Perceptual Operating Systems for the Trade Associations of Cyber Criminals to Scrutinize Hazardous Content

Romil Rawat, Department of Computer Science Engineering, Shri Vaishnav Institute of Information Technology, Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore, India*

Anand Rajavat, Department of Computer Science Engineering, Shri Vaishnav Institute of Information Technology, Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore, India

ABSTRACT

The limits of user visibility have been exceeded by the internet. The “Dark Web” or “Dark Net” refers to certain unknown portions of the internet that cannot be found using standard search methods. A number of computerised techniques are being explored to extract or crawl the concealed data. All users can freely interact on the surface web. Identity identities may be found on the deep web, and the dark web (DW), a hub for anonymous data, is a haven for terrorists and cybercriminals to promote their ideologies and illegal activities. Officials in clandestine surveillance and cyberpolicing are always trying to track down offenders’ trails or hints. The search for DW offenders might take five to ten years.The proposed study provides data from a DW mining and online marketplaces situation from a few domains, as well as an overview for investigators to build an automated engine for scraping all dangerous information from related sites.

KEYWORDS

Content Crawling, Cyber Attack, Dark E-Markets, Darkweb, Healthcare, Terrorist, Tor Network, Web Crawler

1. INTRODUCTION

A shadow economy (Gupta et al, 2021) (Weimann,2016) (Tsuchiya & Hiramoto, 2021) by Squires in 2021 A darknet (DN) business website called (Sonmez & Codal, 2022) leverages services like Tor or I2P (Gupta et al., 2021). (Weimann, 2016). They typically act as “black market places (MPS)”, selling or brokering deals involving illegal items such as drugs, cyber-arms, weapons, hijacked credit card details, forged documents, anabolic steroids, and other illegal assets (Rawat, 2023). DN marketplaces were the second-most popular Tor sites, per a study by the University of Portsmouth’s (Squires, 2021) and (Sonmez & Codal, 2022) researchers.

Numerous evil organisations, including terrorist organisations and hackers, are drawn to the DW’s uncontrolled and unregulated character (Gupta et al., 2021). Terrorist organisations may sell their ideas, recruit, share skills, train, market, finance, target, and develop diverse communities without

DOI: 10.4018/IJCWT.343314

*Corresponding Author

concern for location or even the presence of a local leader thanks to the DW's anonymity features (Weimann, 2016) (Sonmez & Codal, 2022).

Similarly, the DW (Wang et al., 2021) enables anonymous information sharing among hackers. DW forums are frequently the subject of many types of surveillance, ranging from manual observation to crawling mixed with natural language processing (NLP) (Saharan et al., 2024) techniques for automated threat intelligence. Terrorism or cybercrime (Gupta et al., 2021) that individuals or well-organised organisations can carry out (Weimann, 2016)(Tsuchiya & Hiramoto, 2021) on the DW. Cybercrime is becoming more accessible to anybody who wants to engage in low-risk illegal activities while still making a difference (for example, launching DDoS (Alshammery & Aljuboori, 2022) assaults on websites is as simple as contracting a botnet that provides DDoS-as-a-Service). These services allow criminals to take advantage of the ”low hanging fruit” (targets without adequate security controls or training). Tor's hidden services can let attackers and victims maintain command-and-control (C2) (Gupta et al., 2021)(Weimann, 2016) (Sonmez & Codal, 2022) communications. Tor's anonymity (and difficulties in shutting it down) is excellent for C2 servers, and it is one of the most widely used hidden services.

States are concerned about the necessity of preparing for digital warfare (Saharan et al., 2024), particularly when it impacts critical infrastructure (CI) and industrial control systems (ICS) and has the potential to have negative real-world consequences. Because of the asymmetry of the wartime environment, it is even easier to become a cybercriminal (Fu & Li, 2021). Despite the fact that cybercriminals are not as well-funded or resourced as the organisations they target, they have an edge on the digital battlefield because they can select their tactics, timing, and location, whereas defenders must always be alert. Deterrence and dissuasion have been effective military measures in the past for a variety of reasons, including the high barrier to entry into nuclear weapons (Gupta et al., 2021) (Weimann, 2016)(Tsuchiya & Hiramoto, 2021)(Squires, 2021). This does not apply in cyberspace, because a weapon may be simply coded or purchased on the DW. Deterrence is no longer just the realm of states, since non-state entities become active players in cyber warfare against shared adversaries. Terrorist organisations' activity on the surface web has been reduced by law enforcement authorities and hacktivist groups all across the world. These terrorist organisations have shifted to the DW; their followers may anonymously voice their thoughts; their activities can continue to be supported through virtual currencies; and the DW can be used as a possible recruiting (Alshammery & Aljuboori, 2022) and training ground (Alshammery & Aljuboori, 2022). The latter has been linked to a considerable amount of terrorist activity, and NLP is being used to identify it on DW forums (Mili & Rodin, 2022).

The system has an accessibility feature that allows users to register for DW forums with the assistance of a third party. The solution uses a combination of dynamic proxies and topic-specific spidering parameter setups to assure forum access. The URL Ordering (AlKhatib & Basheer, 2019) component uses language-independent URL ordering features to facilitate spidering of DW forums across languages. It is planned to focus on three separate groups from three different regions: the United States, the Middle East, and Latin America/Spain. Along with BFS and DFS crawling crawling(AlKhatib & Basheer, 2019), a rule-based URL sorting technique (Tsuchiya & Hiramoto, 2021) (Cole et al., 2021) is employed (Fu & Li, 2021) to traverse space. This technique is used to limit the number of unwanted web pages acquired. An incremental crawler that decides which threads need to be collected using forum wrappers. The system will contain a recall improvement mechanism that parses the spidering log (Gupta et al., 2021) (Weimann, 2016) and reinserts incomplete downloads into the crawl area. Finally, the system has a collection analyzer that finds duplicate downloads and generates collection statistics (Cole et al., 2021) at the forum, region, and worldwide levels.

Cybercrime poses a severe threat to the institution and its patients by stealing medical records (Ahmad et al., 2022). In many ransomware (Weimann, 2016)(Tsuchiya & Hiramoto, 2021) cases, the hacker would steal patient information and sell it on the DW, which is a secret section of the internet. In 1971 or 1972, Stanford students (Squires, 2021) and their colleagues at the Massachusetts Institute of Technology (MIT) used ARPANET (Advanced Research Projects Agency Network)

Network)(Sonmez & Codal, 2022) accounts in their labs to exchange marijuana. Online drug selling has become increasingly common as the Internet and social media have grown in popularity, with the most well-known network being ”Silk Road (Sonmez & Codal, 2022), which was seized by the FBI in 2013(Squires,2021). Within two years of its launch, the FBI said that this website had a 1.2 billion dollar turnover (Sonmez & Codal, 2022)(Faizan & Khan, 2019)(Hayes et al.,2018), including 80 million dollars in commission for administrators. Nowadays, there are crypto MPS where anybody can buy anything illegal (Hayes et al., 2018), from benzodiazepines (Hayes et al., 2018) to cocaine (Hayes et al., 2018), and have it delivered to practically any location in the world by ordinary mail. Consumers appear to be happy with purchases made on the DW.

Because crypto (Paul, 2018) MPS include potency ratings for medicines and a comprehensive feedback loop, drug purity and vendor accountability are unexpectedly high. As a result, there is less possibility of contamination, dilution with other items, or overdosing than in street transactions. Furthermore, the risk of direct physical assault or harassment by drug-dealing gangs(Saharan et al., 2024) has decreased owing to anonymity. On the DW, there are several harm reduction forums where users discuss information on how to use a drug safely (i.e., the ideal dose, personal experience with a drug, and even advice on how to quit a drug), essentially in a non-judgmental ”Narcotics Anonymous (Robertson et al., 2016) ” forum. Few industries have bigger cybersecurity stakes than healthcare (Alshammery & Aljuboori, 2022), with medical institutions (Alshammery & Aljuboori, 2022) at risk of life-threatening interruptions from malevolent actors. Beyond the more serious consequences, the sensitive nature of the data these organisations store (e.g., social security numbers, blood types, patient history, etc.) means that if their data is exposed through an insecure service provider, patients could become victims of impersonation, fraud, theft, and manipulation.

Understanding the sorts of cyberattacks that the healthcare industry (Robertson et al., 2016) (Alshammery & Aljuboori, 2022) faces is just as critical as understanding the financial motives that drive hackers. Healthcare provider data, coupled with Personal Health Information (PHI), falsified prescriptions, and health insurance login material, is now some of the most valuable information on the DW. Most of the time, this information is in the form of administrative documentation that might be used by a hacker to impersonate a real doctor. Once the hacker (Alshammery & Aljuboori, 2022) has this information, he or she may sell it on the DW to purchasers who would masquerade as the doctor and make fake Medicare or insurance claims, or even claims for pricey, high-end treatments, pocketing the money and leaving the victims to cope with the consequences. This kind of information usually sells for 500 dollars per listing (Alshammery & Aljuboori, 2022)(Bracci et al., 2021). After compromising a web server or credential database, a hacker will sell the target information to a bidder for a low price. The buyer will rapidly log in and acquire access to medical insurance information before the data becomes old or outdated, potentially combining it with fake medical information to obtain services at the victim’s expense. Due to the huge volume and turnover rate of this data, it is sometimes sold on the DW for as low as 3.25 dollars (Heinl et al.,2019)(Lee et al.,2022). In some circumstances, the vendors are supplied with the information they need to falsify a prescription, which they subsequently pass on to their buyer. These fraudulent documents can subsequently be used to carry illegal substances, with a trafficker (Lee et al., 2022) flashing the prescription to legitimise their actions.

