

Bio informatics

Memo

1. Dynamic Programming

- PairwiseAlignment
- Needleman-Wunsch algorithms
- Smith-Waterman algorithms
- Affin gap score

2. Hash

- hash function
- open addressing
- Bloom filter
- Distribution counting sort
- Radix sort

3. Super Computer

- scale-up and scale-out

4. Graph Algorithm

- Undirected graph
- chess's knight
- Path
- Directed graph
- Benzer and interval graph
 - It has one vertex for each interval in the family, and an edge between vertices whose intervals intersect.
- Shortest superstring problem
 - not useful because it must has sequence errors
- Traveling salesman problem

5. ZDD

- introduction
- ZDD's structure
- ZDD's advantages
- How to make ZDD tree
- Spanning Tree

6. How to manage big biological data

- succinct data structure
 - insert + delete + search + rank + select
 - "rank" is a number of 'x's in a given string
 - "select" is a position of 'i'th 'x'
- AND query
 - alternation α
 - range intersection by using Wavelet tree $\rightarrow O(\alpha \log m)$

7. Probability models

- Basic knowledge of DNA
- Find transferred DNA segments by adhoc methods
- A probability model with states and transitions
 - hidden markov model
 - how to calculate the probability
 - Finding the most likely path (using dynamic programming)
 - * the most likely path from the 1st letter to the i th letter, ending in state X
 - * store the X_i , to find the path
 - * maximize the probability
 - * Viterbi Algorithm
- Score-based sequence analysis

8. Evolution

- four ways to determine the family tree
 - Bayes
 - Maximum likelihood
 - Maximum parsimony
 - Distance method