

Προγραμματισμός Συστήματος

Project 1

Ονοματεπώνυμο: Βασιλική Γεωργάρα

A.M.: 1115201400026

Πώς δουλεύει το πρόγραμμα:

Αρχικά διαβάζει τα ορίσματα, είτε 5, είτε 3 (ανάλογα αν θα δοθεί k από τον χρήστη ή θα κρατηθεί το default που είναι 10).

Γίνεται έλεγχος για κάθε όρισμα και φυσικά μπορούν να δοθούν και με αντίθετη σειρά.

Στη συνέχεια δημιουργείται το Trie, το οποίο δημιουργεί την ChildList, PostingList και Tnode. Το Trie αρχικοποιείται δεσμεύοντας τη ρίζα, αρχικοποιώντας το Tnode της ρίζας και δημιουργώντας την ChildList, PostingList με prev=NULL και value = ' '.

Επίσης, δημιουργείται και το Map, όπου δημιουργεί δείκτες Document, start και end και τελικά αρχικοποιείται με count=totalwords=0 και start=end=NULL.

Υστερα, ανοίγω το file μέσα σε συνάρτηση, περνώντας το trie και το map: ανοίγω το file και το διαβάζω μέχρι να τελειώσει, καθαρίζοντας τις γραμμές της μεταβλητής line από σκουπίδια βάζοντας σε όλα 0. Έπειτα, καλείται η parseLineAndAddDocument μέσα στην οποία αποθηκεύεται το id από το κείμενο και γίνεται ο έλεγχος για το αν δόθηκαν σε σειρά και τερματίζει το πρόγραμμα σε αντίθετη περίπτωση.

Δημιουργώ τη δομή Document, περνώντας το id και δημιουργώντας τη λίστα λέξεων.

Για κάθε λέξη του document, βάζω τη λέξη στη λίστα λέξεων του document, καθώς και την προσθέτω στο trie. Για να την προσθέσω στο Trie, παίρνω έναν-έναν χαρακτήρα και αν υπάρχει ήδη στο trie, τότε επιστρέφω τον κόμβο για να συνεχίσω στον επόμενο χαρακτήρα, αλλιώς δημιουργώ τον κόμβο και τον προσθέτω. Όταν τελειώσω με τη λέξη,

δημιουργώ την PostingList και την ενημερώνω, ψάχνοντας αν η λέξη εισάγεται 1η φορά ή όχι, ώστε να ενημερώσω κατάλληλα και τη μεταβλητή frequency.

Για τις ανάγκες του score, υπάρχει η μεταβλητή totalwords που ενημερώνεται με κάθε πλήθος document words. Όταν τελειώσει η διαδικασία για το 1ο document τότε προστίθεται στο map. Τέλος, ξεκινάει η διαδικασία των εντολών. Πάντα θα φαίνεται η γραμμή εντολών “Minisearch>” όπου θα αναμένεται η πληκτρολόγηση των εντολών με “/” πριν από κάθε μία.

Αν δοθεί «/exit» θα τερματίσει η επανάληψη για το διάβασμα των εντολών και θα ακολουθήσουν τα απαραίτητα destructions.

Αν δοθεί σκέτο «/df» τότε θα καλεστεί η df function που καλεί τη συνάρτηση document_frequency, η οποία αν ο τρέχον κόμβος του Trie υπάρχει τότε καλεί την getWord για να επιστρέψει τη λέξη και την τυπώνει και ύστερα, ΑΝΑΔΡΟΜΙΚΑ, καλούνται όλοι οι κόμβοι παιδιά, ώστε τελικά να εκτυπωθούν οι λέξεις με τις συχνότητές τους.

Επίσης, έχει υλοποιηθεί και μια συνάρτηση strrev η οποία καλείται στην getWord ώστε να αντιστρέψει τη λέξη που διασχίζεται στο δέντρο με τη σωστή σειρά. Έτσι, τελειώνει ο έλεγχος για τις μονολεκτικές εντολές.

Όσον αφορά την «/df word» καλείται η df_word function που ψάχνει τη λέξη στο Trie.

Η «/tf id word» αποθηκεύει το id από την εντολή και τη λέξη και καλεί τη συνάρτηση tf η οποία κάνει 2 αναζητήσεις: πρώτα ψάχνει τη λέξη στο Trie και, αν δεν τη βρει, επιστρέφει 0, αν τη βρει όμως, κάνει αναζήτηση και στην PostingList το freq του συγκεκριμένου id.