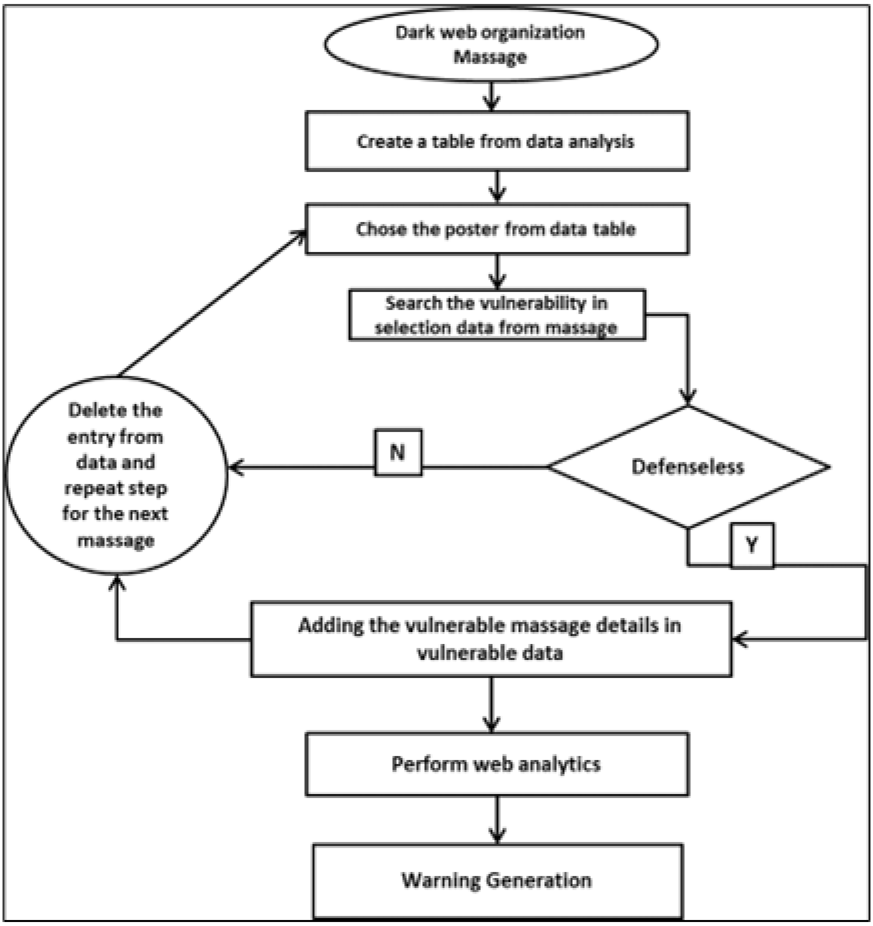
1.1 International Terror

Drug trafficking (Bracci et al., 2021)(Heinl et al., 2019), contract killing (Bracci et al., 2021), weapon sales (Lee et al., 2022), pornography (Lee et al., 2022), child abuse (Lee et al., 2022), malware (Lee et al., 2022), hacking software (Lee et al., 2022), Bitcoin service (Nadini et al., 2022), forgeries, drug trafficking, and identity theft are just a few of the bad activities that employ Darkweb(DW) technologies . The majority of this research focused on terrorism and extremism, including determining terrorist groups’ identities, propaganda, and philosophy.

Extremist organisations (Fu & Li, 2021) and terrorist supporters abuse online social networks (OSN) (Bergeron et al.,2022) by sharing organized posts (propaganda, radicalization or strategic operations, improvised hazardous gadgets (IEDs), including statistics for homemade explosives (HMEs) (Mili & Rodin, 2022) and containing motivating and enticing images and opportunities to join terrorist groups). Instruction to initiate terrorism destruction, extremists, lone-wolf attackers, and sleeper cells always wait for the general command structure of an organization. Figure 1 below shows the DW malicious content alert. DW crime is a major problem faced by several countries, even affecting their economies. Security agencies, along with social network researchers, are developing algorithms and designs so that illicit and malicious content can be traced and warnings generated by associated users. The warnings are generated based on the malicious fingerprinting of cybercriminals and terrorists collected in the data table. The marked suspicious entry is matched with the record (AlKhatib & Basheer, 2019). If a similarity is found, a warning is generated, and the page can be blocked. Figure 1 shows the architecture of the DW malicious content alert (Bergeron et al., 2022).

Cyberspace (Vajjhala and Strang, 2023) is made up of a huge variety of interlinked networks, systems, and gadgets that make communication and data exchange possible. The Internet, wireless networks, and other electronic gadgets like cellphones, computers, and servers are all included.

Figure 1. DW Malicious Content Alert (Rawat et al., 2022)


Dark web organization Massage
Create a table from data analysis
Chose the poster from data table
Search the vulnerability in selection data from massage
Defenseless
N
Y
Delete the entry from data and repeat step for the next massage
Adding the vulnerable massage details in vulnerable data
Perform web analytics
Warning Generation

The usage of the worldwide Internet for many reasons, from business to leisure, is one approach to discussing cyberspace and also the attracting platform for cyber terrorism (Strang and Vajjhala,2023) to conduct illicit events. They observe the existence of cyberspace whenever actors build up virtual meeting places. It might be said that whenever the Internet is used, a cyberspace is created. In a real (albeit somewhat theoretical) sense, cyberspace is expanding due to the widespread usage of desktop computers and smartphones to access the Internet.

1.2 Discussion

Cyberterrorism (Vajjhala and Strang, 2023) is the use of computers and networks to intimidate and coerce a person, group, or even an authority for personal, political, or societal benefit. Cyberterrorism, sometimes referred to as digital terrorism, is the term used to describe disruptive threats made against computer systems by recognised terrorist organisations with the aim of causing fear or physical harm to the information system.

Despite the fact that people are accustomed to hearing about cyberattacks, cyberterrorism causes a distinct kind of concern. Cyberterrorists may also acquire confidential material since computer hackers have long sought to do so for financial gain. Cyberterrorists may plot their incidents, recruit new terrorists, and fund their activities online. Hacking into public or private systems to get access to confidential data or even to steal money for use in extremism is the more common conception of cyberterrorism.

The online gaming platforms that are marketed as having big online players with fraud identities for virtual terrorism events (Vajjhala and Strang, 2022)(Ejazi, 2022) ecosystems are another excellent illustration of cyberspace. These massive gaming categories, also referred to as “meatspace,” build their own virtual worlds while playing alongside those that exist solely online and not in the real world.

Think about what occurs when thousands of individuals who would have previously convened in actual rooms to play a game instead do it by each staring into a gadget from faraway regions in order to truly understand what cyberspace implies and what it is. In a way, game operators are introducing interior design to cyberspace by dressing up the interface to make it engaging and secure (Strang et al., 2018) from the outside world having fake identities (Topor and Pollack, 2022).

The traditional method of computer security (Vajjhala and Strang, 2023) has been referred to as perimeter defence. A sub-network’s entrance point is protected by routers and firewalls to prevent access from outside intruders. The ineffectiveness of the perimeter defence strategy is well known among cybersecurity specialists. Every one of these defences can eventually be broken through or avoided. Systems can nevertheless be infiltrated even in the absence of such breaches, for example, when malicious actors currently exist within the perimeter or when a denial-of-service threat causes servers to fail by flooding them with fictitious requests.

A harmful threat by cybercriminals or terrorists attempting to enter a computer network, destroy data, or steal sensitive information is referred to as a threat in cyberspace.

1.3 Novelty and Contribution

- The work presents an ontology for extracting dark web marketplaces (OEDWM) for cyberspace illicit events by dark web criminals or cyber terrorists.
- For the purpose of illustrating, OEDWM is designed to show the algorithm with a crawling flow in the context of cyberterrorism.
- From online platforms or channels Keywords like cyber terrorism, crime, violence, territory, fraud, propaganda, nation, religion, location, online channels, ransom, time, and language are essential concepts in the OEDWM.

The remainder of the paper is laid out as follows: Section 2 discusses related work; Section 3 highlights Ontology Extracting DarkWeb Marketplaces (OEDWM); Section 4 discusses layers of the internet; Section 5 discusses the Onion Router (TOR) layout; and finally, Section 6 concludes the paper.

2. RELATED WORK

(Topor and Pollack, 2022) provide a framework for ML (machine learning) with real-time forecasting to defend intricate system operations and factory automation chains from cyber-attacks. The study methodology uses multivariate time series characterizations and real-time predictive modelling to forecast risks, estimate the time for identifying cyber-attacks, and ultimately pinpoint the point of failure.

