

# Exercise 4

Ewa Szczurek  
MIM UW

October 27, 2015

**Note 1.** How to hand in homework:

- Format the title of your email: HOMEWORK; your first and last name; lab number
- Put all exercises for this homework into one single email
- Run all programming code checking that it executes correctly before submitting
- Provide calling examples: examples showing how to call your functions, showing that the program does what it is supposed to do
- Run your calling examples before submitting
- Comment your code
- All solutions which do execute and which do what they are supposed to count, although code elegance and efficiency is much appreciated
- Program either in R or in python
- Submit to `szczurek@mimuw.edu.pl` by the next week.

**Exercise 1.** BWT construction.

1. Inspect the code for building BWT over strings
2. How does the *bwtViaBwm* function work?
3. What is the worst case time complexity of this function?
4. What could be its complexity if we used the efficient SA build?

**Exercise 2.** BWT reversing.

1. Inspect the code for reversing BWT
2. How does the *reverseBwt* function work?

3. What is its memory complexity?
4. Time complexity?
5. How could it be implemented in a more memory efficient-manner?

**Exercise 3.** FM index

1. Inspect the code for reversing the FM index
2. Consider a string  $T = \text{'alamakotaimapsa'}$ .
3. Use the checkpoints for  $T$  to answer the question: what is the number of occurrences of 'a' up to and including offset 5?
4. Use the FM index to answer the question whether "chomika" is a substring of  $T$ .
5. Use the FM index to answer the question: Where does the string 'ma' occur in  $T$ ?

**Homework 1.** Word occurrences in text.

1. Split the text in the text.txt file into words.
2. Report the occurrences of five most frequently used words of length larger than 4. Use the FM index for this task.