Τέλος, αν έρθει η εντολή «/search query1 query2 ... queryn», από τις n queries που θα πληκτρολογήσει ο χρήστης, θα διαβαστούν μόνο οι πρώτες 10 και θα αποθηκευτούν σε μια λίστα με queries όπου θα φτιαχτεί και ένας πίνακας για τα scores που θα έχει μέγεθος όσα και τα documents. Ύστερα, καλείται η prepareQuery όπου υπολογίζεται το score, δηλαδή υπολογίζεται ο κάθε όρος και στέλνεται σε κατάλληλη συνάρτηση για να υπολογιστεί το τελικό κλάσμα. Και η printQuery που

εκτυπώνει τα k αποτελέσματα. Η εκτύπωση των αποτελεσμάτων των queries χρησιμοποιεί και την print από το αρχείο ListChar, όπου οι λέξεις που συμπεριλαμβάνονται στη search εντολή και υπάρχουν στα κείμενα, βρίσκονται μεταξύ των «|» για τις λέξεις που έχουν την μεταβλητή marked=true. Για κάθε νέα αναζήτηση όλα «καθαρίζουν» και όσες λέξεις ήταν μαρκαρισμένες, πλέον θα περάσουν από τη συνάρτηση unmark για να εκτελεστεί επιτυχώς η νέα αναζήτηση. Για την mergeSort που χρησιμοποιείται, ο κώδικας πάρθηκε από την ιστοσελίδα <https://www.geeksforgeeks.org/merge-sort/>.

Ακολουθούν ενδεικτικές εκτελέσεις παρακάτω.

Ενδεικτικές εκτελέσεις του προγράμματος:

Η μεταγλώττιση του προγράμματος γίνεται με το «make» και είναι separate compilation, υλοποίηση που επιταχύνει την μεταγλώττιση.

```
valia@valia-Lenovo-G50-70: ~/Desktop/project1
valia@valia-Lenovo-G50-70:~/Desktop/project1$ make
gcc -g ListChar.o ListPost.o Main.o Map.o MergeSort.o Query.o Trie.o -o Minisearch -lm
valia@valia-Lenovo-G50-70:~/Desktop/project1$ ./Minisearch -i Entertainment.txt -k 5
Minisearch> 
```

Ορισμένες εκτελέσεις για τις εντολές /df word και /tf id word για το αρχείο Entertainment.txt.

```
valia@valia-Lenovo-G50-70: ~/Desktop/project1
valia@valia-Lenovo-G50-70:~/Desktop/project1$ ./Minisearch -k 5 -i Entertainment.txt
Minisearch> /tf 6 actress
6 actress 3
Minisearch> /tf 300 zombie
300 zombie 0
Minisearch> /df beautiful
beautiful 4
Minisearch> /df syspro
syspro 0
Minisearch> 
```

```
valia@valia-Lenovo-G50-70: ~/Desktop/project1
valia@valia-Lenovo-G50-70:~/Desktop/project1$ ./Minisearch -i Entertainment.txt -k 5
Minisearch> /df actress
actress 68
Minisearch> /df man
man 42
Minisearch> /df the
the 386
Minisearch> /df zombie
zombie 1
Minisearch> 
```

Αν δοθεί λάθος αρχείο, πχ λόγω τυπογραφικού λάθους:

```
valia@valia-Lenovo-G50-70: ~/Desktop/project1
valia@valia-Lenovo-G50-70:~/Desktop/project1$ ./Minisearch -i Entertainment.txt -k 5
fopen: No such file or directory
valia@valia-Lenovo-G50-70:~/Desktop/project1$
```

Επίσης, αν δοθούν λάθος αριθμημένα αρχεία:

```
valia@valia-Lenovo-G50-70: ~/Desktop/project1
valia@valia-Lenovo-G50-70:~/Desktop/project1$ ./Minisearch -k 5 -i Sport.txt
Documents are not ordered. Exiting program...valia@valia-Lenovo-G50-70:~/Desktop/project1$
```

