

CS17B010

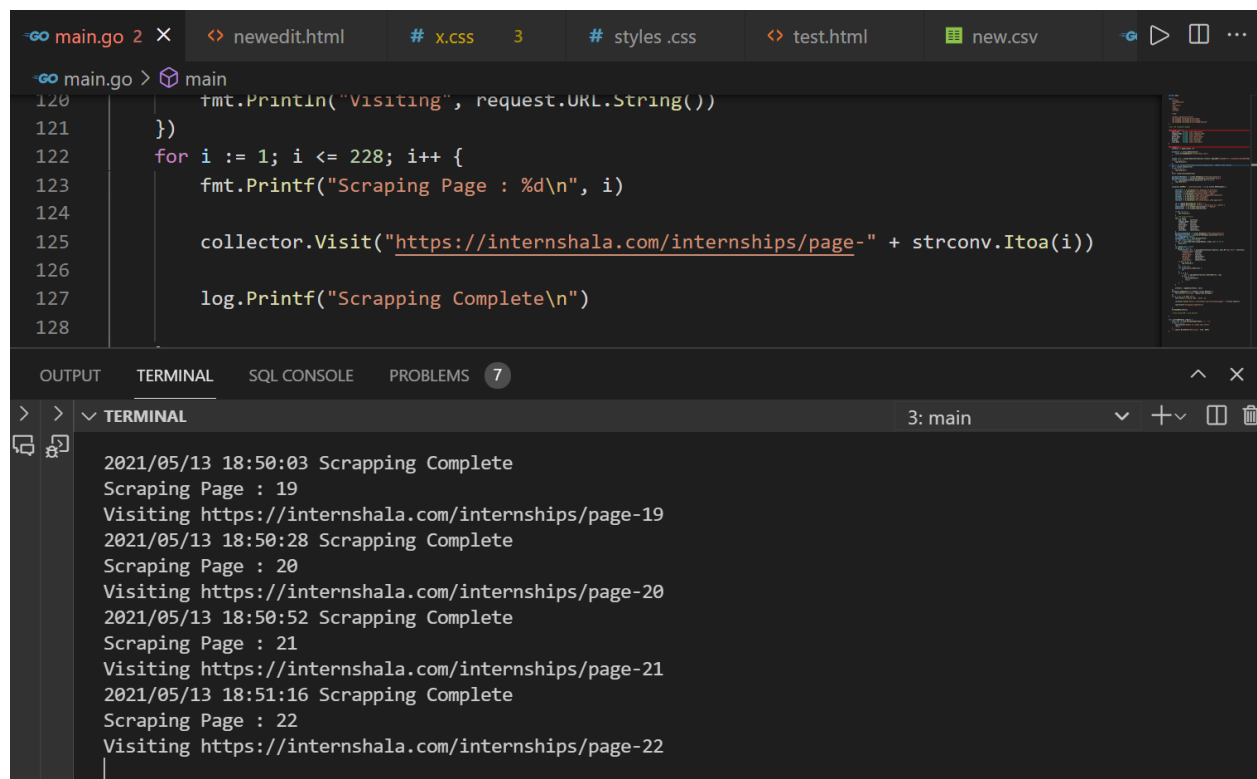
REPORT ON ASSIGNMENT-4

In this assignment I selected **INTERNSHALA** website for web scraping major problem with this website to get specific internship we have to iterate over multiple web pages it may have sometimes more than 200 So to reduce the work of the user we will scrap the all pages and store in database

(scrapping code is present here and stored the data in database)

(https://github.com/CS17B/distrubuted_Systems_Assignment_4/blob/main/main.go)

Terminal displaying scrapping of each time



```
main.go 2 X  <> newedit.html  # x.css 3  # styles .css  <> test.html  new.csv  <> <> ...

main.go > main
120     fmt.Println("Visiting", request.URL.String())
121     })
122     for i := 1; i <= 228; i++ {
123         fmt.Printf("Scrapping Page : %d\n", i)
124
125         collector.Visit("https://internshala.com/internships/page-" + strconv.Itoa(i))
126
127         log.Printf("Scrapping Complete\n")
128     }

OUTPUT  TERMINAL  SQL CONSOLE  PROBLEMS 7  ^  X
> > 3: main  v  +v  <>  <>
2021/05/13 18:50:03 Scrapping Complete
Scrapping Page : 19
Visiting https://internshala.com/internships/page-19
2021/05/13 18:50:28 Scrapping Complete
Scrapping Page : 20
Visiting https://internshala.com/internships/page-20
2021/05/13 18:50:52 Scrapping Complete
Scrapping Page : 21
Visiting https://internshala.com/internships/page-21
2021/05/13 18:51:16 Scrapping Complete
Scrapping Page : 22
Visiting https://internshala.com/internships/page-22
```

