



KIET
GROUP OF INSTITUTIONS
Connecting Life with Learning



**A Project Report
ON
Dark Web Crawler
BACHELOR OF TECHNOLOGY
DEGREE**

SESSION 2023-24

in

Name of discipline

By

Kartikeya Srivastava (2000290100078)

Rishi Srivastava (2000290100116)

Really Singh (2000290100112)

Under the supervision of

Asst.Prof Mr.Gaurav Parashar

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow
(Formerly UPTU)
December, 2023

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature

Name:

Roll No.:

Date:

CERTIFICATE

This is to certify that Project Report entitled “Dark Web Crawler” which is submitted by Student name in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science & Engineering of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

.

Date:

Supervisor Name

(Designation)

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. Final Year. We owe special debt of gratitude to supervisor name, Department of Computer Science & Engineering, KIET, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Vineet Sharma, Head of the Department of Computer Science & Engineering, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members, especially faculty/industry person/any person, of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Date:

Signature:

Name :

Roll No.:

ABSTRACT

The "Dark Web Crawler" project aims to develop a specialized tool for systematically exploring and indexing content within the secretive Dark Web. It employs intelligent crawling algorithms, adapted for the unique structures of hidden websites, leveraging machine learning to optimize information extraction.

Content Extraction and Analysis:

Using advanced techniques like natural language processing and computer vision, the Dark Web Crawler extracts meaningful content, including text, images, and multimedia. The inclusion of sentiment analysis enhances understanding of the content's nature.

Distributed Architecture:

To tackle the challenges of the Dark Web, the project adopts a distributed crawling system for enhanced scalability and reduced processing time through parallel processing and distributed storage.

Webpage Rendering and Dynamic Content Handling:

The Dark Web Crawler utilizes browser automation to navigate dynamically generated content and JavaScript interactions, ensuring comprehensive exploration of hidden websites.

User-Friendly Interface and Customization:

Tailored for Dark Web exploration, the project features a user-friendly interface for configuring crawling parameters specific to this environment, allowing easy customization to meet unique challenges.

Ethical and Responsible Crawling: Due to the sensitive nature of Dark Web content, ethical considerations are a priority. The Dark Web Crawler adheres to responsible crawling practices, maintaining ethical standards in its exploration of this clandestine space.

TABLE OF CONTENTS

Page No.

DECLARATION.....	ii
CERTIFICATE.....	iii
ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xi
LIST OF ABBREVIATIONS.....	xii
1 INTRODUCTION.....	1
2. PROGRESS.....	2

INTRODUCTION

The "Dark Web Crawler" project addresses the intricate task of systematically navigating and cataloging content within the obscured realm of the Dark Web. Unlike the surface web, the Dark Web operates on a hidden network, presenting unique challenges for information retrieval. This project endeavors to develop a specialized web crawling system equipped with intelligent algorithms designed to adapt to the clandestine structures of hidden websites. By leveraging advanced techniques, including machine learning, natural language processing, and computer vision, the Dark Web Crawler aims to extract meaningful content, such as text, images, and multimedia, contributing to the creation of a comprehensive repository of Dark Web data. This endeavor not only involves technical challenges but also emphasizes the ethical responsibility of handling sensitive information within this unconventional online space.

PROGRESS

1. Research paper is in progress.
2. Project is ready but have some bugs and we are fixing it.

THANK YOU