package crawl

import (

"regexp"

“ioutil”

“io”

“log”

“net/http”

)

func ParseUrlListAndContent(contents []byte, linkLimit int) error {

re := regexp.MustCompile(“(?s)<a[ t]+.\*?href=”(http.\*?)”.\*?>.\*?</a>”)

newUrls := re.FindAllSubmatch(contents,-1) //[][][]byte

file, err := os.Create(“link\_output.txt”)

if err != nil {

return err

}

defer file.Close()

// header & body

var header string

var write []byte

for \_, v := range newUrls {

log.Printf(“Fetching %s”, v[1])

\_, err := io.WriteString(file, v[1])

if err != nil {

return err

}

// Make HTTP GET request

resp, err1 := http.Get(v[1])

defer resp.Body.Close()

if err1 != nil {

return err1

} else {

//body

body, err := ioutil.ReadAll(resp.Body)

for h, v := range resp.Header {

for \_, v := range v {

header += fmt.Sprintf("%s %s \n", h, v)

}

}

// Append all to one slice

write = append(write, []byte(header)...)

write = append(write, body...)

// Write it to a file

err = ioutil.WriteFile("content\_output.txt", write, 0644)

if err != nil {

return err

}

)

linkLimit--

if linkLimit == 0 {

break

}

}

return nil

}