Project 0.1 External Hash

資工四 404410076 陳鐸元

Arguments

`-m` : memory limit in bytes
`-h` : number of entries in the hash table
`-s` : largest key size

Structure

- Input: read and parse all files
- hInsert: hash and store the record
  - After hitting memory limit, write to a temporary file

- Output: iterate through each list in each entry
- Sort: use HW1's record sort to sort it
  - `../HW1/record_sort -d $'\n'$'\n' -sd $'\n' -r -n ../HW0/tcount.out -o ../HW0/tout`

Benchmark

Computer Hardware

- CPU: i7 - 6700HQ (4C8T)
- Memory: 16GB (2x8GB)
- Storage: 512Gb M.2 PCIE Gen 3 x4 SSD (Plextor M8pegn)

Test Data

18 Gb Youtube data

In-memory only

- 51s

```
X130[danny@localhost HW0]$ 2018/11/26 20:25:51
~> time ./tcount_v1
0
file ettoday0.rec
1
file ettoday1.rec
2
file ettoday2.rec
3
file ettoday3.rec
4
file ettoday4.rec
5
file ettoday5.rec

real    0m51.859s
user    0m45.404s
sys     0m2.961s
```
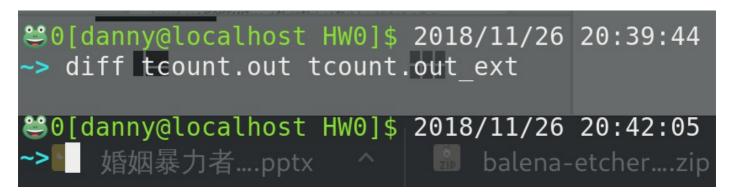
In-memory

- 4m7s

```
X130[danny@localhost HW0]$ 2018/11/24 23:38:24
~> ./tcount -m 3000000000 ettoday0.rec ettoday2.rec  ettoday4.rec ettoday1.rec  ettoday3.rec  ettoday5.re
c && ../HW1/record_sort -d $'\n'$'\n' -sd $'\n' -r -n ../HW0/tcount.out -o ../HW0/tout && less tout
0
file ettoday0.rec
1
file ettoday2.rec
2
file ettoday4.rec
3
file ettoday1.rec
4
file ettoday3.rec
5
file ettoday5.rec
start yo output
0
file: ../HW0/tcount.out
thread_token: 0
thread_token: 1
thread_token: 2
Finished reading file: ../HW0/tcount.out
Finished internal sort
start: 0, ccnt: 3, target: 3, final: ../HW0/tout
stream 0 Finished
stream 1 Finished
stream 2 Finished
Bye~
0[danny@localhost HW0]$ 2018/11/24 23:44:31
```

Limited-memory

```
😛0[danny@localhost HW0]$ 2018/11/24 23:44:31
~> ./tcount -m 1700000000 ettoday0.rec ettoday2.rec  ettoday4.rec ettoday1.rec  ettoday3.rec ettoday5.re
c && ../HW1/record_sort -d $'\n'$'\n' -sd $'\n' -r -n ../HW0/tcount.out -o ../HW0/tout && less tout
file ettoday0.rec
file ettoday2.rec
file ettoday4.rec
file ettoday1.rec
file ettoday3.rec
file ettoday5.rec
start yo output
0
file: ../HW0/tcount.out
thread_token: 0
thread_token: 1
thread_token: 2
Finished reading file: ../HW0/tcount.out
Finished internal sort
start: 0, ccnt: 3, target: 3, final: ../HW0/tout
stream 0 Finished
stream 1 Finished
stream 2 Finished
Bye~
😛0[danny@localhost HW0]$ 2018/11/24 23:50:16
~> ./tcou
```

Validation

```
😛0[danny@localhost HW0]$ 2018/11/26 20:39:44
~> diff tcount.out tcount.out_ext

😛0[danny@localhost HW0]$ 2018/11/26 20:42:05
~>  婚姻暴力者....pptx    ^       balena-etcher....zip
```

Code

https://github.com/Superdanby/Search-Engine/blob/master/Project%202/record_sort.c