



Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati

Department of Artificial Intelligence & Data Science

Report on One Day Workshop Programme

Title: One Day Workshop on Web Retrieval and Crawling

Date: 5th October 2023

Time: 9.30 AM to 4.30 PM

Venue: Main Building, Computer Lab- 105.

Target Audience: B.E AIDS Students

Number of Participants: 70

Objectives of the Programme:

1. To learn web crawling and its role in collecting data from the web
2. To extract text, image, video, ppt from web and save it in working directory
3. To learn Importance of respecting the robots.txt file
4. To learn strategies for dealing with various data formats.
5. To learn page ranking algorithm.

Speakers: Rushikesh Tanksale (Professional Freelancer, Piyush IT Services Pune)

Inauguration: The workshop started with an inauguration session by Dr. P. M. Paithane (HoD, AI&DS, VPKBIET), in the presence of all staff members. Dr. P. M. Paithane (HoD, AI&DS, VPKBIET), Welcomed the guest, all the staff members and participants. He introduced speaker to the audience and also discussed the rationale of the workshop programme and its key components. He motivated to take the benefit of skills through this workshop.

Mrs. Rohini Naik addressed audience with her words of wisdom and expressed that all members have the ability and responsibility to change themselves to changing technology. She motivated the participants, she added that she is fully confident that teachers are always ready to and would successfully cope up with new challenges for the benefit of the students and institution.

The programme was one day Workshop on Web Retrieval and Crawling. It included two short sessions. Both sessions were delivered by the session speakers. It was a full hands-on session.

Session I: Session I, of the programme focussed on "Basics of Web Retrieval and Crawling" and was delivered by Speaker Rushikesh Tanksale, Professional Freelancer Piyush IT Services Pune.

He explain what web retrieval is and why it's essential. Discuss the key concepts, including web resources, URLs, and HTTP requests. Getting Started with Python. He briefly introduced Python as the primary programming language for web crawling and guide participants through setting up their Python environment, making HTTP requests etc...Also, explained use of Python libraries like HTTP requests to web servers. Demonstrated how to retrieve web content (HTML) from a webpage and parsing HTML document which helped to extract meaningful information. Introduced BeautifulSoup as a popular library for parsing HTML. Guided participants through extracting data from a sample webpage a hands-on example.

Session II: In session II Speaker focussed on “Hands-On Web Crawling”

Speaker highlighted some main points on web crawling. He defined web crawling and its role in collecting data from the web. Explained with use cases for web crawling in various industries, Robots.txt and Ethical Crawling. A hands-on through creating a simple web crawler using Python. Setting up initial URLs, crawl through links, and store data. Handling Different Data formats. Strategies for dealing with various data formats, such as JSON and XML. Demonstrate parse and extract data from different file types. Dealing with Pagination and Infinite Scrolling. Also, explained with hands-on the page ranking algorithm. Encouraged participants to explore these areas further after the workshop.

After the session, area was open for Q&A and troubleshooting

The session ended with a vote of thanks by Mr.Yash Padhye, SAAI Committee President, and Atharva Gosavi, B.E Class Representative Dept of AI&DS, VPKBIET. They thanked staff members for their valuable time for the session. They also thanked all the participants for attending the sessions and believed that they will use the tools and technical skills in their day-to-day activities so as to make the process dynamic.

Speaker summarized the key takeaways from the workshop. Shared additional resources and references for participants to continue their learning journey.

Outcomes of Programme:

1. Learned Basics of Web Retrieval and Crawling
2. Understood techniques for handling web pages
3. Learned the importance of respecting the robots.txt file to avoid legal
4. Understood and implemented page ranking algorithm

Glimpses of the event



Inauguration by Dr.P.M Paithane, HoD



Introduction by Mrs. R. S. Naik



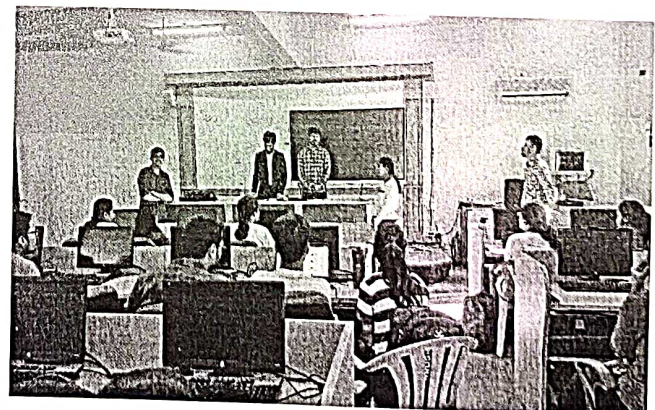
Workshop Speaker Rushikesh Tanksale



Session Participants B.E AIDS Students



Vote of thanks by Yash Padhye & Atharva Gosavi



Students feedback

[Handwritten Signature]