(Solgi et al., 2022) discusses the digital manufacturing (DM) paradigm that, in the context of Industry 4.0, may boost performance and enhance quality. But there are also cybersecurity vulnerabilities associated with DM that must be addressed, which evaluates the production implications, the dangers, and provides perspectives to safeguard the Industrial 4.0 system.

Any planned, political-driven attack on computer networks, programmes, and data that jeopardises or causes violence against banking, e-education, healthcare, industrial IoT (Solgi et al., 2022)., and national economies is commonly referred to as cyberterrorism. Often, the phrase is broadened to cover any cyberattack that causes fear or intimidation among the target population. Attackers frequently do this by destroying or impairing vital infrastructure.

More than 68 nations, including the USA, take part in the European Convention against Cyberterrorism. To stop cyberwarfare (Mohan, 2022) it aims to harmonise international regulations (Khater, 2023), boost investigative and detection capacities, and encourage global collaboration (Solgi et al., 2022).

The author (Gupta et al., 2021) presented a query selection strategy based on the most frequently occurring terms in the collected information (Weimann, 2016). Using the most commonly used terms in a query does not guarantee more unique results from the deep web database. (Tsuchiya & Hiramoto, 2021) introduced a greedy query selection strategy for estimating harvest rates. For the following query, the method selects the one with the greatest expected harvest rate. Each online database was represented as a distinct attribute-value network by (Squires, 2021) in a greedy link-based query selection technique designed for predicting the best values (Sonmez & Codal, 2022). (Faizan & Khan, 2019) used the acquired set-covering optimal sampling technique (Hayes et al., 2018) to tackle the problem.

They enlarged Ntoulas' (Rawat, 2023) approach to the complete form by providing a novel concept termed MEP (Minimum Executable Pattern) (Saharan et al., 2024). After creating an MEP set, viable keywords are selected based on the combined harvest rate, fingerprints, and pattern. The crawler achieves better results by selecting from a large number of MEPs (Robertson et al., 2016).

The hidden web was explained by the author (Alshammery & Aljuboori, 2022) using "the four types of invisibility." These categories originated for one of two reasons: either because search engine crawlers overlook some online content or because of the website's or sections' technological characteristics. Other research looked into the most hidden part of the invisible web, known as "Hidden Services," which is hosted by private networks.

The author (Heinl et al., 2019) offered a "human-aided technique" for hidden crawling (Lee et al., 2022) that involved processing search forms to acquire patterns inside hidden databases and submitting requests that included entries in hidden database local search interfaces, followed by crawling based on the findings.

The researchers (Nadini et al., 2022) began developing approaches for identifying features occurring in large quantities and at a rapid rate on the DW. The majority of research focused on international terrorist organisations. They suggested basic techniques for gathering data from DW sites (Lee et al., 2022), analysing it, and incorporating it into a knowledge management system. These methods begin with the identification of terrorist organisations (Lee et al., 2022) from reliable sources, followed by the construction of a preliminary list of websites, the extension of the list through link-based analysis, the addition of links to the list, and data collection from viewed pages (Nadini et al., 2022).

The author (Bergeron et al., 2022) focused on DW e-markets, particularly "Agora(Alshammery & Aljuboori, 2022)," which sold drugs, fake IDs, and documents. The crawler uses the LAMP Stack to imitate the authentication procedure for obtaining user credentials related to the e-market (traditional web development environment). By browsing relevant ”seed pages,” Rawat et al. (2022) constructed a targeted crawler to study themes related to homemade explosives (HME) lessons (videos and notes), as well as merchant trade channels and equipment used in explosives creation.

The author (Saleem et al.,2022) went into greater detail about the technological features of websites hosted on Tor hidden services, proposing a dark crawler technique based on OSN (social network analysis) (Alshammery & Aljuboori, 2022) with content selection strategies for gathering information from cyber terrorists (Howell et al., 2022) and extremist weblinks. The author (Alshammery & Aljuboori, 2022) worked on a crawler-based system that feeds data to a search engine and specialises in detecting data from dangerous and problematic websites. Its main operation starts with harvesting links to other Tor websites and extracting HTML files from a list of seed URLs (Holt & Lee, 2022). In order to recover information, the system analyses the pages after they’ve been preserved.

To categorise dangerous URLs, static analysis usually uses machine learning and pattern mining. (Iyer et al., 2022) demonstrated a prototype for detecting spam using lexical properties and host-based aspects of suspicious URLs.

A methodology for collecting malvertising postings and communications was presented by Madtracer (Collier, 2021). SpiderWeb (Howell et al., 2022) uses a crowd-sourced redirection chain approach to find good web connections. (Hayes et al., 2018) is a programme for tracking new viruses throughout a big ISP network and controlling domain names and lists. To identify cyber threat patterns, WebWitness (Bracci et al., 2021) tracks malware link download characteristics. The technique of identifying the regular properties of malicious code in JavaScript code snippets on web sites may be done statically (Heinl et al., 2019) or gradually by loading the influenced pages in a mimicked browser.

3. ONTOLOGY EXTRACTING DARKWEB MARKETPLACES (OEDWM)

Crawler (Khan et al., 2023) for Ontology Extracting DarkWeb Marketplaces (OEDWM), which deals with the context of details connected to a domain for cyber terrorism (Yadav et al.,2023) is the foundation of the semantic web. Crawling and extracting information (Panem et al., 2023) is a crucial source of data for knowledge-based systems. However, creating Crawler is a labour-intensive process that strongly depends on the developer’s skill. Here, a semi-automated method of invention in the field of terrorism for cyberspace is proposed. The intelligence gained from terrorist acts may be used to strengthen a nation’s security system. Online social network data (Khan et al., 2023), notably text data from Twitter, is obtained to get the most recent knowledge of the field. Concepts and related relationships are then identified and mapped using formal concept analysis. Numerous user-defined relationships are displayed by the fluent editing tool (FET) (Khan et al., 2023).

3.1 OEDWM for Cyberscape Terrorism’

Finding out the many uses for OEDWM is the first step towards building one. What criteria must be met for OEDWM composition, and what various conclusions may be drawn from them? For the purpose of illustrating their method for OEDWM creation, the algorithm created a crawling flow in the context of cyberterrorism. The work describes each step involved in creating an OEDWM. In order to create the OEDWM, a number of questions are needed to ascertain its needs.

- What does it imply when someone claims they were a victim of a cyber terrorism event, suffered abuse, or were forged?
- Is there a difference between cyberterrorism and crime?
- Who are the individuals involved in cyberterrorism and criminal activity?
- Do you have certain online platforms or channels in mind?
- Is Numerous incidents have been connected to criminal and terrorist activity.

Firstly, identify the essential idea of OEDWM development that relates to the real world in order to arrive at the answers to the questions above. Keywords like cyber terrorism, crime, violence, territory, fraud, propaganda, nation, religion, location, online channels, ransom, time, and language are essential concepts in the OEDWM. It is described how to locate ideas and relationships as well as how to map these associations between concepts using a semi-automated process. The full approach is shown in a process flow with the methodology and results in the terrorist domain. Figure 2 shows the constructed OEDWM.

3.2 Text Analysis and Collection

To identify concepts, relationships, examples, and fundamental information for OEDWM composition, Twitter text data is used. We made use of the Twitter API to obtain the Twitter data. Several R programming tools, including TwitterR and TM, are used to extract NLP (natural language processing) tweets. We took out 3725 tweets to create the corpus. Here is a collection of tweets that contain the hashtags “fraud,” “extremist groups,” “violence,” and “cyber_terror.”

- Deal with the dark web when you have questions about cyber warfare.”
- Market-related terrorist activities are being stopped. The use of violence through cyberterrorism
- Over 1,000 people have been killed in Mozambique by insurgents from extremist organisations.
- Russia was able to put a stop to the conflict, dismantle extremist organisations, and keep hacktivism’s cyberspace statehood intact.

4. LAYERS OF THE INTERNET

To go into the depths of the internet, we must first identify the different portions of the web, or hyperspace. The web is divided into three levels, according to researchers. Table 1 highlights about the web application structure as shown in Figure 3.

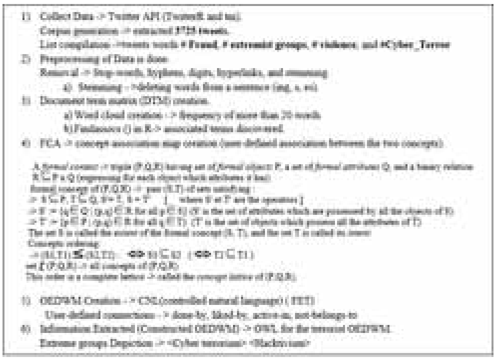
Ross William Ulbricht (Paul, 2018) created the most well-known e-market market (Silk Road) (Collier, 2021) and offers trade in drugs, malware, hacking services, money laundering, account forgery, stolen social and credit cards, multimedia pirated tools, forged passports (Howell et al., 2022), and social card ID (Heinl et al., 2019) on the DW, called ”Dread Pirate Roberts.” The FBI discovered and shut down the website in September 2013 (Ulbricht’s arrest in October). He was sentenced to life imprisonment (2015) after gaining more than 13.2 million dollars (Braatena & Vaughn, 2021) from Silk Road trades (commissions). Researchers work to understand the vision of DW activities, motivations, interests, participation offers, and operating environments using data and statistics to benefit specialists in a variety of fields.