Μια ενδεικτική εκτέλεση του df:

```
valia@valia-Lenovo-G50-70:
Xhosa 2
X. 3
X, 2
Xfn, 1
Xzibit 1
XXL, 1
[Leigh] 1
[Depardieu] 1
[actor 1
[the 1
[Free 2
[older 1
[home], 1
[Appleton] 1
[Joss 1
[in 1
. 6
... 2
06/05 1
06/05, 1
06/05. 1
0130 2
0100 1
007 1
007, 2
0700, 1
03 1
0300 1
0400 1
0845 1
0800 1
0900 1
0200 1
, 2
zombie 1
zombie" 1
zombie." 1
zillion 2
zeitgeist. 1
zero 1
). 1
: 1
:- 1
:-) 1
/ 2
! 1
? 1
'Don't 1
Minisearch>
```


Μια ενδεικτική εκτέλεση του search:

valia@valia-Lenovo-G50-70: ~/Desktop/project1

Minisearch> /search hello from me

1.(349)[1.4837] India to deport Bollywood actress India has ordered the deportation of Iranian-born model and actress Negar Khan to Norway after saying she was working illegally on her visa. Khan has had raunchy roles in music videos and Bollywood films over the past two years. The distressed actress told media she was being driven straight to the airport after a routine appointment at the Bombay immigration office. Immigration officials said she had been warned about her visa last year. Khan told an Indian news channel on her mobile phone: "They are not even letting [me] go home to get my bags. I have no clothes or money on me. "They did not give [me] any notice. I don't know why they are taking [me] away like this. There is nothing wrong with my visa. It says I work here... They are not allowing [me] to even call a lawyer." Bipin Bihari, deputy police chief in Bombay, said Khan was being sent back to Norway, [from] which she holds a passport, on the first available flight. "She was on a visitor's visa but she has engaged in several commercial ventures," he said. "We issued notices last year, in view of which she had gone back to Norway, but she has returned again this year to work." One of Khan's biggest successes was the steamy Hindi song Chaddi Jawani (Rising Youth). Khan was involved in controversy when topless photos alleged to be of her appeared in a Norwegian magazine.

