Προγραμματισμός Συστήματος
Εργασία 3

Καλογερόπουλος Ιωάννης Α.Μ.:1115201500057

Σχόλια-Παραδοχές

α)Web site creator

-Κατά την εισαγωγή των link σε κάθε σελίδα ο σύνδεσμος που χρησιμοποιείται για τα εσωτερικά links είναι της μορφής page_....html,ενώ για τα εξωτερικά ../site_.../page_....html και το όνομα του συνδέσμου για κάθε περίπτωση είναι όλο το μονοπάτι από τον κατάλογο root_dir

-Αρχικά ο πίνακας array περιέχει όλα τα μονοπάτια των σελίδων της μορφής root_dir/site_.../page_....html και στον πίνακα incoming_links αποθηκεύονται όλα τα links που χρησιμοποιούνται στην δημιουργία κάθε σελίδας.
Στο τέλος τα στοιχεία του array ταξινομούνται αλφαριθμητικά και το ίδιο γίνεται στα στοιχεία του incoming_links αφού έχουν αφαιρεθεί τα διπλότυπα.
Αν αυτοί οι 2 πίνακες είναι ίδιοι τότε εκτυπώνεται πως όλες οι σελίδες έχουν έναν τουλάχιστον εισερχόμενο σύνδεσμο

β)Web server

-Με την συνάρτηση poll() ελέγχω ποιός file descriptor είναι διαθέσιμος
Μου φάνηκε χρήσμιο αυτό το link:https://www.ibm.com/support/knowledgecenter/en/ssw_i5_54/rzab6/poll.htm
Αν γίνει σύνδεση για http request τοποθετείται ο fd στον buffer(λίστα χωρίς max μέγεθος) κσι στέλνεται σήμα σε ένα thread για να “ξυπνήσει” και να επικοινωνήσει με μέσω socket.(Για κάθε σύνδεση δημιουργείται μία σύνδεση και άρα ένας fd)
Αν ο fd για εντολή command είναι έτοιμος,τότε μία φορά δημιουργείται ένας fd για εισαγωγή εντολών από το telnet,όπου όταν αυτός είναι έτοιμος ο χρήστης θα μπορεί να δώσει εντολές

-Η δομή thread_arg χρησιμεύει να περαστού πολλές παράμετροι στα threads.Ο ακέραιος shutdown σε αυτήν γίνεται 1 όταν δωθεί SHUTDOWN.

-Το κάθε thread αν ο buffer είναι άδειος περιμένει σήμα για να συνεχίσει.Έτσι ξαναελέγχει αν είναι άδειος η όχι για να κάνει pop το πρώτο στοιχείο.Ο τελευταίος έλεγχος ώστε με το pthread_cond_broadcast(&cv_nonempty); όταν δωθεί SHUTDOWN τα threads που έχουν κολλήσει στο pthread_cond_wait να τερματίσουν χωρίς να επέμβουν στον buffer

-Σε κάθε εισαγωγή fd στην ουρά στέλνεται σήμα σε ένα thread να ξυπνήσει.

Σε κάθε pop και push στην ουρά γίνεται mutex lock,unlock
-Στα threads αν flag==1 σημείναι ότι έχει γίνει pop από την ουρά και καλείται η συνάρτηση manage_request για επικοινωνία με thread από τον crawler

γ)Web crawler

-Η λογική για το ποιός fd είναι διαθέσιμος είναι παρόμοια με τον server μόνο που εδώ δεν δεχόμαστε request

-Όπως και στον server τα threads λειτουργούν παρόμοια μόνο που σήμα σε thread δεν στέλνεται από την main συνάρτηση,αλλά από κάθε thread σε ένα άλλο όταν βάλει στοιχείο στην ουρά

-Η συνάρτηση manage_reply διαχειρίζεται κατάλληλα το http response από τον server δημιουργώντας αντίγραφο του φακέλου root_dir αν όλα τα αρχεία του είναι αναγνώσιμα

-Η ουρά history χρησιμοποιείται ως ιστορικό των σελίδων που έχουν ζητηθεί

Οργάνωση αρχείων:
Ο server και ο crawler έχει ο καθένας τα δικά του αρχεία με τις συναρτήσεις και τους ορισμούς τους,όμως μοιράζονται τα αρχεία common.c και common.h.Κι οι 2 χρησιμοπουούν ουρά οπότε η υπάρχει μια γενική λίστα με ένα Union ως στοιχείο το οποίο αποτελείται από έναν int και ένα char * ώστε κάθε φορά να χρησιμοποιείται αυτό που χρησιμεύει ως στοιχείο λίστας

Γενικά:

-Στον crawler αν ο φάκελος save_dir υπάρχει,ο χρήστης πρέπει να τον διαγράψει για να τρέξει το πρόγραμμα

-Requests που στέλνονται από τον browser άλλες φορές δουλεύουν και άλλες όχι.Αν το αρχείο υπάρχει δεν στέλνεται όλο το περιεχόμενό του αλλά ένα μεγάλο μέρος του

-Για να δουλεύει server και crawler πρέπει το rood_dir να έχει / στο τέλος ενώ το save_dir να μην έχει.Επίσης το root_dir πρέπαι να είναι στο ίδιο directory με τα εκτελέσιμα αρχεία

Περισσότερα σχόλια μέσα στα αρχεία