Dr. P. M. Paithane

Head

Department of Artificial Intelligence

& Data Science,

VPKBIET, Baramati 413133

[Handwritten Signature]

Mrs. R. S. Naik

Assistance Professor

Dept of AI&DS

One Day Workshop on Web Retrieval and Scraping

Class: BE - AI&DS

Date:- 05/10/2023

Time:- 9:30 am TO 4:30 pm

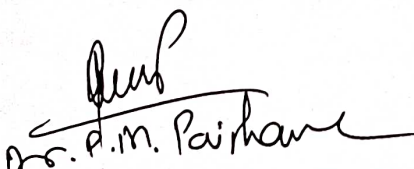
Attendance Sheet

Sr. No.	Roll No	Name of the Student	Signature	
			Session - I	Session - II
1	2347001	BHASKAR SNEHA SANJAYKUMAR		
2	2347002	NIMBALKAR SHIVANJALI VIJAY		
3	2347003	JAWALE VIPASHYANA UTTAM		
4	2347004	PATHARIYA MOHAMMED JOHAR		
5	2347005	DATE ANISH CHARUDATTA		
6	2347006	SUTAR ABHISHEK ASHOK	— AB —	— AB —
7	2347007	MUSKAN JAVEDKHAN PATHAN	— AB —	— AB —
8	2347008	PATIL ANIKET MAHARUDRA		— AB —
9	2347009	HAGARE PRANALI SANDEEP		
10	2347010	TAMBOLI ALFIYA JAVED		
11	2347011	DUCHAL HARSHAD SHAHAJI		
12	2347012	PAWAR VAIBHAV SANJAY		
13	2347013	PAWAR PRATIKSHA SANJAY		
14	2347014	KHANDARE UTKARSH RAJENDRA		
15	2347015	JAGTAP VISHAKHA HANUMANT		
16	2347016	PAWAR RACHANA NANDKUMAR		
17	2347017	KSHIRSAGAR SHARWARI HEMANT		
18	2347018	GONJARI SANIKA ANIL		
19	2347019	KHUSPE HARSHADA NARAYAN		
20	2347020	DHOTRE ROHIT VIKAS		
21	2347021	PAWAR RITIKA AJAY		
22	2347022	CHOUGULE YOGITA DHANANJAY		

Sr. No.	Roll No	Name of the Student	Signature	
			Session - I	Session - II
23	2347023	GADE MAYURI MAHESH	<i>Maade</i>	<i>Maade</i>
24	2347024	SHAIKH KAINA SAILAN	<i>Shaiikh</i>	<i>Shaiikh</i>
25	2347025	SHENDGE VAISHNAVI AJAY	<i>Vaishna</i>	<i>Vaishna</i>
26	2347026	RAUT VEDANT RAJENDRA	<i>Raut</i>	<i>Raut</i>
27	2347027	KULKARNI SHREYAS RAMESH	<i>Kulkarni</i>	<i>Kulkarni</i>
28	2347028	JAGTAP GAURAV MAHENDRA	<i>Jagtap</i>	<i>Jagtap</i>
29	2347029	SHINDE RUSHIKESH SHARAD	<i>Shinde</i>	<i>Shinde</i>
30	2347030	SOHAM VAIDYA	<i>Soham</i>	<i>Soham</i>
31	2347031	KESKAR SHUBHAM DHANANJAY	<i>Keskar</i>	<i>Keskar</i>
32	2347032	BARAWKAR VAIBHAV PRAKASH	<i>Barawkar</i>	<i>Barawkar</i>
33	2347033	GADADE AKASH VIJAY	<i>Akash</i>	<i>Akash</i>
34	2347034	KUMBHAR GAURAV CHANDRAKANT	<i>Kumbhar</i>	<i>Kumbhar</i>
35	2347035	BARAVKAR MAYUR MANSING	<i>Baravkar</i>	<i>Baravkar</i>
36	2347036	SALVE SELVIN PRASHANT	<i>AB</i>	<i>AB</i>
37	2347037	KHARAT TRUPTI MOHAN	<i>Kharat</i>	<i>Kharat</i>
38	2347038	JAGTAP SNEHAL SHYAM	<i>Jagtap</i>	<i>Jagtap</i>
39	2347039	RAUT ABHISHEK RAJENDRA	<i>Raut</i>	<i>Raut</i>
40	2347040	NASHIKKAR RISHI NIRAJ	<i>Nashikkar</i>	<i>Nashikkar</i>
41	2347041	GOSAVI ATHARVA NITIN	<i>Gosavi</i>	<i>Gosavi</i>
42	2347042	DAWGHT ADITYA SUNIL	<i>Daght</i>	<i>Daght</i>
43	2347043	SOMVANSI SHREYASH SHASHIKANT	<i>AB</i>	<i>AB</i>
44	2347044	GAIKWAD SHUBHAM VILAS	<i>Gaikwad</i>	<i>Gaikwad</i>
45	2347045	YASH NIMISH PADHYE	<i>Padhye</i>	<i>Padhye</i>
46	2347046	KHARADE HRUSHIKESH BHARAT	<i>Kharade</i>	<i>Kharade</i>
47	2347047	DAMARE YAISHNAVI DATTATRAY	<i>Damare</i>	<i>Damare</i>
48	2347048	VANARASE RADHIKA SANTOSH	<i>Vanarase</i>	<i>Vanarase</i>
49	2347049	PALIWAL KALYANI KAILAS	<i>Paliwal</i>	<i>Paliwal</i>
50	2347050	SHELKE KANCHAN MILIND	<i>Shelke</i>	<i>Shelke</i>
51	2347051	ARUDE SAKSHEE SUNIL	<i>Arude</i>	<i>Arude</i>