Ross William Ulbricht (Paul, 2018) The DW, known as ”Dread Pirate Roberts (Howell et al., 2022),” created the most well-known e-market market (Silk Road) and trades in drugs, malware, hacking services, money laundering, account forgery, stolen social and credit cards, multimedia pirated tools, forged passports, and social card ID (Heinl et al., 2019). In September 2013, the FBI found and shut down the website (Ulbricht was arrested in October) (Alshammery & Aljuboori, 2022). After collecting more than 13.2 million dollars via Silk Road deals, he was sentenced to life in jail in 2015. (commissions). Researchers use facts and statistics to help professionals in a range of disciplines grasp the vision of DW activities, motivations, interests, participation opportunities, and operational settings, aid cyber security organisations (Alshammery & Aljuboori, 2022) in discovering security weaknesses and faults for anticipating cyber assaults before they occur, as well as support security authorities in obtaining evidence for investigations or prosecutions. Instead of using a central server, they typically use peer-to-peer systems, in which data is stored on a collection of personal computers distributed around the world over a network using the public internet’s infrastructure

Table 1. Web Application Structure

Internet Layers	Details
Surface Web.	Denotes a section of the internet that has been indexed by search engines (Mili & Rodin, 2022). (like Google). The Visible Web, Lightnet Lightnet(Alshammery & Aljuboori, 2022), Clear Web, and other names have been given to it. Web crawlers, which are little pieces of software that assist with indexing, are used by search engines. A crawler's principal task is to locate weblinks on the internet (the total number of indexed pages on the surface web is estimated to be around 4.2 billion pages) (Robertson et al., 2016). When a crawler studies a page, it looks for out-bound links to other sites and visits those pages while sending the information to a search engine, which indexes the data in the form of keywords that precisely describe the webpage. When a user performs standard searching (Iyer et al., 2022), the search engine compares the terms in the query to the index and extracts the relevant pages, after which the user receives the result.
Deep Web(Deepnet, Hidden Web, and Invisible Web).	Un-indexable webpages (Iyer et al., 2022), or those that are not linked to other pages on the Surface Web that are not searchable make up a chunk of the internet. Experts estimate that the Deep Web accounts for 96 percent (Collier, 2021) of the internet. A multitude of circumstances may contribute to the inability to index a webpage, including: ● To prevent crawlers (Howell et al., 2022) from accessing the webpage, the owner uses a password. ● Crawlers may be unable to reach the website after a certain number of access attempts, and the page may become unavailable. It's possible that the page isn't connected or hidden. ● The file (robots.txt) (Braatena & Vaughn, 2021) on the website advises crawlers not to crawl (Howell et al., 2022) that site or certain areas of it.
Darkweb	On this anonymous web space network, the majority of illegal operations take place, including human trafficking (Howell et al., 2022), drug trafficking, pornography (Holt & Lee, 2022), hacking, malware hosting (Iyer et al., 2022), security breaches, contract system exploitation, hitmen hiring (Collier, 2021), criminal idea promotion (Howell et al., 2022), credit card fraud, document forgery, organ trafficking, and so on. We discovered that academics defined the Deep Web and DW differently. The Deep Web, according to some research, refers to anything that search engines can't locate or index, whereas the DW refers to private networks that use particular protocols and software (Alshammery & Aljuboori, 2022). The levels of the internet are depicted in Figure 3.

Figure 2. Constructed OEDWM (Rawat, 2023)



1) Collect Data -> Twitter API (TwitteR and tm).
Corpus generation -> extracted 3725 tweets.
List compilation ->tweets words **# Fraud, # extremist groups, # violence, and #Cyber_Terror**

2) Preprocessing of Data is done
Removal -> Stop-words, hyphens, digits, hyperlinks, and stemming
a) Stemming -> deleting words from a sentence (eg, s, es)

3) Document term matrix (DTM) creation.
a) Word cloud creation -> frequency of more than 20 words
b) Findassocs () in R-> associated terms discovered.

4) FCA -> concept association map creation (user-defined association between the two concepts).

A *formal context* -> triple (P,Q,R) having set of *formal objects* P, a set of *formal attributes* Q, and a binary relation $R \subseteq P \times Q$ (expressing for each object which attributes it has).
formal concept of (P,Q,R) -> pair (S,T) of sets satisfying :
-> $S \subseteq P$, $T \subseteq Q$, $S' = T$, $S = T'$ [where S' et T' are the operators]
-> $S' = \{q \in Q \mid (p,q) \in R$ for all $p \in S\}$ (S' is the set of attributes which are possessed by all the objects of S)
-> $T' = \{p \in P \mid (p,q) \in R$ for all $q \in T\}$ (T' is the set of objects which possess all the attributes of T)
The set S is called the *extent* of the formal concept (S, T), and the set T is called its *intent*
Concepts ordering
-> $(S1,T1) \leq (S2,T2) \iff S1 \subseteq S2$ ($\iff T2 \subseteq T1$)
set £ (P,Q,R) -> all concepts of (P,Q,R)
This order is a complete lattice -> called the *concept lattice* of (P,Q,R).

5) OEDWM Creation -> CNL(controlled natural language) (FET)
User-defined connections -> done-by, liked-by, active-in, not-belongs-to

6) Information Extracted (Constructed OEDWM) -> OWL for the terrorist OEDWM.
Extreme groups Depiction -> <Cyber terrorism> <Hacktivism>

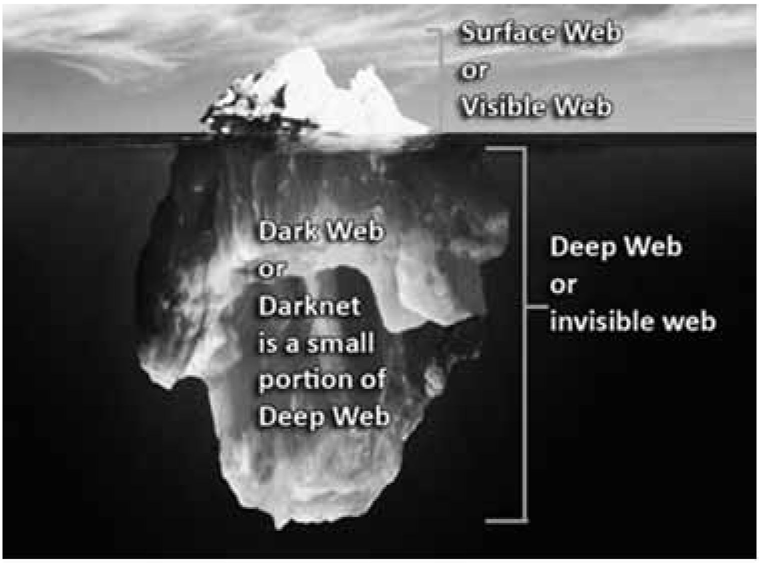
and non-standardised computer protocols (Braatena & Vaughn, 2021) and port services to keep its existence hidden from others.

5. THE ONION ROUTER (TOR) LAYOUT

Tor (The Onion Router) (Howell et al., 2022) is the most well-known anonymous software, along with I2P, Freenet, Hyperboria Network (Holt & Lee, 2022), M-Web, and Shadow Web (Nadini et al., 2022).

Activities on DW are socially dependent on a strong community structure for paying attention to members, and they require an administering body (controlling websites, traffic maintenance, marketing, and trust generation) (Bergeron et al., 2022) (Howell et al., 2022) to ensure the platform's security and anonymity while also allowing sellers and dealers to concentrate solely on their business (Braatena & Vaughn, 2021).

Figure 3. Layers of the Internet (Paul, 2018)



The Onion Router encrypts data using numerous levels of encryption before sending it over the open network, then removes layers of encryption (Howell et al., 2022) to send the server on a randomly determined path. The US Navy developed Tor (Holt & Lee, 2022) (Rawat et al., 2022) in the 1990s to encrypt US intelligence communications (Braatena & Vaughn, 2021) on the internet. They first made it available to the public in 2002, with the primary purpose of permitting open-source data collection (Braatena & Vaughn, 2021).

