CS 3520: Algorithms

Homework 6

**Due Date: Monday, April 22, 2019 at beginning of class**

***Please write legibly. Show your steps to receive partial credit.***

1. (30 points) Horspool’s algorithm
   1. Construct the shift table for Horspool’s algorithm for the pattern:

IMPOSSIBLE

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| B | I | L | M | O | P | S | OTHERS |
| 2 | 3 | 1 | 8 | 6 | 7 | 4 | 10 |

* 1. Show the actual search in the following text by marking where the end of the pattern aligns in the text with a down arrow ↓and writing the corresponding and shift size for each shift. **Notice that punctuation marks and spaces between two words and after each semicolon count as characters**.

↓ ↓ ↓ ↓ ↓ ↓

START BY DOING WHAT'S NECESSARY; THEN DO WHAT'S POSSIBLE;

↓ ↓ ↓ ↓

AND SUDDENLY YOU ARE DOING THE **IMPOSSIBLE**.

IMPOSSIPLE 10

IMPOSSIBLE 10

IMPOSSIBLE 10

IMPOSSIBLE 6

IMPOSSIBLE 10

IMPOSSIBLE 10

POSSIBLE 10

IMPOSSIBLE 10

IMPOSSIBLE 10

IMPOSSIBLE 3

IMPOSSIBLE ✓

1. (40 points) Boyer-Moore algorithm
   1. The alphabet for DNA sequences is {}. Construct the bad-symbol table and the good-suffix table for Boyer-Moore algorithm for the following gene segment.

Bad: Good:

|  |  |  |
| --- | --- | --- |
| K | Pattern | D2 |
| 1 | TCCTA**TT** | 1 |
| 2 | **T**CCTA**TT** | 6 |
| 3 | **T**CCT**ATT** | 6 |
| 4 | **T**CC**TATT** | 6 |
| 5 | **T**C**CTATT** | 6 |
| 6 | **TCCTATT** | 6 |

|  |  |  |  |
| --- | --- | --- | --- |
| A | C | T | OTHERS |
| 2 | 4 | 1 | 7 |

* 1. Apply Boyer-Moore algorithm to locate the above pattern in the following DNA sequence. Mark where the end of the pattern aligns in the text with a down arrow ↓and write the corresponding shift size for each step.

↓ ↓ ↓ ↓ ↓ ↓ ↓

TTATAGATCTCGTATTCTTTTATAGATC**TCCTATT**CTT

TCCTATT 2

TCCTATT 4

TCCTATT 6

TCCTATT 6

TCCTATT 7

TCCTATT ✓

At A, d1 = mac{t(A) – 0,1} = max{2, 1} = 2

At A, d1 = mac{t(C) – 0,1} = max{2, 1} = 2

At A, d1 = mac{t(G) – 0,1} = max{2, 1} = 2

At A, d1 = mac{t(C) – 0,1} = max{2, 1} = 2

At A, d1 = mac{t(A) – 0,1} = max{2, 1} = 2

1. (30 points) For the input 49, 20, 56, 75, 89, 88, 62 and hash function ,

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Key | 49 | 20 | 56 | 75 | 89 | 88 | 62 |
| *h*(K) | 10 | 7 | 4 | 10 | 11 | 10 | 10 |

* 1. Construct the open hash table.

0 1 2 3 4 5 6 7 8 9 10 11 12

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

↓ ↓ ↓ ↓

56 20 49 89

↓

75

↓

88

↓

62

* 1. Construct the closed hash table.

0 1 2 3 4 5 6 7 8 9 10 11 12

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  | 49 |  |  |
|  |  |  |  |  |  |  | 20 |  |  | 49 |  |  |
|  |  |  |  | 56 |  |  | 20 |  |  | 49 |  |  |
|  |  |  |  | 56 |  |  | 20 |  |  | 49 | 75 |  |
|  |  |  |  | 56 |  |  | 20 |  |  | 49 | 75 | 89 |
| 88 |  |  |  | 56 |  |  | 20 |  |  | 49 | 75 | 89 |
| 88 | 62 |  |  | 56 |  |  | 20 |  |  | 49 | 75 | 89 |