Roll No	Name of the Student	Signature		
		Session - I	Session - II	
52	2347052	MANE VEDANTI DILIP	AB	AB
53	2347053	LAWAND AASTHA RAVINDRA	Lawand	Lawand
54	2347054	RUPALI PALIWAL	Rupalwal	Rupalwal
55	2347055	JAGTAP SAYALI SHAMRAO	Jagtap	Jagtap
56	2347056	JADHAV VIJAYRAJE CHANDRAKANT	AB	AB
57	2347057	GHUMARE SUMEET SANTOSH	Ghumare	Ghumare
58	2347058	SHINDE YADNYASEN SANGRAM	Shinde	Shinde
59	2347059	PRAJWAL RUDRAPWAR	Prajwal	AB
60	2347060	PAWAR SUYOG SUDHIR	Pawar	Pawar
61	2347061	GADHAVE SIDDHESH RAJENDRA	Gadhav	Gadhav
62	2347062	SUJAY TANAJI SHINDE	Sujay	Sujay
63	2347063	BHISE KARAN SAMPAT	Bhise	Bhise
64	2347064	PAWAR SAMIKSHA SANTOSH	Pawar	Pawar
65	2347065	JADHAV VISHAKHA SANJAY	V.S. Jadhav	V.S. Jadhav
66	2347066	JALKOTE PRATYUSH BASAVRAJ	Pratyush	Pratyush
67	2347067	CHOTHAVE PRANAV RAJU	Pranav	Pranav
68	2347068	KAMBLE SAHIL SOMAJI	Kamble	Kamble
69	2347069	PATIL SIDDHESHWAR SHIVRAJ	Patil	Patil
70	2347070	MEHETRE ROHIT RAJHANS	Rohit	Rohit
71	2347071	KUMBHAR AVDHOOT RAJENDRA	Kumbhar	Kumbhar
72	2347072	VEER TEJAS BABASAHEB	Veer	Veer
73	2347073	KUBDE ABHISHEK RAJKUMAR	Kubde	Kubde
74	2347074	MAMADGE TUSHAR VISHWAS	Mamadge	Mamadge
75	2347075	DEV RAT YASHRAJ DEEPAK	Devrat	Devrat
76	2347076	KULKARNI VEDANT MAHENDRA	AB	AB
77	2347077	WAGHCHOUDE HARSHADA PRAVIN	AB	AB