5.1 The Layered Methodology

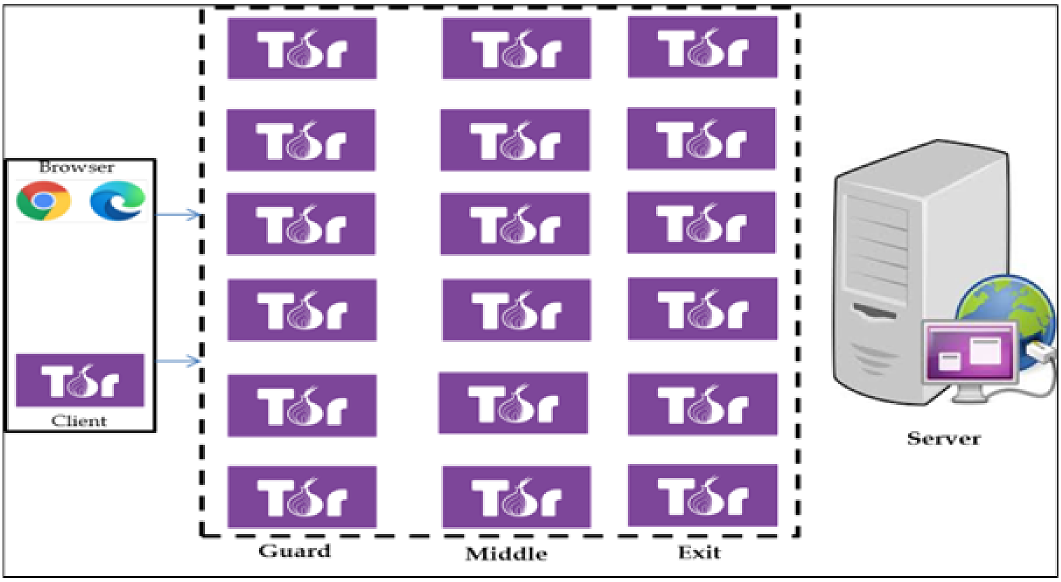
Tor has become the ideal venue for many websites to undertake illicit (Rawat et al., 2022) and destructive acts while keeping their users' anonymity, despite the initial goal of developing such a network. Silk Road (Bergeron et al., 2022) was a well-known electronic bazaar that was shut down by the FBI but revived after barely a month. Some of these websites are easily accessible by going to certain web pages. These sites can be used as dictionaries (similar to Hidden Wiki) (Howell et al., 2022) or to search the Tor network using specialised but limited search engines like Grams, DuckDuckGo(Howell et al., 2022), and TorSearch (Saleem et al., 2022), but only for limited anonymous services.

5.2 Crawled Data From DW – Examples

The author (Bracci et al., 2021) in Figure 4: Crawled Data from the Drug Market Highlighting Some Key Features of a Chloroquine Listing in the DarkBay/Dbay MPS (Bracci et al., 2021) were emphasised.

The author (Heinl et al., 2019) used a crawling algorithm (Mili & Rodin, 2022) (Tsuchiya & Hiramoto, 2021) to extract seller (Figure 5) and buyer (Figure 7)) advertisement on the DW platform for illegally selling organs (Gupta et al., 2021) (Squires, 2021) in the case of organ trafficking at the black market. Transplant tourism and contemporary slavery have become major concerns across the world. For people with end-stage renal disease, transplantation is regarded as the best option. In 2018, over 21,000 kidney transplants were performed (Squires, 2021) (Paul, 2018), and performed in the United States; nevertheless, there were around 100,000 transplant candidates on the waiting list.

Figure 4. Tor layered encryption (Iyer et al., 2022)



Around 10 percent of patients (Tsuchiya & Hiramoto, 2021) on the waiting list die or become too ill to be transplanted each year. Wealthy candidates, predictably, will look for alternatives. The crypto market is anticipated to account for 5 to 10 percent of the kidney transplant market. Between 2005 and 2013, the growth (Lee et al., 2022) of Internet technology shifted the trend of organ trafficking from local to worldwide.

From 2008 to 2015 (Alshammery & Aljuboori, 2022), the price of a transplantable organ (Iyer et al., 2022) on the DN quadrupled, reaching roughly 40,000 dollars (Robertson et al., 2016). Because such transplants require a high degree of patient care for both preoperative and postoperative treatment, they pose a significant potential risk for both donors and recipients. As a result, it is the obligation of all clinicians who work with transplant candidates to educate them about the dangers of organ ads on social media and the DW.

Tor networks are incognito shields for the Onion Routing Protocol (Howell et al., 2022), which uses several layers of multilayer encryption to transport data between consumers and administrators. Tor works in a unique way at the TCP transport layer (Heinl et al., 2019), depending on socket connections for communication. It builds a virtual circuit environment on the whole network using randomly selected nodes. It spends around 10 minutes (Howell et al., 2022) with virtual circuits before moving on to building a new circuit (for a new user), and so on.

Tor encrypts (Iyer et al., 2022) data using various layers of encryption, sends it to the entry node, and transports it throughout the whole network via a number of randomly selected intermediary nodes. With each jump to a node, the current accessible node removes one layer of encryption before delivering the encrypted request to the next intermediate node (Collier, 2021). When the request reaches the exit node, the remaining levels of encryption are removed from it, and it is forwarded (unencrypted) (Howell et al., 2022) to the original web server on the internet. All information about the origin is lost, identification is ambiguous, and only the final node in the circuit may be reached using this method (Howell et al., 2022). Tor also allows for the deployment of sites without identifying the location of hosting servers (anonymous browsing with secure communication among users), as well as sites with the ".onion" suffix (Braatena & Vaughn, 2021), which cannot be processed outside of the Tor network.

Table 2. ToR Structure and Terminologies (Alshammery & Aljuboori, 2022) (Holt & Lee, 2022)(Iyer et al., 2022) (Collier, 2021)

Data	Structure
Tor Layered Methodology.	● Tor client:- a Tor web-net client. ● Server: - This is the target web server. ● Tor (onion) router:- A Tor web-special net's intermediate junction. ● Directory server:-A server that provides a Tor web-net junction.
Three Nodes in the Circuit.	● Customer entry junction - Tor circuit that processes incoming data and transfers it to the intermediate levels. ● Middle junction - Data is sent through intermediary nodes to several nodes. ● Exit junction - The last junction receives data from the centre junction, decrypts the final layer of encryption, and exits into the open environment. .
Crawler	According to Heinl et al. (2019), crawlers (Bracci et al., 2021) are "software programmes that navigate the internet information space by following hypertext links and acquiring web material using the standard HTTP protocol." ● Crawlers are employed in a wide range of applications and research domains, including search engines that require updated data, keeping a copy of every visited page for later processing through search engines, and indexing webpages (Heinl et al., 2019) quickly and easily when a user searches for a certain topic. Crawlers gather and archive huge groups of pages on a regular basis for future use as online monitoring services, enabling users to input inquiries and serving as triggers alerting crawlers to continually search the web and present related new sites (Lee et al., 2022). ● Crawlers are also used by web administrators to maintain a website automatically, such as by checking hyperlinks (Alshammery & Aljuboori, 2022) and validating HTML parts, or to gather certain types of data, such as email addresses and particularly harmful or spam emails (Howell et al., 2022). ● The main problem is scalability (Bergeron et al., 2022), which may be solved by creating a single repository for storing webpages for a range of calculations. starts with the development of a URL database structure, then retrieves content from the specified links, updates the repository with new links, and so on. Researchers (Howell et al., 2022) refer to this as "crawling (Rawat et al., 2022)" or "spidering (AlKhatib & Basheer, 2019). ● In the recent two decades, there has been a boom in interest in crawling software development in the DW, but owing to the technological peculiarities of that segment, creating it requires extra strategies combining for crawlers to detect harmful websites and access them for reporting. ● Proxy software (such as Privoxy) (Bergeron et al., 2022)(Rawat et al., 2022)(Howell et al., 2022) (Collier, 2021) is used to make HTTP proxy connections (Wang et al., 2021) without maintaining any data cache about the current connection. The crawler is connected to the Tor network via this proxy (Iyer et al., 2022).
Challenges	A theoretically easy crawler duty is to download all pages (seed URLs) under the supplied ids, extract links from the pages and associate them with the list of addresses, and constantly crawl the resulting links. Despite its apparent simplicity, web crawling (Rawat et al., 2022) has a variety of challenges, the most prominent of which are: ● Due to the large size of the web and its constant growth, crawling approaches must solve the problem of operating the crawler (Rawat et al., 2022) on several devices by separating the URL data (Howell et al., 2022) and assigning each device to a subset of URLs. ● Websites hosted on a private, encrypted network (Holt & Lee, 2022) have a shorter lifespan than those hosted on the public internet since they move many addresses, making operation time (Iyer et al., 2022) unreliable. Furthermore, online administrators employ a variety of web domains to distribute webpages, notably at DW Electronic Markets (Paul, 2018), to circumvent surveillance. ● Systems that employ encrypted networks have technological obstacles such as bandwidth constraints, which cause Tor-hosted (Paul, 2018) websites to load more slowly than websites that use direct network (Bracci et al., 2021) connections. ● The bulk of websites require human input (Wang et al., 2021). The regulations of their community must be registered and approved. Frequently, the registration and login processes are complex. To avoid automated logins (Mili & Rodin, 2022) and DDoS assaults (Mili & Rodin, 2022), users must complete CAPTCHA (Howell et al., 2022), graphical puzzles, and other similar challenge quizzes. ● Web administrators keep an eye on the professionalism and efficiency of the electronic (Gupta et al., 2021) community in which they work. This might entail creating a social layering system depending on how involved their members are, as well as their abilities and professional standing. A technique is employed to deactivate the accounts of unused and inactive users to discourage clandestine surfing habits (Tsuchiya & Hiramoto, 2021; Squires, 2021). .
Crawling and Data Extraction	TOR (Heinl et al., 2019) is a well-known programme that allows for anonymous communication and is getting more popular as DW sites become more prominent. Because the web servers are hidden on the TOR network (Tsuchiya & Hiramoto, 2021) and need unique protocols to be accessed, "DW" sites are seldom examined by standard crawlers. Domain addresses under the top-level domain .onion (Squires, 2021) are used to visit sites hidden on the TOR network (Heinl et al., 2019). ● External HTTP proxies, such as Privoxy(Sonmez & Codal, 2022)configured to route traffic over the TOR network, are used by the crawler to scan such sites. We just need to configure route requests to onion addresses (Faizan & Khan, 2019) via the TOR proxy after establishing the proxy (Hayes et al., 2018). ● One of the most well-known onion names is facebookcorewwwi.onion (Paul, 2018), which refers to a hidden service on the TOR network that hosts an instance of Facebook's website. Users must utilise the TOR browser, which is a customised version of Mozilla Firefox (Robertson et al., 2016), to gain access to it. When using the TOR browser, the .onion name may be entered into the address bar just like any other URL. ● Users (through the TOR browser) (Bracci et al., 2021) and content producers (via TOR hidden services) (Lee et al., 2022) may both behave anonymously using TOR.
Tools	● By relocating the IP address, the author (Heinl et al., 2019) developed a system,Darky, using Scrapy (a Python programming library) (Weimann, 2016)(Tsuchiya & Hiramoto, 2021), as we provided it with a connection to DWsites on the Tor network through Tor software integrated with Privoxy (a software for Virtual Private Networks (VPN)) (Squires, 2021) to ensure the crawler's security and anonymity against those sites. ● Following the establishment of the Tor-Privoxy connection (Faizan & Khan, 2019), the crawler is started from the website URL (Hayes et al., 2018), and the login interface is processed using the credentials that we generated previously on the website for this reason. ● A crawler must be aware of the properties of the crawled network when developing it. Proxy software (such as Privoxy) (Rawat, 2023) should be used to offer a proxy connection on the HTTP protocol (Bracci et al., 2021) without keeping any data cache about the presently arising connection, and this proxy links the crawler to the Tor network anonymously (Heinl et al., 2019). ● Scrapy (https://scrapy.org/) (Alshammery & Aljuboori, 2022) is a high-level web crawling and scraping framework (Heinl et al., 2019) that is used to crawl websites and extract structured data from their pages. It has a wide range of applications, including data mining, monitoring, and automated testing. ● Privoxy (https://www.privoxy.org/) (Bracci et al., 2021) is a free non-caching web proxy with filtering features for boosting privacy, managing cookies, and changing web page data and HTTP headers before the page is viewed by the browser (Heinl et al., 2019). ● Privoxy is a "privacy-enhancing proxy" that filters on-line pages and blocks ads. Users may modify Privoxy for single-user computers as well as multi-user networks and Privoxy may be linked.

