A flexible tool for Shape Analysis

IDP Presentation

Emanuel Laude
Zorah Lähner

Technische Universität München

January 9, 2015

Motivation

Observation 1

Analyzing biological sequences for their statistical properties can lead to interesting conclusions:

- Coding vs. non-coding regions (CpG content, entropy, . . .)
- Species identification
- Number and length of repeats

Observation 2

Many algorithms and data structures depend on statistical properties of the underlying sequence:

- Algorithms for sequence assembly: repeats
- Suffix tree: repeats
- q-gram index: number and distribution of q-grams
- Compressed index structures: entropy

Motivation

Observation 1

Analyzing biological sequences for their statistical properties can lead to interesting conclusions:

- Coding vs. non-coding regions (CpG content, entropy, . . .)
- Species identification
- Number and length of repeats

Observation 2

Many algorithms and data structures depend on statistical properties of the underlying sequence:

- Algorithms for sequence assembly: repeats
- Suffix tree: repeats
- q-gram index: number and distribution of q-grams
- Compressed index structures: entropy

Logo

The logo of the Technische Universität München looks as follows:

Logo

The logo of the Technische Universität München looks as follows:

Technische Universität München 

Discussion

- Summary
  - First
  - Second
- Outlook
  - First
  - Second

Discussion

- Summary
  - First
  - Second
- Outlook
  - First
  - Second

- Thank you!
- Questions?

- Thank you!
- Questions?