21 Feb ASS

March 16, 2023

[ ]:
```
Q1. What is Web Scraping? Why is it Used? Give three areas where Web Scraping␣
 ↪is used to get data.
```

[ ]:
```
ANS -
```

[ ]:
```
Web scraping is the process of using bots to extract content and data from a␣
 ↪website.
unlike screen scraping,which only copies pixies displayed onscreen, web␣
 ↪scraping extracts underlying HTML code and ,
with it data stored in a database. The scraper can then replicate entire␣
 ↪website content elsewhere.
```

[ ]:
```
web scraping is used in a variety of digital businesses that rely on data␣
 ↪harvesting.
```

[ ]:
```
. Search engine bots crawling a site, analyzing its content and then ranking it.
```

[ ]:
```
. Price comparison sites deploying bots to auto-fetch prices and product␣
 ↪descriptions
 for allied seller websites.
```

[ ]:
```
. Market research companies using scrapers to pull data from forums and social␣
 ↪media
 (e.g., for sentiment analysis).
```

[ ]:
```
Web scraping is also used for illegal purpose, including the undercutting of␣
 ↪prices and the theft of copyrighted content
An online entity target by a scraper can suffer severe finincial by a scraper␣
 ↪can suffer severe finincal losses,especially if its a
business strongly relying on competitive pricing models or deals in content␣
 ↪distribution.
```

[ ]:
```

```

[ ]:
```

```

[ ]:
```
Q2. What are the different methods used for Web Scraping?
```

[ ]:
```
ANS -
```

[ ]:
```
Web scraping is an automated method of obtaining large aounts of data from␣
 ↪websites.most of this data is unstructured data in HTML format,
which is then converted in to structured data in a spreadsheet or database so␣
 ↪that it can be used in various applications,there are many
ways to perform web scrapping to get data from websites.
These include using online services, special APIs, or even creating code for␣
 ↪web scraping from scratch.
```

[ ]:
```
1. Human Copy and Paste :
                    Manually copying and pasting data from a web page into␣
 ↪text file or spreadsheet is the most basic form of web scraping.
Even the best web-scraping technology cannot always replace a human manual␣
 ↪examination and copy and paste and this may be the only viable
option when the websites for scraping   explicity prohibit machine automation.
```

[ ]:
```
2 . Text Pattern Matching :
                      The UNIX grep command or regular expression matching␣
 ↪facilities of programming language can be used
to extract information from web pages in a simple yet powerful way HTTP␣
 ↪programming.
```

[ ]:
```
3. HTML  Parsing :
                  Many websites contain large collection of pages that are␣
 ↪dynamically generated from an underlying structured
source,such as a database a common script or template is typically used to␣
 ↪encode data from the same category into similar pages.
A wrapper is a program in data mining that detect such templates in a specific␣
 ↪information sources,extract its content and convert
it to a relational form.
```

[ ]:
```

```

[ ]:
```

```

[ ]:
```
Q3. What is Beautiful Soup? Why is it used?
```

[ ]:
```
ANS -
```

[ ]:
```
Beautiful Soup is a python package for parsing HTML and XML documents␣
 ↪(including having maiformed markup,
i.e., non-closed tags, so named after tag soup). It creates a parse tree for␣
 ↪parsed pages that can be used
to extract data from HTML , which is useful for web scraping.
```

[ ]:

[ ]:

[ ]: Q4. Why is flask used in this Web Scraping project?

[ ]: ANS -

[ ]:
```
Flask is a lightweight framework to build websites. We use this to parse our␣
 ↪collected data and display it as HTML in a new HTML files.
The requests module allows us to send http requests to the webaites we want to␣
 ↪scrape. The first line imports the Flask class
and the render_templates methods from the Flask library.
Flask is a microframework for developers, designed to enable them to create and␣
 ↪scale web apps quickly and simply.
```

[ ]:

[ ]:

[ ]: Q5. Write the names of AWS services used in this project. Also, explain the use␣
 ↪of each service.

[ ]: ANS -

1 AWS Elastic BeanStalk

[ ]:
```
This AWS service supports running and managing web applications. Elastic␣
 ↪BeanStalk allows for the easy deployment
of applications from capacity provisioning , load balancing , and auto-scaling␣
 ↪to application health monitoring.
With its auto-scaling properties, this service simplifies demands in scaling to␣
 ↪adjust to the needs of the business.
it help to manage peaks in workloads and traffic with minimum costs. Basically,␣
 ↪AWS Elastic Beanstalk is a developer-friendly
tool since it manages serves,load balancers, firewalls, and networks simply, AS␣
 ↪a result , this service allows developers to
show much more focus on coding.
```

[ ]:
```
1. AWS EC2 - Elastic Compute cloud :
                                    Amazon Elstic compute cloud is a web␣
 ↪service that provides secure, resizable compute capacity in the cloud
amazon EC2 simple web service interface allows you to obtain and configure␣
 ↪capacity quickly and with minimum efforts.
```

[ ]:
```
2. AWS Lambda :
                Lambda is a great technology choice for background processing␣
 ↪that is triggered by events.
- image transformation for newly uploaded images.
- Real-time metric data processing.
- Streaming data validation, filtering, and transformation.
```

[ ]:
```
3. AWS ECS - Elastic Container Service :
                                        Amazon Elastic Container service is a␣
 ↪highly scalable, fast container management
service that makes it easy to run,stop and manage containers on a cluster. ECS␣
 ↪comes with two lunch types: ECS and fargate.
The container can run on a serveless infrastructure,you can run your tasks and␣
 ↪service on a cluster of Amazon EC2 instances that you manage.
for more control over your ingfrastructure,you can run your tasks and service␣
 ↪on a  cluster of amazon EC2 instance that you manage.
```

[ ]:
```
4. AWS S3 - Simple Stroge Service :
                                    Archive old data that is infrequently␣
 ↪accessed. good alternative for on-premisis
NAS or external hard disks.Helps keep your data safe and secure without the␣
 ↪risk of data loss.
with S3 intelligent-tiering,you can automatically move data to the most␣
 ↪cost-effective access tier without performance
impact or operational overhead.
```