Figure 5. Examples of Crawled Data (Bracci et al., 2021) (Heinl et al., 2019)

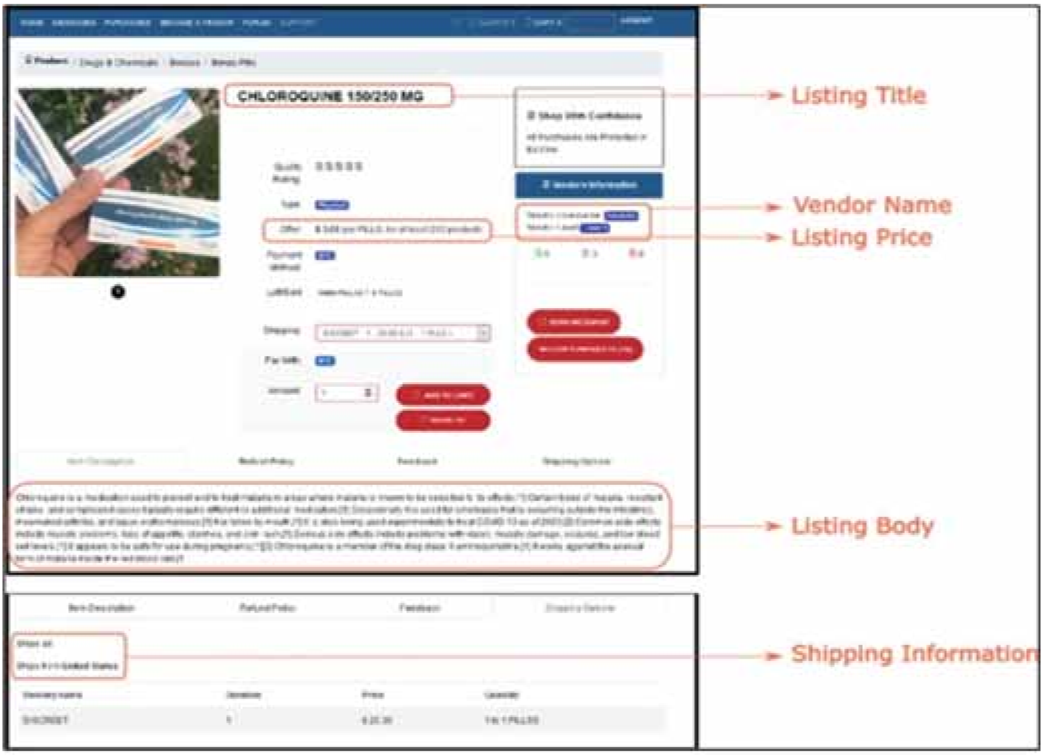
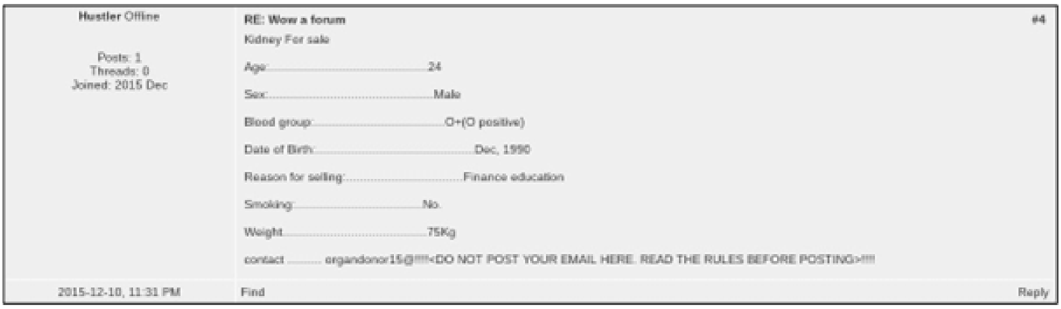


Figure 6. Seller's advertisement on the DW forum Moneybook (Heinl et al., 2019) (Rawat et al., 2022)



Hustler Offline

Posts: 1
Threads: 0
Joined: 2015 Dec

RE: Wow a forum #4

Kidney For sale

Age...24

Sex..Male

Blood group:...O+(O positive)

Date of Birth:...Dec, 1990

Reason for selling:...................................Finance education

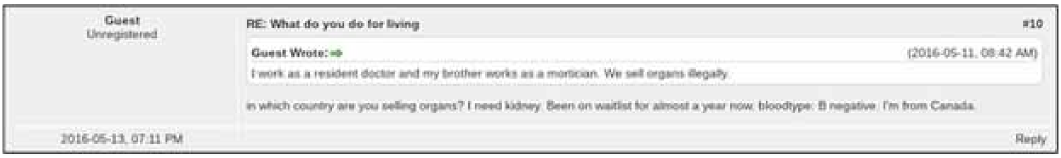
Smoking:.......................................No.

Weight...75Kg

contact .... organdonor15@!!!!<DO NOT POST YOUR EMAIL HERE. READ THE RULES BEFORE POSTING>!!!!

2015-12-10, 11:31 PM Find Reply

Figure 7. Buyer's advertisement on the DW forum Moneybook (Heinl et al., 2019) (Howell et al., 2022)



Guest
Unregistered

RE: What do you do for living #10

Guest Wrote: ⇨ (2016-05-11, 08:42 AM)

I work as a resident doctor and my brother works as a mortician. We sell organs illegally

in which country are you selling organs? I need kidney. Been on waitlist for almost a year now. bloodtype: B negative. I'm from Canada.

2016-05-13, 07:11 PM Reply

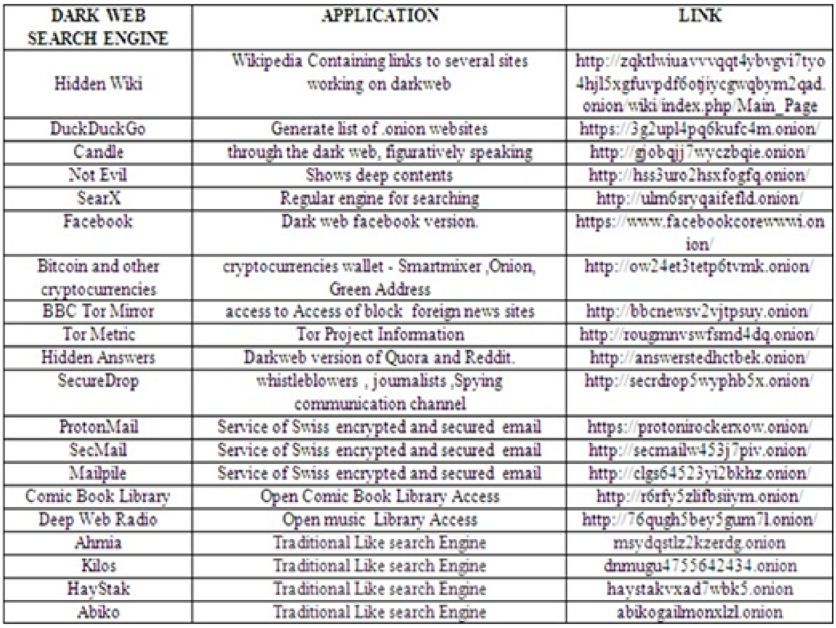
The Figure 8 shows about the DW platform Tools and Applications(Saharan et al., 2024) and the Table 3 shows DW Anonymous markets and Crime (Bracci et al., 2021).

