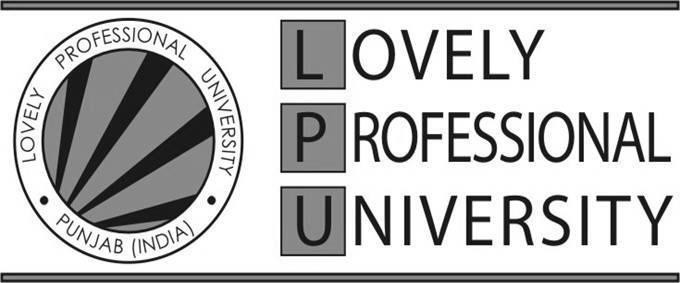
**PROJECT REPORT**

(Project Term November 2020)



Submitted by

**(Potnuru. Sandeep) Registration Number:11910309 (Kondaveeti Lakshman Kumar) Registration Number: 11910186**

**(Kuriti Jayasai) Registration Number: 11910262**

**Course Code INT-213**

Under the Guidance of

**(Mrs. Sagar Pandey, Professor)**

**School of Computer Science and Engineering**

# **TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **1. Abstract ………………………………. 4**  **2. Introduction ………………………………. 5**  **3. Literature ………………………………. 6**  **4. Conclusion ………………………………. 7** |  |  |
| **5. Result ………………………………. 8 - 10** |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**ACKNOWLEDGEMENT**

In this present world of competition there is a race of existence in which those are having will to come forward and succeed. Project is like a bridge between theoretical and practical working. With this willing we joined this project. We would like to express our deepest appreciation to all those who provided us the possibility to complete this report. A special gratitude we give to our final year project mentor, [Mrs. Usha Mittal], whose contribution in stimulating suggestions and encouragement, helped us to coordinate my project especially in writing this report and completion of project. It was so helpful for us and good to working with her.

ABSTRACT

As the Internet and World Wide Web continue to expand and amass more users, increased rates of crime occur. One such crime is modern slavery, or human trafficking. Social media and web forums are often employed by traffickers to recruit and advertise victims anonymously. While this issue continues to propagate through both the surface web and the dark web, web scraping tools must be developed to extract and analyse the information on these websites to identify traffickers and victims of human trafficking. This project aims to investigate how crime continues to occur on the Internet and where on the web this is. Solutions to the problem using current web scraping technology are researched, along with other, more freely available web scrapers. The proposed system for this project is a web scraper that can access and extract data from websites using a web application as an interface for user interaction. The extracted data is then stored in a database, as the web application allows the user to search through and query the saved findings. When the system has been fully implemented, a reflection on the completed system takes place, judging to see if a web scraper can successfully be implemented to combat the issue of human trafficking.

INTRODUCTION

Since the dawn of the digital age, many people have used the Internet and the World Wide Web (web) in their everyday lives. These users often use the web as a means of communication, with forums, chatrooms and even video hosting sites being the key platforms. Users of the Internet are, for the most part, anonymous, which allows users to openly discuss personal topics and allows users to be free of discrimination, as people are equal under anonymity (Palmer and Berglund, 2004). While encouraging users to become anonymous has allowed the World Wide Web to become a success, it has also allowed criminals to use the web to commit crime while hiding behind a wall of anonymity. A major crime that is committed using the web as a platform, is modern slavery, or human trafficking. More accurately, the web is used to recruit potential victims, luring them to take risks with the promise of a better life (Logan et al., 2009). In 2012, there was an estimated 20.9 million victims of forced labor in the world, many of which are also victims of human trafficking (ILO, 2012). As this is an estimate, real victims are often hard to spot with current methods of detection, as only 21,251 victims were detected between 2014 and 2016 (UNODC, 2016a). As human traffickers, and perpetrators of other crimes, use the web to commit offences, specialist tools must be used to track down these instances. One such method is to use a web scraper. Web scrapers are closely related to web crawlers and while there are no real industry definitions for these terms, a distinction is often made. Web crawlers, or spiders, are often used in search engines, as they crawl or scan through a website, looking for links for indexing purposes. Web scrapers may still crawl through a website; however a web scraper is a software that extracts data from webpages and stores the data in a database for further manipulation. Due to the current number of victims and perpetrators of human trafficking that are unaccounted for, law enforcement agencies must turn to specialized tools, like web scrapers, to have a chance of intercepting these cases. Government bodies and non-governmental organizations alike have been recently researching and developing web scrapers for use in law enforcement, as manually searching through websites is highly inefficient and wastes time and money.

LITERATURE

As the Internet sets out to connect the world, more people than ever are using the Internet and the World Wide Web in their everyday lives. Estimates have found that in 2000, there were around 400 million users of the Internet, while in 2015, there were around 3.2 billion users, an increase of 2.8 billion in 15 years (ITU, 2015).