2.(253)[1.4359] Scissor Sisters triumph at Brits US band Scissor Sisters led the winners at the UK music industry's Brit Awards, walking off with three prizes. The flamboyant act scored a hat-trick in the international categories, winning the best group, best album and best newcomer awards. Glasgow group Franz Ferdinand won two prizes, as did Keane and Joss Stone, who was voted best urban act by digital TV viewers. Robbie Williams' Angels was named the best song of the past 25 years. Scissor Sisters frontwoman Ana Matronic collected the best international album prize [from] singer Siouxsie Sioux. She told the audience: "If you told us a year ago we would be getting these awards today we would have called you crazy. You guys made our dream come true." The band - whose self-titled LP was 2004's biggest-selling album - thanked "all the members of the sisterhood", adding: "We wouldn't be here without you." The US band, who opened the show with Take Your Mama, won the best international act and newcomer awards, as well as best international album. Franz Ferdinand, who were shortlisted in five categories, won best rock act and best British group, an award they dedicated to late DJ John Peel. But they missed out on best British live act, which went to Muse. Keane won best British album and breakthrough act. Will Young won the best single prize for Your Game. McFly won the best pop act prize, and Gwen Stefani picked up the best international female artist award. Eminem won the male prize. Best British male artist winner Mike Skinner - aka The Streets - does not usually attend award ceremonies, but at the Birmingham hip-hop artist performed his hit Dry Your Eyes at the ceremony. However, he did not collect his prize. A bandmate informed the crowd Skinner was "in the toilet". After beating Amy Winehouse, Jamelia, Natasha Bedingfield and PJ Harvey to the best British female prize, Joss Stone said: "I don't know what to say. I don't like doing this at all. I'd like to thank my family for being really supportive and everybody that made my record with me." "I don't even know what to do right now. Thank you all you guys for voting for me, I feel sick right now." Viewers of digital music TV channel MTV Base voted Stone the winner in the best urban act category. Little Britain comedy duo Matt Lucas and David Walliams presented the best song prize to Robbie Williams dressed as his former Take That colleague Gary Barlow and Howard Donald, leading him to quip he was "always the talented man of the band". Williams' track beat songs by Will Young, Queen, Kate Bush and Joy Division in a vote by BBC Radio 2 listeners to mark 25 years of the UK music industry ceremony. It is his 15th Brit award, having already received 10 solo awards and four with Take That. He told the audience: "I'm just amazed that my career keeps going." Keane frontman Tom Chaplin thanked fans for enduring "rubbish gigs" after they won the British breakthrough act prize. He added: "A lot of people don't think it's cool that we've had the guts to be ourselves but it's a vital part of who we are as a band and receiving this is recognition of that." Natasha Bedingfield - in the running for best British female and best pop act - performed with her brother Daniel for the first time at Wednesday's event. The chart-topping siblings duetted on the Chaka Khan hit Ain't Nobody. Meanwhile, Joss Stone performed Right To Be Wrong backed by a gospel choir, while Lemar and Jamelia performed the Robert Palmer track Addicted To Love. Bob Geldof won a prize for his outstanding contribution to music. Of the 15 Brit awards for achievements in 2004, 10 were won by artists tipped in the BBC News website's Sound of 2004 list of artists to watch, published at the start of last year. Scissor Sisters, Franz Ferdinand, Keane, Joss Stone and McFly were all in the Sound of 2004 top 10. The other five Brits winners were already established before Sound of 2004 was compiled. The ceremony will be televised on ITV1 on Thursday. I'm speechless. Best song of the last 25 years? Yeah right. I very much doubt that 'Angels' was even the best song of the week that it came out. Like every track Robbie has released as a single, it's a blatant but poor facsimile of something that someone else has done better before. Give us a break....!!! Best song in 25 years, you must be joking. Its good if you like that sort of thing, but really! Listened to Angels on Radio 1 this morning when I was driving into work. Had not heard it for a while. I love Robs voice, the lyrics and tune. Perfection! As usual, the public have short memories when it comes to voting for "the greatest". There must be more than a dozen songs in the last 25 years that deserve this award more. It's not exactly groundbreaking. Presumably, the age range that could be bothered to vote is pretty low... I'm actually embarrassed to be British if that is the best song we have produced in the last 25 years!! What about The Specials - Ghost Town, The Buzzcocks - Ever Fallen in Love With Someone... Happy Mondays - Kinky Afro, McAlmont & Butler - Yes, Joy Division - Love Will Tear Us Apart... Angels is middle-of-the-road rubbish. Angels is a awful piece of sentimental claptrap. It's musically and lyrically inept; and fantastically overrated, a bit like Mr Williams himself. This result isn't very surprising though, The Brits has a long history of celebrating rubbish music! Best of the last 25 years? Maybe. Cunning to make the timescale not include S'tairway to Heaven or Bohemian Rhapsody, but it does kind of make it a bit of a hollow award really. Not much competition in the last 25 years after all. It's alright for a pop song - but the best song of the last 25 years??? There is no way on earth that song should have been voted the best of the last 25 years....it's a travesty. Rubbish! Who voted it for it to be included in any list? I am a regular listener to Radio 2 but I don't recall the invitation to vote for this bland, slushy rubbish which might appeal to the masses who wouldn't know a good song if it jumped up and bit them on the nose but is certainly NOT the best song of the last 25 years. How depressing and just when we thought manufactured 'pop' was on the way out - where on earth did this dreadful list appear from? While I am biased i