Timestamp	Name of Student	Roll Number	Workshop objectives were stated clearly and met.	Workshop was well organized.	The information and/or skills presented were relevant and useful.	The resource person provided adequate time for questions and answered them satisfactorily.	Would you like to have such kind of workshops in future?	What topics would you like to see covered in future workshops?	Overall Rating for this workshop	Any Suggestion/Comments:
2023/10/06 11:38	Kharade Hrushikesh Bharat	2347046	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Machine learning / IIOT	5	Excellent
2023/10/06 11:40	Pawar Suyog Sudhir	2347060	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Quantum Neural Networks	5	NA
2023/10/06 11:41	Shinde Sujay Tanaji	2347062	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	IR In detail	5	
2023/10/06 11:41	Kubde Abhishek Rajkumar	2347073	Strongly Agree	Strongly Agree	Agree	Strongly Agree	Yes	Data Science , Data analytics	5	No
2023/10/06 11:43	Gosavi Atharva Nitin	2347041	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Data science and analytics	5	It was a wonderful and informative session.
2023/10/06 11:53	Gadade Akash Vijay	33	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Web Scraping related concepts , Unix internals , Operating system Internals, and Android applications	5	They just showed the codes on screen, and give task to copy the same and implement it. Then they
2023/10/06 12:28	Paliwal Rupali Galendra	54	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Linux system programming, OS Internal,	5	NA
2023/10/06 12:32	Kharat Trupti mohan	37	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Internet of things	5	No
2023/10/06 12:40	Dae Anish Charudatta	2347005	Agree	Strongly Agree	Agree	Strongly Agree	Yes	Any Emerging technology	5	Thank you mam for arranging this workshop
2023/10/06 12:50	Barawkar Vaibhav Prakash	2347032	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Data Science	5	
2023/10/06 1:00:3	Duchal Harshad Shahaji	2347011	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	NLP	5	
2023/10/06 1:01:0	Nashikkar Rishi Niraj	40	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	IIoT	5	It was amazing
2023/10/06 1:16:2	Vaidya Soham Nilesh	2347030	Strongly Agree	Strongly Agree	Strongly Agree	Agree	Yes	Quantum Computing, Deep learning	5	
2023/10/06 1:29:0	Tamboli Alfiya Javed	10	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Machine Learning, Data Visualization	5	
2023/10/06 5:28:3	Pawar Vaibhav Sanjay	2347012	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Machine Learning and Data visualisation	5	
2023/10/06 5:29:0	Pawar Pratiksha Sanjay	2347013	Strongly Agree	Strongly Agree	Strongly Agree	Agree	Yes	Machine learning	5	
2023/10/06 5:40:3	Dawghat Aditya Sunil	42	Agree	Agree	Agree	Agree	Yes	Mlops	5	
2023/10/06 6:57:0	Shendge Vaishnavi Ajay	25	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Cloud for Data science/Data analytics	5	
2023/10/06 7:14:4	Jagtap Vishakha Hanumant	2347015	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Distributed computing	4	Improvement for explaining concept in detail .
2023/10/06 9:13:3	Devrat Yashraj Deepak	75	Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	ML	5	
2023/10/07 8:10:5	Nimbalkar Shivanjali Vijay	2347002	Strongly Agree	Strongly Agree	Strongly Agree	Agree	Yes	Cloud Computing & Deep Learning	5	
2023/10/07 10:57	KUMBHAR AVDHOOT RAJENDRA	71	Strongly Agree	Agree	Agree	Strongly Agree	Yes	MLOps	5	
2023/10/07 11:45	Padhye Yash Nimish	45	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	UNIX SYSTEM INTEGRALS	5	
2023/10/07 11:45	Kumbhar Gaurav Chandrakant	34	Agree	Agree	Neutral	Agree	Maybe	Machine learning	3	
2023/10/07 11:46	Lawand Aastha Ravindra	53	Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	-	5	
2023/10/07 11:48	Shaikh Kaina Sallan	2347024	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	IIOT	5	
2023/10/07 11:56	rudrapwar prajwal rajendra	2347059	Strongly Agree	Agree	Strongly Agree	Agree	Yes	IIot	4	
2023/10/07 12:25	Arude Sakshee Sunil	51	Agree	Strongly Agree	Agree	Agree	Yes	Virtual Reality	4	
2023/10/07 4:48:5	Kshirsagar Sharwari Hemant	17	Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes	Machine learning	5	
2023/10/07 4:49:4	Gonjari Sanika Anil	18	Agree	Strongly Agree	Agree	Agree	Yes	Gen AI	5	No
2023/10/07 6:46:3	Damare Vaishnavi Dattatray	2247047	Agree	Agree	Agree	Agree	Yes	Advanced scraping techniques,	5	
2023/10/07 7:34:2	Pawar Rachana Nandkumar	16	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Yes			
2023/10/08 11:20	Pawar Samiksha Santosh	64	Agree	Agree	Agree	Agree	Yes			
2023/10/08 11:57	Veer Tejas Babasaheb	72	Agree	Agree	Strongly Agree	Strongly Agree	Yes			


 Dr. P. M. Patil
 Head (A. I. D. S.)
 Department of Artificial Intelligence
 & Data Science,
 VPKBIET, Baramati 413 133