

Sequence Analysis 3
Summer 2024

Exercises

Number 1, Discussion: 2024-April-18

1. Answer the following simple questions:
 - (a) What is the maximum number of non-empty substrings a string s of length n can have?
 - (b) What is the maximum number of non-empty subsequences a string s of length n can have?
 - (c) What is the number of prefixes of a string s of length n ? And the number of proper prefixes?
 - (d) What is the number of suffixes of a string s of length n ? And the number of proper prefixes?
 - (e) Given an alphabet of size σ , how many different strings of length n are there?
2. If they exist, find words
 - (a) of lengths 2, 4 and 10 that are primitive;
 - (b) of lengths 2 and 5 with a period of length 1;
 - (c) of lengths 5 and 10 whose only period is of length 4;
 - (d) of lengths 6 and 10 whose only periods are of lengths 3 and 5.
3. Compute the
 - (i) Hamming distance
 - (iii) (unit cost) edit distance
 - (ii) subsequence distance
 - (iv) prefix distancebetween the following two pairs of strings:
 - (a) aabbabaabababa and abbaabaabaabba,
 - (b) abababaaababab and abbbbbaabbab.
4. Compress the string
aabaaaababbabaaabaaaababbabbabbbbabaaaabaaaabb
using
 - (i) Run-Length Encoding,
 - (ii) Move-To-Front + Run-Length Encoding,
 - (iii) Delta Encoding,
 - (iv) Move-To-Front + Delta Encoding,
 - (v) Lempel-Ziv (1977) Compression with infinite window size,
 - (vi) Lempel-Ziv (1978) Compression