

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ



ΥΛΟΠΟΙΗΣΗ ΣΥΣΤΗΜΑΤΩΝ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ
ΑΣΚΗΣΗ 1/2

ΚΟΥΤΣΟΦΙΟΣ ΔΗΜΗΤΡΙΟΣ
ΠΟΛΥΧΡΟΝΑΚΗΣ ΣΑΒΒΑΣ

1115201200076
1115201200150

Στο παραδοτέο αρχείο περιλαμβάνονται τα παρακάτω αρχεία:

- main.c
- ht_functions.c
- ht_functions.h
- sht_functions.c
- sht_functions.h
- BF.h
- BF_64.a

το readme και ο φάκελος record_examples.

Οι εισαγωγές διαβάζονται από τα αρχεία που βρίσκονται στον φάκελο record_examples. Η εναλλαγή των αρχείων πρέπει να γίνει χειρονακτικά στη main, λόγω ενός προβλήματος που αντιμετωπίσαμε όταν αυτά δίνονταν από τη γραμμή εντολών.

Όσον αφορά τις δομές και τις συναρτήσεις, επιλέξαμε να μην κάνουμε καμία αλλαγή μετά από σχετικό νήμα που βρήκαμε στο eclass (παρ' όλο που στη δομή **SecondaryRecord** δε χρειαζόντουσαν όλες οι πληροφορίες του **Record**).

Όλες οι ζητούμενες συναρτήσεις υλοποιήθηκαν πλήρως και προστέθηκαν και οι 2 hash_functions.

Στον πρωτεύον κατακερματισμό, αφού το κλειδί ήταν int, επιλέξαμε να τα ταξινομήσουμε ομοιόμορφα, χρησιμοποιώντας την :

```
int hash_function(int id, int buckets){  
  
    return id % buckets;  
  
}
```

Στο δευτερεύον κατακερματισμό, χρησιμοποιήθηκε μία hash_function που βρέθηκε στο wikipedia και μετατρέπει το δοθέν string σε $0 < \text{int} < \text{buckets}$.

```
int shash_function(char* str, int buckets){  
  
    unsigned long hash = 5381;  
    int c;  
  
    while (c = *str++)  
        hash = ((hash << 5) + hash) + c;  
        /* hash * 33 + c */  
  
    return hash%buckets;  
  
}
```

Για τις διαγραφές παίρνουμε από τη γραμμή εντολών έναν αριθμό x και διαγράφουμε τις πρώτες x εγγραφές του αρχείου. Οι διαγραφές, όπως αναφέρθηκε σε μια συζήτηση στο eclass, γίνονται "εικονικά", δηλαδή απλά μαρκάρουμε την εγγραφή ως διεγραμμένη, θέτοντας το id της -1 και κάνοντας τα name, surname, address '\0'.

Στη main μας υπολογίζονται οι γραμμές του input file και μία-μία γραμμή γίνεται εισαγωγή στο πρωτεύον ευρετήριο. Στη συνέχεια, κατασκευάζεται το δευτερεύον ευρετήριο και ταυτόχρονα αντιγράφεται το πρωτεύον σε αυτό. Έπειτα, γίνονται άλλες n εισαγωγές (n = οι γραμμές-εγγραφές του input file) και στα δύο ευρετήρια.

Το πρόγραμμα μεταγλωττίζεται με την εντολή:

```
gcc -o main main.c ht_functions.c sht_functions.c BF_64.a -static
```

και εκτελείται με την εντολή:

```
./main <αριθμός_διαγραφών>
```

Υλοποιήθηκε σε Ubuntu 18.04 (VirtualBox).