valia@valia-Lenovo-G50-70: ~/Desktop/project1

3.(201)[2.5993] Triple triumph for Little Britain Top TV cult hit Little Britain has triumphed over The Office at the 15th annual British Comedy Awards. The BBC show beat The Office Christmas Special in the best TV comedy category and took the people's choice award. Matt Lucas and David Walliams also jointly won best TV comedy actor over The Office creator Ricky Gervais, who won a special award for writing. It was also a good night for Ant and Dec who clinched two awards, including best comedy entertainment personality. The Geordie duo also secured best comedy entertainment programme for Saturday Night Takeaway, beating I'm A Celebrity... Get Me Out Of Here!, which they also host, and the BBC's Have I Got News For You. The ceremony was presented by Jonathan Ross from ITV's London Studios headquarters on the capital's South Bank. Little Britain began life as a radio series on Radio 4, with Walliams and Lucas playing a range of misfit characters from an alternative Britain. It was later commissioned for BBC Three, where it proved a huge success, with the second series also shown on BBC One. After receiving his award for best TV comedy actor with Walliams, Lucas said: "It's a great honour to be nominated alongside Ricky Gervais and Martin Clunes." The show also won the people's choice award, which is the only gong on the night voted for by viewers. Speaking after receiving the award for best comedy, Walliams said: "It was a real shock but I don't think we deserved to beat The Office which is a great show and more than just a comedy. It's a beautiful piece of work." Gervais and The Office co-writer Stephen Merchant received a special award for writing, presented by Noel Gallagher and Tom Stoppard. Gervais said: "The show has won an awful lot of awards in the last couple of years, too much. It's got embarrassing to be honest." Other winners at the awards included Caroline Quentin, who took home best TV comedy actress for Life Begins and Von Trapped. Catherine Tate, from The Catherine Tate Show, was also nominated for the same award, but she did walk away with best comedy newcomer, over James McAvoy from Shameless (Channel 4) and Julian Barratt and Noel Fielding from The Mighty Boosh (BBC Three). Nighty Night was named Britain's best new TV comedy over ITV1 shows Doc Martin and Life Begins. But Doc Martin won best TV comedy drama, over Jonathan Creek (BBC One) and Shameless. Winner in the best international comedy category was The Simpsons, beating Sex and the City and Curb Your Enthusiasm. The prize was accepted by Matt Groening, creator of The Simpsons, who also received an outstanding contribution award from scientist Stephen Hawking, who has been featured in the series. School of Rock was voted best comedy film beating Shaun of the Dead and Shrek 2. Matt Lucas & David Walliams (Little Britain, BBC Three) Caroline Quentin (Life Begins/ Von Trapped, ITV1) Ant & Dec (Ant & Dec's Saturday Night Takeaway/ I'm A Celebrity, Get Me Out Of Here!, ITV1) Catherine Tate (The Catherine Tate Show, BBC Two) Nighty Night (Baby Cool Productions, BBC Three) Little Britain (BBC Three) Doc Martin (Buffalo Productions, ITV1) Ant & Dec's Saturday Night Takeaway (Granada, ITV1) The Simpsons (Twentieth Century Fox, BBC Two/Sky One/Channel 4) School Of Rock

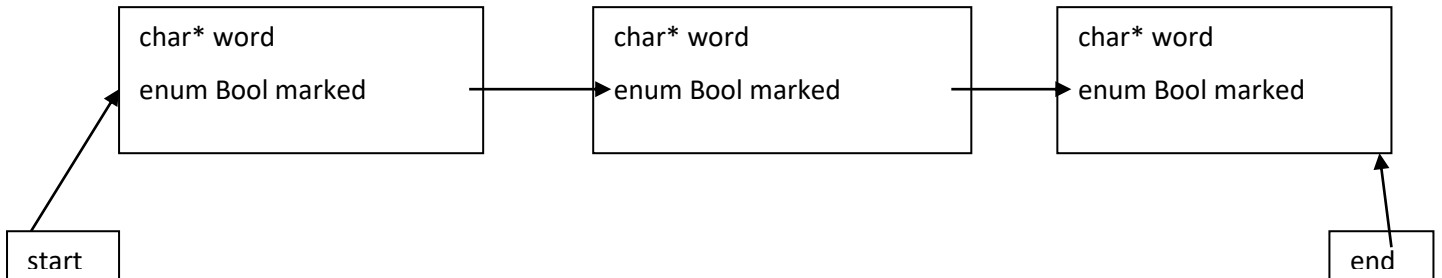
4.(143)[1.9306] Blair buys copies of new Band Aid Prime Minister Tony Blair purchased two copies of the charity single Band Aid 20 in Edinburgh on Friday. Staff were surprised when the Prime Minister walked into HMV at 0900 GMT, accompanied by aides and local police. "When Mr Blair came in unannounced, we were all pretty gobsmacked," said HMV manager Clive Smith. "Our customer helper approached him... it was only then we realised he wanted to buy copies of the Band Aid single, rather than the latest Eminem album." Predicted chart-topper Do They Know It's Christmas? is expected to sell at least 300,000 copies by the time the new chart is announced on Sunday. However, the new version of the 1984 single is not going to be released in the US, despite being sold in many countries around the world. US record shops are stocking an import version of Do They Know It's Christmas, which is said to be selling very well in Los Angeles and New York. The original track was released in the US, and reached number 13 in the singles chart. British stars who appear on the current recording, such as Dido and Coldplay's Chris Martin, are well-known to music fans across the Atlantic, along with U2 frontman Bono. Record company Universal is responsible for the global distribution of the single, which will be available across Europe, Asia, South America and Canada. But music fans in the US are still able to access the song and download it on Band Aid 20's official website. In 1985, a group of high-profile American stars known as USA For Africa came together to record their own fund-raising single, We Are The World. The song was written by Lionel Richie and Michael Jackson, with Quincy Jones as producer. It topped the US charts for three weeks and went on to win Grammy awards for best record and song. Dionne Warwick, Diana Ross and Tina Turner were among the line-up of performers. It is predicted that the Band Aid 20 song will sell 300,000 copies in the UK by the time the new chart is announced on Sunday. The record is also tipped to become this year's Christmas number one, as the original version did in 1984. Proceeds from the sales are going towards relief for the Darfur region of Sudan and to combat HIV and Aids across Africa.