Following Silk Road's paradigm (Weimann, 2016) (Tsuchiya & Hiramoto, 2021)(Squires, 2021), modern MPS are defined by the use of DN anonymous access (usually Tor), Bitcoin or Monero payment with escrow services, and eBay-like vendor feedback systems. The first open cyber-arms market for software vulnerabilities and pharmaceuticals, TheRealDeal(Weimann, 2016) (Sonmez & Codal, 2022), was introduced to the delight of computer security professionals. DDOS (Alshammery & Alju-Boori, 2022) assaults were launched against many markets in May, including The Real Deal

Table 3. DW Anonymous markets and Crime (Bracci et al., 2021)(Howell et al., 2022)(Howell et al., 2022)

Keywords	Use by Organization – Category
Dark-net, Dark-web, Hidden web, Deep-web, Violence, war on drugs, drug trafficking, Drug Market, DN markets, clandestine market, Grey market, cannabis, modafinil, LSD, cocaine, designer drugs, MDMA, Temazepam, Illegal logging, Silk Road, black markets.	Drug Trafficking crime
unreported economy, Online extremist recruitment, Anonymous network, Bitcoin Fraud, Intelligence agency, CIA, Interpol, United Nations Office on Drugs and Crime, terrorism, cyber threat, crowd funded assassinations, Social Network propaganda	Cyber Terrorism
Cyber crime, cyber attack, Darkweb crypto-markets, Hidden ToR Market	Organ Trafficking
illegal drug economy, underground economy, shadow economy, tax gap, Online Social Network (OSN) Illicit Business	Financial Crime
Criminal, Illicit Drugs, smuggling, Arms trafficking, Biological organs Trading, illegal drug trade	Weapon Trafficking
Child pornography, Human Trafficking, organized crime	Human Trafficking and Pornography

Figure 8. DW platform Tools and Applications (Howell et al., 2022) (Braatena & Vaughn, 2021)



DARK WEB SEARCH ENGINE	APPLICATION	LINK
Hidden Wiki	Wikipedia Containing links to several sites working on darkweb	http://zqktlwiuavvvqqt4ybvgvi7tyo4hjl5xgfuvpdf6otjiycgwqbym2qad.onion/wiki/index.php/Main_Page
DuckDuckGo	Generate list of .onion websites	https://3g2upl4pq6kufc4m.onion/
Candle	through the dark web, figuratively speaking	http://gjobqjj7wyczbqie.onion/
Not Evil	Shows deep contents	http://hss3uro2hsxfogfq.onion/
SearX	Regular engine for searching	http://ulm6sryqaifefld.onion/
Facebook	Dark web facebook version.	https://www.facebookcorewwwi.onion/
Bitcoin and other cryptocurrencies	cryptocurrencies wallet - Smartmixer ,Onion, Green Address	http://ow24et3tetp6tvmk.onion/
BBC Tor Mirror	access to Access of block foreign news sites	http://bbcnewsv2vjtpsuy.onion/
Tor Metric	Tor Project Infomation	http://rougmnvswfsmd4dq.onion/
Hidden Answers	Darkweb version of Quora and Reddit.	http://answerstedhctbek.onion/
SecureDrop	whistleblowers , joumalists ,Spying communication channel	http://secrdrop5wyphb5x.onion/
ProtonMail	Service of Swiss encrypted and secured email	https://protonirockerxow.onion/
SecMail	Service of Swiss encrypted and secured email	http://secmailw453j7piv.onion/
Mailpile	Service of Swiss encrypted and secured email	http://clgs64523yi2bkhz.onion/
Comic Book Library	Open Comic Book Library Access	http://r6rfy5zlifbsiiym.onion/
Deep Web Radio	Open music Library Access	http://76qugh5bey5gum7l.onion/
Ahmia	Traditional Like search Engine	msydqstlz2kzerdg.onion
Kilos	Traditional Like search Engine	dnmugu4755642434.onion
HayStak	Traditional Like search Engine	haystakvxad7wbk5.onion
Abiko	Traditional Like search Engine	abikogailmonxlzl.onion

Table 4. DW Crime Statistics (Lee et al., 2022) (Nadini et al., 2022) (Bergeron et al., 2022)

Reported Crime	Crime Statistics
Chainalysis- cryptocurrency crime.	1.7 Dollar billion fraud bitcoin in cryptocurrency transaction.
DW Price Index- stolen PayPal account details.	Stolen credit card data sells for just 25 Dollar.
SpyCloud- Passwords breaches	breached assets of Fortune 1000 companies increased by 35 Dollar.
illegal drugs and other illicit goods	9.5 million sales in Bitcoin.
Secure World- Hydra - largest DN market for illicit events	1.5 billion Dollar MPS revenue generation.
Identity Theft Resource Center- ransomware-related data breaches	largest increase by 217 Dollar for illegal data purchase.
Chainalysis- Russia -biggest market for dark net spending	around 115 million Dollar illegal cryptocurrency sent to DN vendors.
PT Security- DW forums are from buyers to contact a criminal	90 percent of posts relating for criminal activities.
RAND- weapons sold on DW.	60 Percent weapons market.
Wilson Center-50,000 extremist groups on the DW	terrorist activities for recruitment .

(Iyer et al., 2022). The market owners put up a phishing website to obtain the attacker's password, which indicated collusion between the attacker and Mr. Nice Guy's (Howell et al., 2022) market administrator, who was also attempting to swindle his customers. DeepDotWeb (Collier, 2021), a news site, received this information. The darkweb markets are listed as (Silk Road, Black Market Reloaded,Sheep,DeepBay, Agora, Pandora, Evolution, TOM, Middle Earth, Nucleus, Abraxas, Black Bank, Alpha Bay) (Bracci et al., 2021) (Howell et al., 2022) (Cole et al., 2021).

An integrated framework based on machine learning provides a holistic view of the cyber-threat intelligence (Fu & Li, 2021) process, allowing security analysts to easily identify, collect, analyze, extract, integrate, and share cyber-threat intelligence from a variety of online sources, such as clear/deep/DW sites, forums (Mili & Rodin, 2022), and market places. Crawling (Gupta et al., 2021) enables customers to quickly build up and deploy automated data gathering crawlers that can explore the open, social, and DWs to find and harvest relevant material. The Crawling submodule (Rawat et al., 2022) allows the user to choose from a variety of options, including focused/topical crawling guided by appropriate machine learning methods, domain downloads based on powerful yet simple to set up in-depth crawlers, TOR-based DW spidering (Squires, 2021), and semi-automated handling of authentication methods based on cookie management. Table 4 shows the DW Crime Statistics (Lee et al., 2022) (Nadini et al., 2022) (Bergeron et al., 2022).

6. CONCLUSION AND FUTURE WORK

The work describes the methods for crawling DW information from marketplaces. Tor encrypts data using various layers of encryption, sends it to the entry node, and transports it throughout the whole network via a number of randomly selected intermediary nodes. With each jump to a node, the current accessible node removes one layer of encryption before delivering the encrypted request to the next intermediate node. When the request reaches the exit node, the remaining levels of encryption are removed from it, and it is forwarded (unencrypted) to the original web server on the internet. All information about the origin is lost, identification is ambiguous, and only the final node in the circuit may be reached using this method (Howell et al., 2022). Tor also allows for the deployment of sites without identifying the location of hosting servers (anonymous browsing with secure communication among users), as well as sites with the “onion” suffix, which cannot be processed outside of the Tor network. We described how to apply the crawling framework and procedures to hidden websites, as well as how to extract important information to help security agencies and law enforcement (LE) authorities investigate DW operations. We discussed a crawler that could imitate a user’s credentials on a black market site, browse the site’s comprehensive contents, and retrieve the information needed from its connected links and pages. The created data will be modified in the future for data mining procedures to be crawled into the system.

CONFLICTS OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

FUNDING STATEMENT

No funding was received for this work.

