Question 2: Horspool’s Algorithm

Create shift table for the pattern:

**TCCTATTCTT**

The shift table shows the number of characters to shift by in the pattern, until you get to the desired character. This is the number of characters to shift until the first occurrence of the target.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Character** | T | C | A | \* |
| **Shift** | 1  Expect 1 | 2  Expect 2 | 5  Expect 5 | 10  Expect 10 |

Second for loop:

When j = 0: Got a T: Table[T] = 10 – 1 – 0 = 9

When j = 1: Got C: Table[C] = 10 – 1 – 1 = 8

When j = 2: Got C: Table[C] = 10 – 1 – 2 = 5

When j = 3: Got T: Table[T] = 10 – 1 – 3 = 6

When j = 4: Got A: Table[A] = 10 – 1 – 4 = 5

When j = 5: Got T: Table[T] = 10 – 1 – 5 = 4

When j = 6: Got T: Table[T] = 10 – 1 - 6 = 3

When j = 7: Got C: Table[C] = 10 – 1 – 7 = 2

When j = 8: Got T: Table[T] = 10 – 1 – 8 = 1

Stop, since j == 10 – 2.

Apply the Horspool’s algorithm

TTATAGATCTCGTATTCTTTTATAGATCTCCTATTCTT

TCCTATTCTT

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Character** | T | C | A | \* |
| **Shift** | 1  Expect 1 | 2  Expect 2 | 5  Expect 5 | 10  Expect 10 |

TTATAGAT**CT**CGTATTCTTTTATAGATCTCCTATTCTT

TCCTATTC**TT**

* T matches with T, so move on. C does not match with T, so shift by Table[T] = 1

TTATAGATCT**C**GTATTCTTTTATAGATCTCCTATTCTT

TCCTATTCT**T**

* C does not match with T immediately, so shift by Table[C] = 2

TTATAGATCTC**GT**ATTCTTTTATAGATCTCCTATTCTT

TCCTATTC**TT**

* T matches with T, so move on. G does not match with T, so shift by Table[T] = 1

TTATAGATCTCGT**A**TTCTTTTATAGATCTCCTATTCTT

TCCTATTCT**T**

* A does not match with T immediately, so shift by Table[A] = 5

TTATAGATCTC**GTATTCTT**TTATAGATCTCCTATTCTT

TC**CTATTCTT**

* From the right, T,T,C,T,T,A,T all match. G does not match with C, so shift by Table[T] = 1

TTATAGATCTCGTATTC**TTT**TATAGATCTCCTATTCTT

TCCTATT**CTT**

* From the right, T, T all match. C does not match with T, so shift by Table[T] = 1
* For the next step, also shift by 1.

TTATAGATCTCGTATTCTTTT**A**TAGATCTCCTATTCTT

TCCTATTCT**T**

* A mismatched T immediately, shift by Table[A] = 5

TTATAGATCTCGTATTCTTTTATAG**AT**CTCCTATTCTT

TCCTATTC**TT**

* T matches, so move on. A mismatched T, so shift by Table[T] = 1

TTATAGATCTCGTATTCTTTTATAGAT**C**TCCTATTCTT

TCCTATTCT**T**

* C does not match T immediately, so shift by Table[C] = 2

TTATAGATCTCGTATTCTTTTATAGATCT**C**CTATTCTT

TCCTATTCT**T**

* C does not match T immediately, so shift by Table[C] = 2

TTATAGATCTCGTATTCTTTTATAGATCTC**CT**ATTCTT

TCCTATTC**TT**

* T matches with T. C does not match with T, so shift by Table[T] = 1

TTATAGATCTCGTATTCTTTTATAGATCTCCT**A**TTCTT

TCCTATTCT**T**

* A does not match with T immediately, so shift by Table[A] = 5

TTATAGATCTCGTATTCTTTTATAGATC**TCCTATTCTT**

**TCCTATTCTT**

* All matches, and pattern found at index 29

Question 3:

001001001

00001

Bad symbol shift table:

[0, 1], [\*, 5]

Good suffix table:

[1, 5], [01, 5], [001, 5], [0001, 5]

Task 1 & 2: Since the text only contains 0, always shift by the bad symbol 0 (by 1 each time).

Number of comparisons = 1000 – 4 = 996