5.(159)[1.9306] Band Aid 20 single storms to No 1 The new version of the Band Aid song Do They Know It's Christmas? has gone straight in at number one in the UK singles chart. The charity record is also tipped to be this year's Christmas number one. It features vocals from the likes of Chris Martin, Dido, Robbie Williams and the Sugababes. The original version - which was the Christmas number one in 1984 - sold 750,000 copies in its first week and 3.5 million in total. It was released in the US, and reached number 13 in the singles chart. However, Band Aid 20 is not going to be released in the US, despite being sold in many countries around the world. US record shops are stocking an import version of Do They Know It's Christmas, which is said to be selling very well in Los Angeles and New York. British stars who appear on the current recording, such as Dido and Coldplay's Chris Martin, are well-known to music fans across the Atlantic, along with U2 frontman Bono. Record company Universal is responsible for the global distribution of the single, which will be available across Europe, Asia, South America and Canada. But music fans in the US are still able to access the song and download it on Band Aid 20's official website. In 1985, a group of high-profile American stars known as USA For Africa came together to record their own fund-raising single, We Are The World. The song was written by Lionel Richie and Michael Jackson, with Quincy Jones as producer. It topped the US charts for three weeks and went on to win Grammy awards for best record and song. Dionne Warwick, Diana Ross and Tina Turner were among the line-up of performers. Proceeds from the sales of the Band Aid 20 single will go towards aid relief in Africa, in countries such as Ethiopia and Sudan. The money raised will also be used to help combat HIV and Aids across the continent. The Band Aid Trust which was set up 20 years ago, when the original single was released, handed out \$144m (£75m) to famine relief projects across Africa between January 1985 and November 2004.

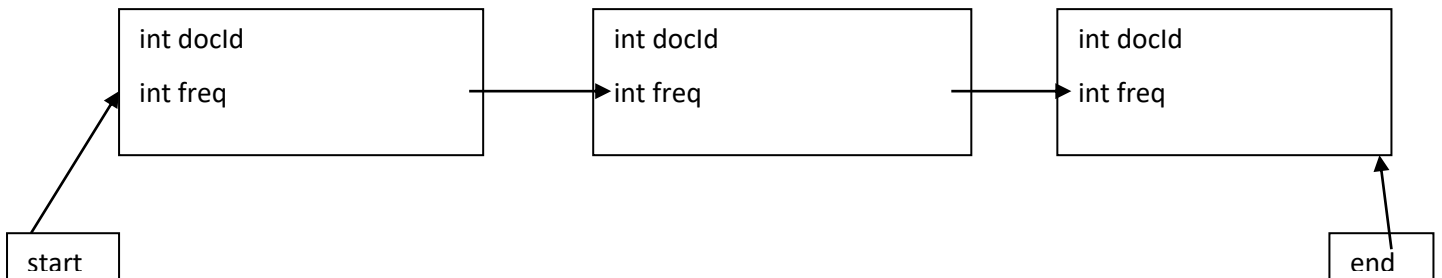
Minisearch>

Οι δομές του προγράμματος:

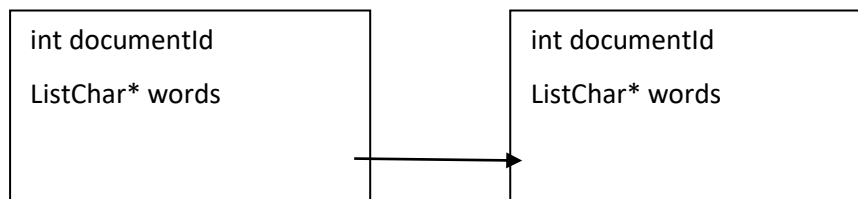
ListChar: (λίστα λέξεων)



