

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

1 GB news records

In-memory only

- 51s

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.rec
c && ./HW1/record_sort -d '$\n'$'\n' -sd '$\n' -r -n ../HW0/tcount.out -o ../HW0/tout && less tout
0 total 11943276
file ettoday0.rec danny danny 596449659 Nov 26 15:45 Camd640S_R69-10895.B-Vanilla.7
1 file ettoday0.rec danny danny 8470 Nov 26 11:25 chromeify.sh
2 -rw-r--r-- 1 danny danny 2271232512 Sep 13 18:38 chromeos_10895.56.0_caroline_r
file ettoday4.rec channel.mp.bin
3 -rw-rw-r-- 1 danny danny 1146792061 Nov 26 17:23 chromeos_10895.56.0_caroline_r
file ettoday1.rec channel.mp.bin.zip
4 -rw-r--r-- 1 danny danny 1526743552 Sep 13 18:08 chromeos_10895.56.0_peppy_reco
file ettoday3.rec channel.mp.v3.bin
5 -rw-rw-r-- 1 danny danny 613955837 Nov 26 11:57 chromeos_10895.56.0_peppy_reco
file ettoday5.rec channel.mp.v3.bin.zip
start yo output danny danny 8751432192 Nov 26 17:42 chromiumos_image.img
0
file: ../HW0/tcount.out 2018/11/26 17:44:41
thread_token: 0
thread_token: 1
thread_token: 2
Finished reading file: ../HW0/tcount.out
Finished internal sort
start: 0, ccnt: 3, target: 3, final: ../HW0/tout
stream 0 Finished
stream 1 Finished
stream 2 Finished
Bye~
0[danny@localhost HW0]$ 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/tree/master/Project%204>