Advanced Algorithms in Bioinformatics (P4)

Sommersemester 2012

1. Exercise sheet, 10. April 2012

1. Horspool
	1. Algorithm-example

|  |  |  |  |
| --- | --- | --- | --- |
| bmBc:a c g t2 1 6 6attempt 1:acactcccccgacaaccACAa.CShift by 1 (bmBc[c])attempt 2:acactcccccgacaacc a....CShift by 1 (bmBc[c]) | attempt 3:acactcccccgacaacc ACa..CShift by 1 (bmBc[c])attempt 4:acactcccccgacaacc a....CShift by 1 (bmBc[c])attempt 5:acactcccccgacaacc a....CShift by 1 (bmBc[c]) | attempt 6:acactcccccgacaacc .....cShift by 6 (bmBc[g])attempt 7:acactcccccgacaacc ACAACCShift by 1 (bmBc[c])acactcccccgACAACC | String length: 17Pattern length: 6Attempts: 7Charactercomparisons: 22 |

* 1. Performance
		1. Performs so bad, because of the fact that the last character of the pattern occurs so often in the text.
		2. Would perform much better, if the pattern would be compared completely from the end
1. Wu-Manber
	1. Example
		1. AGATAGACGATATATACG -> Shift[TA]=0, Hash[TA]=3, no match
		2. AGATAGACGATATATACG -> Shift[AG]=3
		3. AGATAGACGATATATACG -> Shift[CG]=0, Hash[CG]=1, match
		4. AGATAGACGATATATACG -> Shift[GA]=2
		5. AGATAGACGATATATACG -> Shift[TA]=0, Hash[TA]=3, no match
		6. AGATAGACGATATATACG -> Shift[AT]=0, Hash[AT]=2, no match
		7. AGATAGACGATATATACG -> Shift[TA]=0, Hash[TA]=3, no match
		8. AGATAGACGATATATACG -> Shift[AT]=0, Hash[AT]=2, match
		9. AGATAGACGATATATACG -> Shift[TA]=0, Hash[TA]=3, match
		10. AGATAGACGATATATACG -> Shift[AC]=1
		11. AGATAGACGATATATACG -> Shift[CG]=0, Hash[TA]=1, no match
	2. Performance
		1. If the Suffix is of one pattern is the suffix-1 of another pattern
	3. Prove
		1. Yes, this case does occur, because the block size would be again 2.
		2. log4(2\*5\*3) = 2,453 = 2
2. Multiple Horspool
	1. Example