



Algoritmi di compressione

BURROWS-WHEELER TRANSFORM CON HUFFMAN CANONICO

Bonesana Claudio

Nathan Koefer

SUPSI DTI

21.01.2012

INTRODUZIONE

○ Richieste:

- Creazione di un compressore
- Implementazione dell'algoritmo BWT
- Implementazione dell'algoritmo di Huffman Canonico
- Esecuzione test
- Analisi dei risultati
- (ev.) Ottimizzazione del compressore

SCELTE COMPIUTE: BWT

- Algoritmo di ordinamento
 - Lavora su singoli caratteri (byte)
 - Non comprime
 - Restituisce un testo ottimizzato per una successiva compressione
 - Può essere invertito

SCELTE COMPIUTE: HUFFMAN CANONICO

- Algoritmo di compressione
 - Lavora con i caratteri (byte)
 - Sfrutta la frequenza di apparizione
 - Assegna codici corti a caratteri frequenti
 - Sfrutta un albero per la generazione dei codici
 - Necessita di informare il decompressore sui codici utilizzati
 - Non sfrutta vocabolari

SCELTE COMPIUTE: STRUTTURE DATI

○ Scelte:

- Array come buffer dell'intero file 🙌(°°🙌)
- Array per conteggio
- Array per le frequenze
- Albero per l'ordinamento
- Albero per Huffman
- Array per i codici di Huffman

○ Utilizzate:

- Struct per le frequenze
- Albero per Huffman con struct delle frequenze
- Struct per i codici
- Array di struct per codici

ALGORITMI AUSILIARI UTILIZZATI

- Scelti
 - QuickSort
 - Creazione albero dall'alto
- Utilizzati
 - MergeSort (stabile)
 - Creazione albero dalle foglie

PROBLEMI INCONTRATI

- Utilizzo non efficiente degli array di contatori
 - Occupazione inutile di memoria
 - Cicli inutili e spesso errati
- Codifiche errate
 - Albero creato in maniera errata
- Gestione dei file
 - Memorizzazione dei caratteri
 - Scrittura dei codici
- Scrittura di singoli bit
 - C può scrivere al minimo byte
- Trasferimento dei codici

PROBLEMI RISOLTI

- Utilizzo non efficiente degli array di contatori
 - Passaggio alle struct
 - Utilizzo di un sort corretto
- Codifiche errate
 - Albero creato a partire dalle foglie
- Gestione dei file
 - Lettura senza memorizzazione
 - Utilizzo di buffer temporanei di bit
- Scrittura di singoli bit
 - Utilizzo di maschere di bit
- Trasferimento dei codici
 - Trasmissione delle singole lunghezze del set di caratteri pre-impostato

PROBLEMI ANCORA APERTI

- Decompressore
- Come unire in maniera utile il BWT ad Huffman
 - BWT ordina i caratteri
 - Huffman comprime i caratteri
 - $(\cup \circ \square \circ) \cup \frown \text{—}\text{—}\text{—}\text{—}$
- Utilizzo di un Huffman a più caratteri?
 - Conviene?
 - Con quanti caratteri?
 - Su quali grandezze di file?

ESEMPI

- Compressione di alcuni file solo con Huffman

```
dna@Mordor:~/workspace/huffman/Debug$ ls -l
totale 2780
-rw----- 1 dna dna 613730 2011-10-28 10:49 Avast Fluttershy's Yay.mp3
-rw-rw-r-- 1 dna dna 604501 2011-11-15 15:08 Avast Fluttershy's Yay.mp3.thuf
-rw-rw-r-- 1 dna dna 97590 2011-10-21 18:24 fluttershy-yay.jpg
-rw-rw-r-- 1 dna dna 96308 2011-11-15 15:08 fluttershy-yay.jpg.thuf
-rwxrwxr-x 1 dna dna 7241 2011-11-06 13:23 generarogna
-rw-rw-r-- 1 dna dna 173 2011-11-06 13:22 generarogna.c
-rwxrwxr-x 1 dna dna 109096 2011-11-10 10:53 huffman
-rw-rw-r-- 1 dna dna 184265 2011-11-06 13:17 inferno.txt
-rw-rw-r-- 1 dna dna 104368 2011-11-15 15:08 inferno.txt.thuf
-rw-rw-r-- 1 dna dna 2484 2011-11-06 13:17 introduz.txt
-rw-rw-r-- 1 dna dna 1806 2011-11-15 15:08 introduz.txt.thuf
-rw-rw-r-- 1 dna dna 984 2011-11-10 11:03 makefile
-rw-rw-r-- 1 dna dna 231 2011-11-10 10:53 objects.mk
-rw-rw-r-- 1 dna dna 181870 2011-11-06 13:17 paradiso.txt
-rw-rw-r-- 1 dna dna 102062 2011-11-15 15:08 paradiso.txt.thuf
-rw-rw-r-- 1 dna dna 192226 2011-10-28 10:38 ProgMC_serie5-7.pdf
-rw-rw-r-- 1 dna dna 188442 2011-11-15 15:08 ProgMC_serie5-7.pdf.thuf
-rw-rw-r-- 1 dna dna 184858 2011-11-06 13:17 purgator.txt
-rw-rw-r-- 1 dna dna 104235 2011-11-15 15:08 purgator.txt.thuf
-rw-rw-r-- 1 dna dna 256 2011-11-06 13:23 roгна.txt
-rw-rw-r-- 1 dna dna 512 2011-11-15 15:08 roгна.txt.thuf
drwxrwxr-x 2 dna dna 4096 2011-11-10 10:53 source
-rw-rw-r-- 1 dna dna 395 2011-11-10 11:03 sources.mk
-rw-rw-r-- 1 dna dna 852 2011-11-06 13:15 test2.txt
-rw-rw-r-- 1 dna dna 362 2011-11-15 15:08 test2.txt.thuf
-rw-rw-r-- 1 dna dna 33 2011-11-06 00:19 test.txt
-rw-rw-r-- 1 dna dna 266 2011-11-15 15:08 test.txt.thuf
dna@Mordor:~/workspace/huffman/Debug$
```

GRAZIE PER L'ATTENZIONE

