Università degli Studi di Milano - Bicocca
Scuola di Scienze
Dipartimento di Informatica, Sistemistica e Comunicazione
Corso di Laurea Magistrale in Informatica

Algoritmi per la trasformata di Burrows-Wheeler Posizionale con compressione run-length, RLPBWT

Relatore: *Prof.ssa Raffaella Rizzi*

Correlatore:

Tesi di Laurea Magistrale di:
Davide Cozzi
Matricola 829827

Anno Accademico 2021-2022

Abstract

Indice

Capitolo 1

Introduzione

1.1 Motivazioni Biologiche

1.2 Bitvector sparsi

1.3 Straight-Line Program

1.3.1 Random access

1.3.2 Longest Common Extension

1.4 Trasformata di Burrows-Wheeler

1.4.1 Trasformata di Burrows-Wheeler run-length

1.4.2 Matching Statistics

1.4.3 R-index

1.4.4 MONI

1.4.5 PHONI

1.5 Trasformata di Burrows-Wheeler posizionale

1.5.1 Implementazione originale

Gli algoritmi di Durbin

Limiti spaziali

1.5.2 Varianti della PBWT 5

PBWT multi-allelica

PBWT con struttura LEAP

PBWT dinamica

PBWT bidirezionale

Recenti sviluppi

Capitolo 2

Metodo

2.1 Introduzione agli strumenti usati

2.1.1 SDSL

2.1.2 BigRepair

2.1.3 ShapedSlp

Ricostruzione del panel

2.2 Introduzione alle varianti della RLPBWT

2.2.1 Perché un'implementazione run-length

2.3 Mapping nella RLPBWT

2.4 RLPBWT naive

2.4.1 Algoritmo per match massimali

2.5 RLPBWT con bitvectors

2.5.1 Algoritmo per match massimali

2.6 RLPBWT con pannello

2.6.1 Algoritmo con matching statistics

2.7 RLPBWT con SLP

2.7.1 Algoritmo con matching statistics

2.8 Funzione Phi

2.8.1 Costruzione della struttura di supporto

2.8.2 Estensione dei match

Capitolo 3

Risultati

3.1 Ambiente di benchmark

3.1.1 Descrizione input

3.2 Analisi temporale

3.3 Analisi spaziale

Capitolo 4

Conclusioni

4.1 Sviluppi futuri

4.1.1 K-mems

4.1.2 RLPBWT multi-allelica

4.1.3 RLPBWT con dati mancanti

Appendice A

Pseudocodici

Appendice B

Tabelle