

Scraping Web Sites in Go

Reading in a webpage

```
package main

import (
    "fmt"
    "io/ioutil"
    "log"
    "net/http"
)

func main() {

    resp, err := http.Get("https://www.w3schools.com/html/html_tables.asp")
    if err != nil {
        log.Fatal(err)
    }

    defer resp.Body.Close()

    // Read the response body and convert it to a string
    body, err := ioutil.ReadAll(resp.Body)
    if err != nil {
        log.Fatal(err)
    }

    html := string(body)
    fmt.Println(html)

}
```

Install goquery

```
go get github.com/PuerkitoBio/goquery
```

Now Read in the response body into goquery

```
doc, err := goquery.NewDocumentFromReader(resp.Body)

if err != nil {

    log.Fatal(err)

}
```

Using Find:

```

doc.Find("table").Each(func(i int, sel * goquery.Selection) {

    // For sake of simplicity taking the first table of the page
    if i == 0 {

        // Looping through headers
        headers: = sel.Find("th").Each(func(_ int, sel * goquery.Selection) {
            if sel != nil {
                fmt.Print(sel.Text())
                fmt.Print(" ")
            }
        })
        fmt.Println()

        // Looping through cells
        sel.Find("td").Each(func(index int, sel * goquery.Selection) {
            if sel != nil {
                fmt.Print(sel.Text())
                fmt.Print(" ")
            }

            // Printing columns nicely
            if (index + 1) % headers.Size() == 0 {
                fmt.Println()
            }
        })
    }
})

```

References

Reference	URL
Find out how to scrape HTML tables with Golang	https://www.webscrapingapi.com/find-out-how-to-scrape-html-tables-with-golang
goquery	https://github.com/PuerkitoBio/goquery
GoQuery Docs	https://pkg.go.dev/github.com/PuerkitoBio/goquery