Assignment 5 Problem 4

a) Construct the last occurrence function *L* for pattern *P* = *adobodoa* where Σ = *a, b, c, d, o, t*.

b) Trace the search for *P* in *T* = *dotadotadotdotadobodoadot* using the Boyer-Moore algorithm. Indicate in a table such as Table **2** which characters of *P* were compared with which characters of *T*. Follow the example from the lecture video in Module 9. Place each character of *P* in the column of the compared-to character of *T* . Put brackets around the character if they are known to match from the previous step (similar to the examples in the slides). Use a new row when sliding the pattern. You may not need all rows in the table. Add more rows to the table if you need more.

c) For any $m \geq 1$ and any $n \geq m$, give a pattern *P* and a text *T* such that the Boyer-Moore algorithm looks at exactly $\Theta(n/m)$ characters. Justify your answer.

Solution:

a) L(a) = 7
L(b) = 3
L(c) = −1
L(d) = 5
L(o) = 6
L(t) = −1

b)

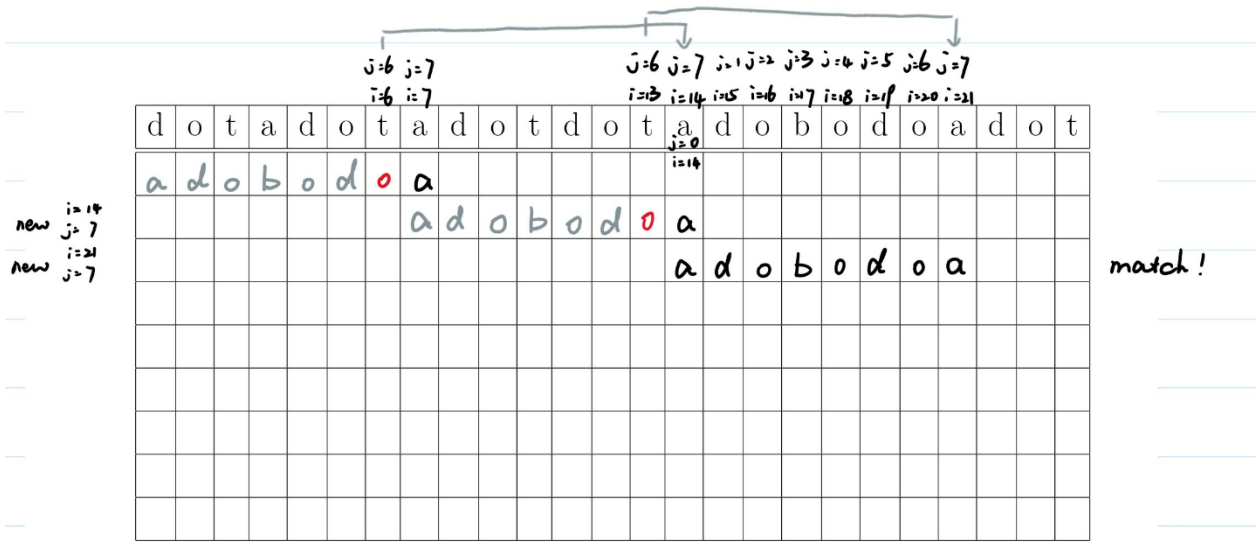

Table 2: Table for Boyer-Moore problem.

c) For any pattern P of length m, and text T of length n. If all characters on position p in T where ($p \equiv m - 1 \mod m$) are the characters that does not occur in P (i.e. L(T[p]) = -1). Then the Boyer-Moore algorithm looks at exactly $\Theta(n/m)$ characters because in this case, Boyer-Moore algorithm will only look at the characters that are in position p, and will find that it does not occur, then it will shift i into i + m. Since each time it shifts m position, totally, it will only check $\Theta(n/m)$ times.