Exercise Sheet 3
complete until Thursday, November 9th

Exercise 1

For each of the following four index data structures, state construction time, space requirement, and query time. Space requirement should be measured in bits, using the fact that numbers from a range $1..n$ can be represented with $\log_2 n$ bits. Give an estimate of the constant factor for each of the three bounds, that is, don’t just write $O(\ldots)$. Assume that basic operations like comparing two numbers, looking up a number in a table, etc. take one unit of time. Give a brief explanation of your estimate. Here are the four index data structures:

(a) The suffix array from the Manber/Myers paper (with lcps but without the modifications needed for expected linear running time).

(b) The same suffix array, but taking a sequence of words (not letters) as input, and considering the set of distinct words as alphabet. Each word is then one symbol.

   Which queries does such a suffix array naturally support? Query time should be with respect to these queries.

(c) An ordinary inverted index for a sequence of words (for each word, a list of the occurrences of that word in the sequence). Query time should be with respect to the same queries as in (b).

(d) An extension, of your choice, of an inverted index that can deal with arbitrary substring queries (substrings within a single word, not substrings spanning several words).

   Give a concise description of your extension.

Make a nice, uniform $4 \times 3$ table out of these figures. Stare at the table for at least a minute and then discuss it a little bit.

There will be a template for such a table on the Wiki. If there are cells which are not yet filled, pick one and write your entry there (and make sure it is correct).

Also upload a PDF with all your solutions for this sheet to the Wiki.