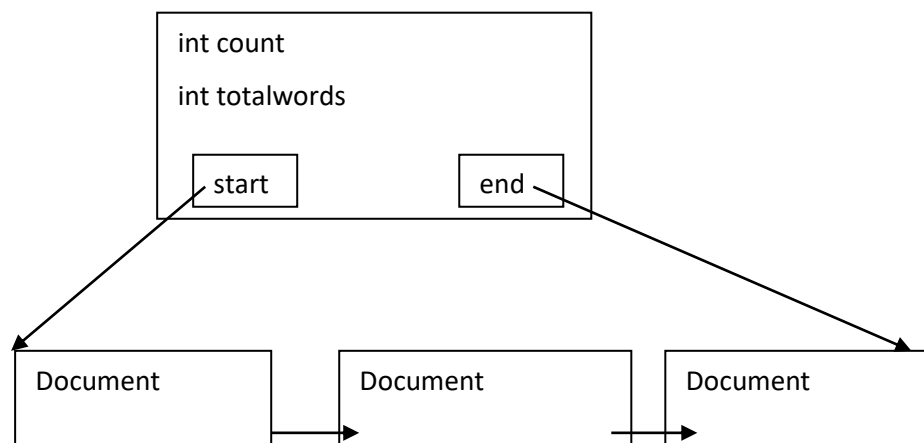
ListPost: (posting list)



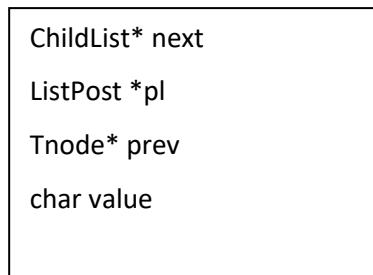
Document:



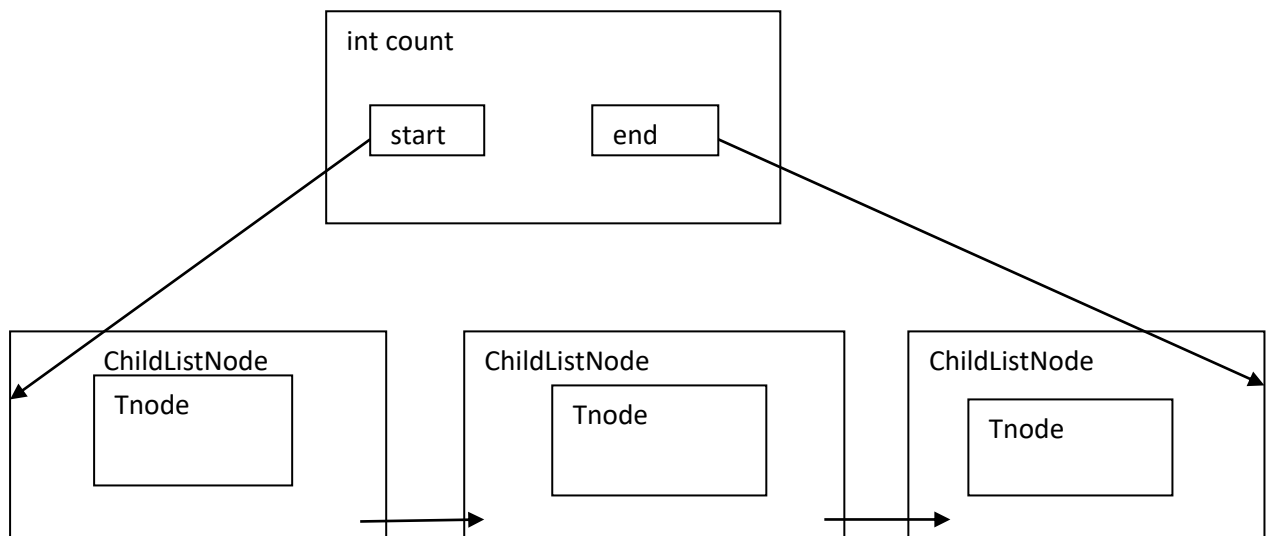
Map



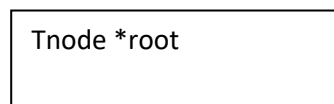
Tnode:



ChildList:



Trie:



Query:

