 0$ $\quad \boxed{P_1 = a}$

$\Pi(2) = 0$ (No es 1 porque $P_1 = a$ no es sufijo de $P_2 = ab$) $\boxed{P_2 = ab}$

$\Pi(3) = 1$ (No es 2 porque $P_2 = ab$ no es sufijo de $P_3 = aba$ y $P_1 = a$ sí) $\boxed{P_3 = \underbrace{a}_{P_1}b\underbrace{a}_{P_1}}$

$\Pi(4) = 2$ ($P_2 = ab$ es sufijo de $P_4 = abab$ y $P_3 = aba$ no). $\boxed{P_4 = \underbrace{ab}_{P_2}\underbrace{ab}_{P_2}}$

$\Pi(5) = 0$ (Ningún prefijo acaba en bb). $\boxed{P_5 = ababb}$

$\Pi(6) = 1$ (Los únicos prefijos que acaban en a son P_1 y P_3 pero P_3 no es sufijo de P_6).

$\boxed{P_6 = \underbrace{a}_{P_1}babb\underbrace{a}_{P_1}}$

$\Pi(7) = 2$ (Los únicos prefijos que acaban en ab son P_2 y P_4 y P_4 no es sufijo de P_7).

$\boxed{P_7 = \underbrace{ab}_{P_2}abb\underbrace{ab}_{P_2}}$

$\Pi(8) = 0$ (El único sufijo que acaba en bb es P_5 pero no es sufijo de P_8.

$\boxed{P_8 = ababbabb}$

$\pi(9) = 1$ (P_6 es el único que acaba en bba pero no es sufijo por la primera letra)

$$\boxed{P_9 = \underbrace{a}_{P_1}babbabb\underbrace{a}_{P_1}}$$

$\pi(10) = 2$ (P_4 y P_7 acaban en bab pero no son sufijos de P_{10} y P_2 sí lo es)

$$\boxed{P_{10} = \underbrace{ab}_{P_2}abbabb\underbrace{ab}_{P_2}}$$

$\pi(11) = 3$ (El único prefijo que acaba en aba es P_3)

$$\boxed{P_{11} = \underbrace{aba}_{P_3}bbabb\underbrace{aba}_{P_3}}$$

$\pi(12) = 4$ (El único prefijo que acaba en abab es P_4)

$$\boxed{P_{12} = \underbrace{abab}_{P_4}babb\underbrace{abab}_{P_4}}$$

$\pi(13) = 5$ (Los prefijos que terminan en abb son P_8 y P_5 pero P_8 no es sufijo y P_5 sí)

$$\boxed{P_{13} = \underbrace{ababb}_{P_5}abb\underbrace{ababb}_{P_5}}$$

Ejercicio 2.- Calcular los conjuntos $\pi^+(q)$ y E_q para dicho patrón.

Recordamos que $\pi^+(q) = \{\pi^i(q) \mid i \geq 1, \pi^{i-1}(q) \neq 0\}$ con π^i definido de manera usual $\pi^i(q) = \begin{cases} \pi(\pi^{i-1}(q)) & \text{si } i \neq 0 \\ q & \text{si } i = 0 \end{cases}$

Está probado que $\pi^+(q) = \{k \mid k < q, P_k \sqsupset P_q\}$ es decir, $\pi^+(q)$ es el conjunto de las longitudes de los prefijos de P que son sufijos de P_q.

Por otro lado, dado $q \in \{2 \cdots m\}$ se define

$E_{q-1} = \{k \mid k \in \pi^+(q-1), P[k+1] = P[q]\} = \{k \mid k < q-1, P_k \sqsupset P_{q-1}, P[k+1] = P[q]\}$,

es decir, el conjunto de las longitudes de los prefijos de P que son sufijos de P_{q+1} y se pueden extender con un caracter más para ser sufijos de P_q.

($E_q \subset \pi^+(q)$ y $\forall k \in E_q \Rightarrow P[k+1] = P[q+1]$)

Con estas ideas podemos calcular E_q y $\Pi^+(q)$ $\forall q \in \{1 \dots m\}$.