As the Internet and the World Wide Web continue to grow and amass more users, there is a higher risk of criminal activity. Currently, the web is uncontrolled and unregulated, which allows anyone to utilize it for communication and create their own websites with documents, links, and their own content (Albert et al., 1999). This is, in part, due to the lack of centralized control on the web, with no governing bodies to regulate content (Castillo, 2004). What this means is that it is possible for someone with intent to commit a crime to use an existing website or create a website to use as a platform for criminal activities. There are many types of crime committed on the Internet, ranging from viruses and malware to fraud and extortion. Extortion is defined as “an incident when a cybercriminal demands something of value from a victim by threatening physical or financial harm or the release of sensitive data” (IC3, 2017). This can include denial of-service (DoS) attacks, government impersonations schemes and sextortion. Sextortion is defined by the Internet Crime Complaint Center (IC3) as “a situation in which someone threatens to distribute your private and sensitive material if you don’t provide them images of a sexual nature, sexual favors, or money” (IC3, 2017).

There have been numerous estimates of the number of victims of human trafficking in the world. In 2012, there were an estimated 20.9 million victims worldwide (ILO, 2012), while the United Nations High Commissioner for Refugees (UNHCR) estimated that in 2015, more than 65 million victims may exist in the world (UNODC, 2016b). These, however, are estimates, as the number of detected victims between 2014 and 2016 was only 21,251 (UNODC, 2016a). What this shows is that with current methods of detection, many victims go unnoticed. How many people become victims of trafficking is through the promise of a better life. Victims are told to forward their information to make forged documents with the promise that they will have a better life in a new country (Logan et al., 2009). These are simply promising, as in reality, the victims are often paid equivalent to the minimum standard wages in their countries of origin (Europol, 2017). As the world moves to make use of the Internet and the web in everyday lives, traffickers have also moved to this digital platform. Traffickers are using social media and web forums as online advertisements to recruit victims and using the web to purchase tickets to transport the victims (McAlister, 2015). Due to the large proportion of victims of human trafficking who are undetected by the United Nations (UN), the 2015 estimated over 65 million victims (UNODC, 2016b) versus the 21,251 detected victims between 2014 and 2016 (UNODC, 2016a), it can be gathered that the current methods of detection are not able to spot victims who are recruited online.

CONCLUSION

While this project may not be as sophisticated as web scrapers made by large corporations, there is enough scope in this application to make a significant impact in the world.

Result

from bs4 import BeautifulSoup Import requests

Wiki = requests.get(“https://en.wikipedia.org/wiki/Coronavirus”)

Soup=BeautifulSoup (wiki . text, ’html’ )

Print (soup . find (‘title’ ))

………………………………………………………………………………………………………

<title>coronavirus</title>

covid = soup. find\_all("div",class\_='toc’)

for i in ww2\_contents:

print (i . text)

………………………………………………………………………………………………………

Contents

1 Etymology

2 History

3 Microbiology 3.1 Structure 3.2 Genome 3.3 Replication cycle

3.3.1 Cell entry 3.3.2 Genome translation 3.3.3 Replicase-transcriptase 3.3.4 Assembly and release

3.4 Transmission

4 Classification

5 Origin

6 Infection in humans 6.1 Common cold 6.2 Severe acute respiratory syndrome (SARS) 6.3 Middle East respiratory syndrome (MERS) 6.4 Coronavirus disease 2019 (COVID-19)

7 Infection in animals

7.1 Farm animals 7.2 Domestic pets 7.3 Laboratory animals

8 Prevention and treatment

9 See also

10 References

11 Further reading

covid = soup. find\_all (“div", class\_='catlinks’)

for i in ww2\_contents:

print (i.text)

………………………………………………………………………………………………………

Categories: Animal virologyCoronaviridaeHidden categories: CS1 French-languag e sources (fr)Webarchive template wayback linksArticles with short descriptio nShort description is different from WikidataWikipedia extended-confirmed-pro tected pagesUse dmy dates from March 2020Articles with 'species' microformats Articles containing Ancient Greek (to 1453)-language textAll articles with un sourced statementsArticles with unsourced statements from July 2020Commons ca tegory link is on WikidataCS1: long volume valueTaxonbars desynced from Wikid ataTaxonbars using multiple manual Wikidata itemsTaxonbars on possible non-ta xon pagesWikipedia articles with GND identifiers

# findall function is used to fetch all tags at a single time.

links=soup. find\_all('a') # here i extracted all the anchor tags of my website total\_links=len(links) # len function is used to calculate length of your array

print ("total links in my website :",total\_links)

print ()

for i in links [:2]: # here i use slicing to fetch only first 6 links from rest of the m.

print(i)

………………………………………………………………………………………………………

total links in my website: 1981

<a id="top"></a>

<a href="/wiki/Wikipedia:Protection\_policy#extended" title="This article is e

xtended-confirmed protected"><img alt="Extended-protected article" data-fileheight="512" data-file-width="512" decoding="async" height="20" src="//uploa

d.wikimedia.org/wikipedia/en/thumb/8/8c/Extended-protection-shackle.svg/20pxExtended-protection-shackle.svg.png" srcset="//upload.wikimedia.org/wikipedi

a/en/thumb/8/8c/Extended-protection-shackle.svg/30px-Extended-protection-shac

kle.svg.png 1.5x, //upload.wikimedia.org/wikipedia/en/thumb/8/8c/Extended-pro

tection-shackle.svg/40px-Extended-protection-shackle.svg.png 2x" width="20"/></a>