

Net-centric Programming

Lab: Web Crawler

I. Objective:

- Implement a crawler to get data from webpages
- The following code connect to webtoons.com to get all the links on frontpage

```
package main

import (
    "bytes"
    "fmt"
    "io/ioutil"
    "net/http"
    "golang.org/x/net/html"
)

func main() {
    // Make a GET request to the WEBTOON homepage
    resp, err := http.Get("https://www.webtoons.com/en/")
    if err != nil {
        fmt.Println("Error fetching WEBTOON homepage:", err)
        return
    }
    defer resp.Body.Close()

    // Read the response body
    body, err := ioutil.ReadAll(resp.Body)
    if err != nil {
        fmt.Println("Error reading response body:", err)
        return
    }

    // Parse the HTML content
```

```
doc, err := html.Parse(bytes.NewReader(body))

if err != nil {
    fmt.Println("Error parsing HTML:", err)
    return
}

// Find all the links on the page
links := findAllLinks(doc)

// Print the links
for _, link := range links {
    fmt.Println(link)
}

}

func findAllLinks(n *html.Node) []string {
    var links []string

    var walk func(n *html.Node)
    walk = func(n *html.Node) {
        if n.Type == html.ElementNode && n.Data == "a" {
            for _, attr := range n.Attr {
                if attr.Key == "href" {
                    fmt.Println(attr.Val)
                }
            }
        }
        for c := n.FirstChild; c != nil; c = c.NextSibling {
            walk(c)
        }
    }

    walk(n)
}
```

```
return links
```

```
}
```

II. Questions:

Implement a web crawler to get at least 10 manga titles from each genres of at least 10 genres on webtoons.com (e.g. Drama, Fantasy,)

The information should be stored in a JSON file

The end.