$\Pi^+(1) = \{0\}$; $E_1 = \emptyset$ $(P_1 = a)$

$\Pi^+(2) = \{0\}$; $E_2 = \{0\}$ $(P_2 = ab)$ (P_1 no es sufijo) $(P[1] = P[3])$

$\Pi^+(3) = \{0,1\}$; $E_3 = \{1\}$ $(P_3 = aba)$ (P_2 no es sufijo, P_1 sí) $(P[1] \neq P[4])$
$(P[2] = P[4])$

$\Pi^+(4) = \{0,2\}$; $E_4 = \emptyset$ $(P_4 = abab)$ (P_1 no es sufijo, P_2 sí, P_3 no)

$\Pi^+(5) = \{0\}$; $E_5 = \{0\}$ $(P_5 = ababb)$

$\Pi^+(6) = \{0,1\}$; $E_6 = \{1\}$ $(P_6 = ababba)$

$\Pi^+(7) = \{0,2\}$; $E_7 = \emptyset$ $(P_7 = ababbab)$

$\Pi^+(8) = \{0\}$; $E_8 = \{0\}$ $(P_8 = ababbabb)$

$\Pi^+(9) = \{0,1\}$; $E_9 = \{1\}$ $(P_9 = ababbabba)$

$\Pi^+(10) = \{0,2\}$; $E_{10} = \{0,2\}$ $(P_{10} = ababbabbab)$

$\Pi^+(11) = \{0,1,3\}$; $E_{11} = \{1,3\}$ $(P_{11} = ababbabbaba)$

$\Pi^+(12) = \{0,2,4\}$; $E_{12} = \{4\}$ $(P_{12} = ababbabbabab)$

$\Pi^+(13) = \{0,5\}$; E_{13} no está def. $(P_{13} = ababbabbababb)$

Ejercicio 3.- Repetir el cálculo de Π, pero ahora utilizando la fase de precondicionamiento del algoritmo KMP.

El algoritmo de preprocesado era:

```
preprocesado-Knuth-Morris-Pratt (P[1..m], out pi[1..m])
  p[1]=0;
  K=0;
  for (q=2 to m){
      do while (k>0 ^^ P[k+1] != P[q]){
            K= pi[K];
      }
      if (P[k+1]= P[q]){ k++;}
      pi[q]=K;
  }
  return pi;
}
```

El valor inicial de pi es: pi = | 0 | - | - | - | - | - | - | - | - | - | - | - | - |

y su evolución en el paso q al principio del bucle for es:

q	1	2	3	4	5	6	7	8	9	10	11	12	13	K
2	0	-	-	-	-	-	-	-	-	-	-	-	-	0
3	0	0	-	-	-	-	-	-	-	-	-	-	-	0
4	0	0	1	-	-	-	-	-	-	-	-	-	-	1
5	0	0	1	2	-	-	-	-	-	-	-	-	-	2
6	0	0	1	2	0	-	-	-	-	-	-	-	-	0
7	0	0	1	2	0	1	-	-	-	-	-	-	-	1
8	0	0	1	2	0	1	2	-	-	-	-	-	-	2
9	0	0	1	2	0	1	2	0	-	-	-	-	-	0
10	0	0	1	2	0	1	2	0	1	-	-	-	-	1
11	0	0	1	2	0	1	2	0	1	2	-	-	-	2
12	0	0	1	2	0	1	2	0	1	2	3	-	-	3
13	0	0	1	2	0	1	2	0	1	2	3	4	-	4
/	0	0	1	2	0	1	2	0	1	2	3	4	5	5
	1	2	3	4	5	6	7	8	9	10	11	12	13	

El vector pi toma los mismos valores que en el ejercicio 1.

A modo de resumen reunimos la información de Π, Π^+ y E en una misma tabla.

q	$\Pi(q)$	$\Pi^+(q)$	E_q	P_q
1	0	$\{0\}$	$\emptyset$	a
2	0	$\{0\}$	$\{0\}$	ab
3	1	$\{0,1\}$	$\{1\}$	aba
4	2	$\{0,2\}$	$\emptyset$	abab
5	0	$\{0\}$	$\{0\}$	ababb
6	1	$\{0,1\}$	$\{1\}$	ababba
7	2	$\{0,2\}$	$\emptyset$	ababbab
8	0	$\{0\}$	$\{0\}$	ababbabb
9	1	$\{0,1\}$	$\{1\}$	ababbabba
10	2	$\{0,2\}$	$\{0,2\}$	ababbabbab
11	3	$\{0,1,3\}$	$\{1,3\}$	ababbabbaba
12	4	$\{0,2,4\}$	$\{4\}$	ababbabbabab
13	5	$\{0,5\}$		ababbabbababb

1 2 3 4 5 6 7 8 9 10 11 12 13

Observamos que:

$$\Pi(q) = \max_{K \in \Pi^+(q)} \{K\} = \begin{cases} 0 & \text{si } E_{q-1} = \emptyset \\ 1 + \max\limits_{K \in E_{q-1}} \{K\} & \text{si } E_{q-1} \neq \emptyset \end{cases}$$