

ΑΝΑΠΤΥΞΗ ΛΟΓΙΣΜΙΚΟΥ ΓΙΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

Μέλη Ομάδας:

Καρατσιώλης Ευστάθιος – 1115201200057

Κωστοπούλου Καλλιόπη 1115201200084

Λουκίδη – Παπανικολή Μαρία – 1115201200095

Για το πρώτο μέρος του project λήφθηκαν υπόψη μας όλες οι οδηγίες/συμβουλές που δόθηκαν στο piazza και τα φροντιστήρια.

- Το μέλος children της TrieNode δομής είναι ένας δείκτης σε πίνακα από TrieNodes προκειμένου να έχουμε καλύτερο «locality» για την cache.
- Το parsing ενός ngram γίνεται εξίσου με τη βοήθεια ενός πίνακα – vector για τον παραπάνω λόγο.
- Εν γένει, γίνεται εκτεταμένη χρήση πινάκων όπου αυτό είναι δυνατό για καλύτερο «locality».
- Αποφύγαμε τη χρήση αναδρομικών συναρτήσεων για καλύτερη απόδοση. Υλοποιήθηκε αναδρομικά μόνο η συνάρτηση που είναι υπεύθυνη για την αποδέσμευση της μνήμης κατά τον τερματισμό της εκτέλεσης του προγράμματος.
- Δεν έχουμε memory leaks.
- Υλοποιήσαμε λίγο κώδικα για “unit testing” όπως μας προτάθηκε και τον χρησιμοποιήσαμε κατά το debugging του προγράμματος μας.
- Η χρήση του πίνακα σαν δομή για τα «παιδιά» ενός TrieNode κάνει πιο δύσκολη τη διαδικασία της διαγραφής ή εισαγωγής ενός παιδιού (δεδομένου ότι αυτά θα πρέπει να είναι ταξινομημένα αλφαβητικά). Για να περιορίσουμε τα memory movements κατά τη διαγραφή/εισαγωγή κόμβων ακολουθούμε την εξής πολιτική:

ΑΝΑΠΤΥΞΗ ΛΟΓΙΣΜΙΚΟΥ ΓΙΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

Όταν ένας κόμβος διαγράφεται «χρωματίζεται» ως deleted (deleted μεταβλητή στη δομή TrieNode) χωρίς να αποδεσμεύεται η λέξη του, παρα μόνο ο χώρος που είναι δεσμευμένος για τα παιδιά του.

Έτσι δεν επηρεάζεται η εκτέλεση της binary search επάνω στον πίνακα των παιδιών και δεν απαιτείται άμεση μετακίνηση μέρους του πίνακα παιδιών για την κάλυψη της καινούριας θέσης που δημιουργείται.

Αντ' αυτού η trieDeleteNgram(..) συνάρτηση είναι υπεύθυνη για την εκκαθάριση του εκάστοτε πίνακα παιδιών, δηλαδή την αληθινή διαγραφή όλων των “deleted” κόμβων όταν το πλήθος αυτών γίνει αρκετά μεγάλο. Αυτό μειώνει σημαντικά το πλήθος των κλήσεων της memmove(..) κατά τη μέση περίπτωση, όπου συναντώνται πολλοί διαδοχικοί deleted κόμβοι.

- Υλοποιήθηκε makefile.
- Ο τρόπος εκτέλεσης του προγράμματος είναι όμοιος με αυτόν της εκφώνησης :

```
> ./ngrams -i <init_file> -q <query_file> ή
```

```
> ./ngrams -q <query_file> -i <init_file>
```

- Ενδεικτικοί Χρόνοι Εκτέλεσης στους υπολογιστές μας:
Χωρίς optimization flag κατά τη μεταγλώττιση:
Output στην κονσόλα: **1.429s**
Output με ανακατεύθυνση σε αρχείο: **0.971s**

Με optimization flag κατά τη μεταγλώττιση:

Output στην κονσόλα: **1.264s**

Output με ανακατεύθυνση σε αρχείο: **0.777s**

Insert

Η `trieInsertNgram(..)` συνάρτηση έχει υλοποιηθεί έτσι ώστε να λαμβάνει υπόψη της όλες τις υποπεριπτώσεις που προκύπτουν εξαιτίας της παραπάνω πολιτικής που υιοθετήσαμε. Εκμεταλλεύεται την περίπτωση που ένας κόμβος πρόκειται να μπει σε θέση που είναι «deleted», αναλαμβάνει να κάνει “realloc” όπου απαιτείται περισσότερος χώρος για τα παιδιά και μετά το πέρας της προσθέτει επιτυχώς ένα ngram στο trie, αφήνοντας πλήρως και ορθά ταξινομημένους τους πίνακες `TrieNodes` κάθε κόμβου.

Search

Η `trieSearchNgram(..)` συνάρτηση βρίσκει όλα τα ngrams που προκύπτουν από το ngram του `query`, όπως ζητείται στην εκφώνηση. Δεν εμφανίζονται διπλότυπα ngrams και η εκτύπωση των αποτελεσμάτων γίνεται με τη σειρά με την οποία αυτά συναντώνται στο δέντρο.

Έγινε αλλαγή του πρωτοτύπου της συνάρτησης σε σχέση με αυτού της εκφώνησης, μιας και τουλάχιστον για το πρώτο μέρος του project δεν απαιτείται η αποθήκευση των αποτελεσμάτων σε buffer αλλά αρκεί η απλή εκτύπωσή τους.

Delete

Σε αυτή τη συνάρτηση γίνεται η διαγραφή ενός n-gram (αν υπάρχει) από το trie. Σε κάθε επίπεδο του trie αν έχει διαγραφεί κανονικά κόμβος που είναι τελευταίος στον πίνακα των children nodes του προηγούμενου επιπέδου, τότε διαγράφονται όλα τα deleted nodes από το τέλος μέχρι το τελευταίο ενεργό και αυξομοιώνονται οι κατάλληλοι μετρητές. Επίσης, αν ακόμα και μετά από αυτές τις διαγραφές υπάρχει ένα ποσοστό από κόμβους μαρκαρισμένους ως deleted, τότε χρησιμοποιείται η `trieCompactSpace` για να μαζέψει όλους τους ενεργούς κόμβους στην αρχή του πίνακα και να αποδεσμεύσει οριστικά τους deleted.