

Лабораторна робота № 5. Алгоритми злиття та пошуку.

Мета роботи: Навчитись реалізовувати різні алгоритми пошуку.

Теоретичні відомості.

Алгоритм Рабіна-Карпа.

Ідея, запропонована Рабіном і Карпом, полягає в тому, щоб поставити у відповідність кожному рядку деяке унікальне число, і замість того, щоб порівнювати самі рядки, порівнювати числа, що набагато швидше.

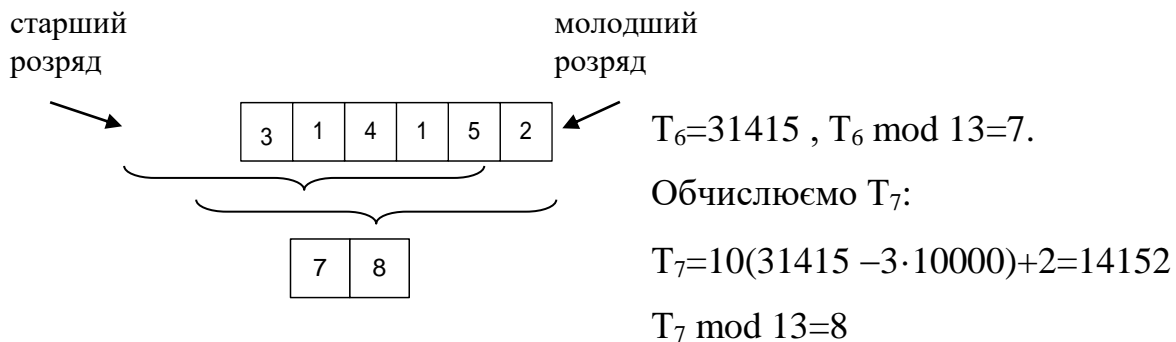
Задано рядок $T=abcdeabfgfcakmbddaf$ і підрядок, який треба знайти $P=bfgfc$ ($m=5$). $\{0,1\dots 9\}$ – алфавіт, $d=10$, $h=d^{m-1}=10^4=10000$; $q=13$.

b	f	g	f	c
3	1	4	1	5

$$P=31415.$$

$$P_q=(5+10(1+10(4+10(1+10\cdot 3)))) \bmod 13=31415 \bmod 13=7$$

Необхідно шукати підрядки, які дорівнюють 7 по модулю 13.



Обчислимо значення T_S для всього тексту:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
a	b	c	d	e	a	b	f	g	f	c	a	k	m	b	d	d	a	f
2	3	5	9	0	2	3	1	4	1	5	2	6	7	3	9	9	2	1

...

...

...

8	9	3	11	0	1	7	8	4	5	10	11	7	9	11
---	---	---	----	---	---	---	---	---	---	----	----	---	---	----

Відповідь: Зразок P входить в текст T зі зсувом $S=6$, позиція $i=7$.

Алгоритм Кнута-Морриса Пратта (КМП).

Алгоритм КМП складається з двох етапів: підготовчого (побудова префікс-функції) і основного (пошуку).

Задано: текст $T= ababaababababcbababababcbabb$ ($\text{length}(T)=27$).

підрядок, який треба знайти $P=ababababca$ ($\text{length}(P)=10$).

Підготовчий етап.

Для довільного слова X розглянемо всі його префікси, що одночасно є суфіксами і серед них оберемо найдовший (не враховуючи самого X). Його будемо позначати $n(X)$ і називати найбільшим префікс-суфіксом.

Знайти повну префіксну функцію f для зразка $P=ababababca$.

$n(a)=L, \quad f(1)=0$

$n(ab)=L, \quad f(2)=0$

$n(aba)=a, \quad f(3)=1$

$n(abab)=ab, \quad f(4)=2$

$n(ababa)=aba, \quad f(5)=3$

$n(ababab)=abab, \quad f(6)=4$

$n(abababa)=ababa, \quad f(7)=5$

$n(abababab)=ababab, \quad f(8)=6$

$n(ababababc)=L, \quad f(9)=0$

$n(ababababca)=a, \quad f(10)=1$

або

i	1	2	3	4	5	6	7	8	9	10
$P[i]$	a	b	a	b	a	b	a	b	c	a
$f[i]$	0	0	1	2	3	4	5	6	0	1

Основний етап (пошук).

$f[i]=[0,0,1,2,3,4,5,6,0,1]$.

1) Зсув $S_0=0, q=5, f[q]=f[5]=3,$

$S_1=S_0+(q-f(q))=0+(5-3)=2, i=3.$

2) Зсув $S_1=2, q=3, f[q]=f[3]=1,$

$S_2=S_1+(q-f(q))=2+(3-1)=4, i=5.$

3) Зсув $S_2=4, q=1, f[q]=f[1]=0,$

$S_3=S_2+(q-f(q))=4+(1-0)=5, i=6.$

4) Зсув $S_3=5, q=9, f[q]=f[9]=0,$

$S_4=S_3+(q-f(q))=5+(9-0)=14, i=15.$

5) Зсув $S_4=14, q=0,$

$S_5=S_4+1=15, i=16.$

6) Зсув $S_5=15, q=10, f[q]=f[10]=1,$

$S_6=S_5+(q-f(q))=15+(10-1)=24, i=25.$ (Пошук зупиняємо, так як закінчився текст T).

T	ababaababababcbababababcbabb
P	<u>ababababca</u> <u>ababababca</u> <u>ababababca</u> <u>ababababca</u> ababababca <u>ababababca</u>

Відповідь: Зразок P входить в текст T зі зсувом $S=15$, позиція $i=16$.

Завдання.

Скласти схеми алгоритмів та написати програми реалізації алгоритмів Рабіна-Карпа та КМП. Провести порівняльний аналіз алгоритмів пошуку в рядках.

Контрольні запитання.

1. Ідея алгоритмів Рабіна-Карпа та КМП.
2. З яких етапів складається алгоритм КМП? Охарактеризувати кожний етап.
3. Що таке префіксна функція?