And Data stored in database:

In below image format how the data is stored in database using the table schema as

Job_Title	Company_Name	Location	Start_Date	Duration	Stipend	Last_Name
-----------	--------------	----------	------------	----------	---------	-----------

After entering the required details like category like(Job_Title,Company_Name,Location,Start_Date,Duration,Stipend,Last_Date) And details of category For example for category of Company the details of the category be like company names (Urban Company,Nykaa etc) After entering the details the corresponding data is searching in the database and corresponding data is fetched and displayed on web page as below

	JOB TITLE	COMPANY NAME	LOCATION	START DATE	STIPEND	DURATION	LAST DATE
0:	Operations	Urban Company	Bangalore	Immediately	8000 /month	2 Months	24 May' 21
1:	Operations	Urban Company	Bangalore	Immediately	7000 /month	3 Months	19 May' 21
2:	Marketing	Nykaa	Bangalore		3500-5000 /month	5 Months	19 May' 21
3:	Web Development	Ericsson India Global Services Private Limited	Bangalore	Immediately	15000 /month	6 Months	17 May' 21
4:	Market Research	Shades Of Spring	Bangalore	Immediately	20000 /month	6 Months	28 May' 21
5:	Research & Development	Mobisy Technologies Private Limited	Bangalore	Immediately	12000-15000 /month	6 Months	28 May' 21
6:	Digital Marketing	EthnicArk	Bangalore	Immediately	12000 /month	1 Months	28 May' 21
7:	Sales Coordination (Inside Sales)	VDISTIE	Bangalore	Immediately	10000 /month	6 Months	28 May' 21
8:	Mechanical Design Engineering	Euphotic Labs Private Limited	Bangalore	Immediately	10000 /month	3 Months	28 May' 21
9:	Business Development (Sales)	TheHouseMonk	Bangalore	Immediately	12000 /month	3 Months	28 May' 21
10:	Data Science - NLP	PiChain Innovations Private Limited	Bangalore	Immediately	15000 /month	6 Months	28 May' 21
11:	Data Science (Computer Vision)	PiChain Innovations Private Limited	Bangalore	Immediately	15000 /month	6 Months	28 May' 21
12:	Python Development	PiChain Innovations Private Limited	Bangalore	Immediately	15000 /month	6 Months	28 May' 21
13:	YouTube Marketing	TopRankers	Bangalore	Immediately	6000 lump sum	1 Months	27 May' 21
14:	Software Development	TechMeridian Solutions Private Limited	Bangalore	Immediately	10000 /month	6 Months	27 May' 21
15:	Computer Vision	Modaka Technologies Private Limited	Bangalore	Immediately	3000-5000 /month	4 Months	27 May' 21
16:	Graphic Design	Tenshi Life Care	Bangalore	Immediately	10000 /month	3 Months	27 May' 21
17:	Human Resources (HR)	Duru Cooperation Private Limited	Bangalore	Immediately	25000 /month	6 Months	27 May' 21
18:	Business Development (Sales)	Greendzine Technologies	Bangalore	Immediately	18000 /month + Incentives	6 Months	27 May' 21
19:	Social Media Marketing	Greendzine Technologies	Bangalore	Immediately	15000 /month + Incentives	6 Months	27 May' 21
20:	Social Media Marketing	TrustLogics	Bangalore	Immediately	15000 /month	6 Months	27 May' 21
21:	Operations	Matter Gifting	Bangalore	Immediately	2000-6000 /month	1 Months	27 May' 21
22:	Marketing	DhiSigma Systems Private Limited	Bangalore	Immediately	5000-10000 /month	1 Months	27 May' 21
23:	Business Development (Sales)	Fidelitus Corporation Property Services Private Limited	Bangalore	Immediately	10000 /month	1 Months	27 May' 21
24:	Business Development (Sales)	Taught	Bangalore	Immediately	8000 /month + Incentives	3 Months	27 May' 21

This output corresponding to the input For INTERNSHIP CATEGORY : Location DETAILS OF CATEGORY:Bangalore

Currently 50 pages of the internshala are scrapped and stored in database. And by changing the count of no of pages any no of pages can be scrapped and stored And can be retrieved easily