PROCESS DATES

Received: 8/29/2023, Revisions: 1/23/2024, Accepted: 3/21/2024

REFERENCES

Ahmad, M., Al-Amri, J. F., Subahi, A. F., Khatri, S., Seh, A. H., Nadeem, M., & Agrawal, A. (2022). Healthcare device security assessment through computational methodology. *Computer Systems Science and Engineering*, *41*(2), 811–828. doi:10.32604/csse.2022.020097

AlKhatib, B., & Basheer, R. (2019). Crawling the dark web: A conceptual perspective, challenges and implementation. *J. Digit. Inf. Manag.*, *17*(2), 51. doi:10.6025/jdim/2019/17/2/51-60

Alshammery, M. K., & Aljuboori, A. F. (2022). Crawling and mining the dark web: A survey on existing and new approaches. *Iraqi Journal of Science*, ●●●, 1339–1348. doi:10.24996/ijs.2022.63.3.36

Bergeron, A., D'ecary-H'etu, D., & Ouellet, M. (2022). Conflict and victimization in online drug markets. *Victims & Offenders*, *17*(3), 350–371. doi:10.1080/15564886.2021.1943090

Braatena, C. N., & Vaughn, M. S. (2021). Convenience theory of cryptocurrency crime: A content analysis. *Deviant Behavior*, *42*(8), 958–978. doi:10.1080/01639625.2019.1706706

Bracci, A., Nadini, M., Aliapoulios, M., McCoy, D., Gray, I., Teytelboym, A., Gallo, A., & Baronchelli, A. (2021). Dark web marketplaces and covid-19: Before the vaccine. *EPJ Data Science*, *10*(1), 6. doi:10.1140/epjds/s13688-021-00259-w PMID:33500876

Cole, R., Latif, S., & Chowdhury, M. M. (2021). *Dark web: A facilitator of crime. In 2021 international conference on electrical, computer, communications and mechatronics engineering (iceccme).*

Collier, B. (2021). Infrastructural power: dealing with abuse, crime, and control in the tor anonymity network. In *Cybercrime in context* (pp. 283–301). Springer. doi:10.1007/978-3-030-60527-8_16

Ejazi, F. (2022). Performance of Virtual Terrorism in Cyber Space. In *Media and Terrorism in the 21st Century* (pp. 224–236). IGI Global. doi:10.4018/978-1-7998-9755-2.ch013

Faizan, M., & Khan, R. A. (2019). Exploring and analyzing the dark web: A new alchemy. *First Monday*. Advance online publication. doi:10.5210/fm.v24i5.9473

Fu, R., & Li, X. (2021). Malicious attacks on the web and crawling of information data by python technology. *Security and Privacy*, *4*(5), e173. doi:10.1002/spy2.173

Hayes, D. R., Cappa, F., & Cardon, J. (2018). A framework for more effective dark web marketplace investigations. *Information (Basel)*, *9*(8), 186. doi:10.3390/info9080186

Heinl, M. P., Yu, B., & Wijesekera, D. (2019). A framework to reveal clandestine organ trafficking in the dark web and beyond. *Journal of Digital Forensics*. *Security and Law*, *14*(1), 2.

Holt, T. J., & Lee, J. R. (2022). A crime script model of dark web firearms purchasing. *American Journal of Criminal Justice*, ●●●, 1–21.

Howell, C. J., Maimon, D., Perkins, R. C., Burruss, G. W., Ouellet, M., & Wu, Y. (2022). Risk avoidance behavior on darknet marketplaces. *Crime and Delinquency*, ●●●, 00111287221092713.

Iyer, S., Rajeswari, G. R., & Jayakrishnan, B. et al.. (2022). The new deep web drug marketplace (and the role of bitcoin as a currency for drugs). *Estudios de Economía Aplicada*, *40*(S1). Advance online publication. doi:10.25115/eea.v40iS1.5471

Khan, Z. C., Mkhwanazi, T., & Masango, M. (2023, August). A Model for Cyber Threat Intelligence for Organisations. In *2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (pp. 1-7). IEEE. doi:10.1109/icABCD59051.2023.10220503

Khater, M. H. (2023). International Perspective on Securing Cyberspace Against Terrorist Acts. [IJSKD]. *International Journal of Sociotechnology and Knowledge Development*, *15*(1), 1–11. doi:10.4018/IJSKD.318706

Lee, J. R., Holt, T. J., & Smirnova, O. (2022). An assessment of the state of firearm sales on the dark web. *Journal of Crime and Justice*, ●●●, 1–15.

Mili, S. E., & Rodin, V. (2022). Wot search engine based on multi agent system: A conceptual framework. *International Journal of Interactive Mobile Technologies*, *16*(5).

Mohan, V. (2022). Cyberspace Based Cross-Border Terrorism: An Overview of Global and Indian Legal Regime. *Indian JL & Just.*, *13*, 273.

Nadini, M., Bracci, A., ElBahrawy, A., Gradwell, P., Teytelboym, A., & Baronchelli, A. (2022). Emergence and structure of decentralised trade networks around dark web marketplaces. *Scientific Reports*, *12*(1), 1–9. doi:10.1038/s41598-022-07492-x PMID:35361797

Panem, C., Gundu, S. R., & Vijaylaxmi, J. (2023). The Role of Machine Learning and Artificial Intelligence in Detecting the Malicious Use of Cyber Space. *Robotic Process Automation*, 19-32.

Paul, K. A. (2018). Ancient artifacts vs. digital artifacts: New tools for unmasking the sale of illicit antiquities on the dark web. In Arts (Vol. 7, p. 12).

Rawat, R. (2023). Logical concept mapping and social media analytics relating to cyber criminal activities for ontology creation. *International Journal of Information Technology : an Official Journal of Bharati Vidyapeeth's Institute of Computer Applications and Management*, *15*(2), 893–903. doi:10.1007/s41870-022-00934-9

Rawat, R., Garg, B., Pachlasiya, K., Mahor, V., Telang, S., Chouhan, M., Shukla, S. K., & Mishra, R. (2022). Scnta: Monitoring of network availability and activity for identification of anomalies using machine learning approaches. [IJITWE]. *International Journal of Information Technology and Web Engineering*, *17*(1), 1–19. doi:10.4018/IJITWE.297971

Robertson, J., Diab, A., Marin, E., Nunes, E., Paliath, V., Shakarian, J., & Shakarian, P. (2016). Darknet mining and game theory for enhanced cyber threat intelligence. *The Cyber Defense Review*, *1*(2), 95–122.

Saharan, S., Singh, S., Bhandari, A. K., & Yadav, B. (2024). The Future of Cyber-Crimes and Cyber War in the Metaverse. In *Forecasting Cyber Crimes in the Age of the Metaverse* (pp. 126–148). IGI Global.

Saleem, J., Islam, R., & Kabir, M. A. (2022). The anonymity of the dark web: A survey. *IEEE Access : Practical Innovations, Open Solutions*, *10*, 33628–33660. doi:10.1109/ACCESS.2022.3161547

Solgi, R., Khodaverdi, H., & Poustinchi, Z. (2022). Pathology of the New Cyber Terrorism Threat to Iran's National Security. *International Journal of Poultry Science*, *12*(1), 61–80.

Sonmez, E. D. A., & Seçkin Codal, K. (2022). Terrorism in cyberspace: A critical review of dark web studies under the terrorism landscape. *Sakarya University Journal of Computer and Information Sciences*, (5).

Squires, P. (2021). Illegal firearms, illicit markets and weapon trafficking. In *Firearms* (pp. 51–71). Routledge. doi:10.4324/9780429316951-4

Strang, K. D., Korstanje, M. E., & Vajjhala, N. (Eds.). (2018). *Research, practices, and innovations in global risk and contingency management*. IGI Global. doi:10.4018/978-1-5225-4754-9

Strang, K. D., & Vajjhala, N. R. (2023). Why Cyberattacks Disrupt Society and How to Mitigate Risk. In *Cybersecurity for Decision Makers* (pp. 1–28). CRC Press. doi:10.1201/9781003319887-1

Topor, L., & Pollack, M. (2022). Fake identities in social cyberspace: From escapism to terrorism. [IJCWT]. *International Journal of Cyber Warfare & Terrorism*, *12*(1), 1–17. doi:10.4018/IJCWT.295867

Tsuchiya, Y., & Hiramoto, N. (2021). Dark web in the dark: Investigating when transactions take place on cryptomarkets. *Forensic Science International Digital Investigation*, *36*, 301093. doi:10.1016/j.fsidi.2020.301093

Vajjhala, N. R., & Strang, K. D. (Eds.). (2022). *Global Risk and Contingency Management Research in Times of Crisis*. IGI Global. doi:10.4018/978-1-6684-5279-0

Vajjhala, N. R., & Strang, K. D. (Eds.). (2023). *Cybersecurity for Decision Makers*. CRC Press. doi:10.1201/9781003319887

Wang, D., Zhang, Q., & Hong, S. (2021). Research on crawling network information data with scrapy framework. *International Journal of Network Security*, *23*(2), 326–331.

Weimann, G. (2016). Going dark: Terrorism on the dark web. *Studies in Conflict and Terrorism*, *39*(3), 195–206. doi:10.1080/1057610X.2015.1119546

Yadav, A., Kumar, A., & Singh, V. (2023). Open-source intelligence: A comprehensive review of the current state, applications and future perspectives in cyber security. *Artificial Intelligence Review*, *56*(11), 1–32. doi:10.1007/s10462-023-10454-y PMID:37362900

Romil Rawat is a Research scholar and faculty, He attended several research programs and received research grants from USA, Germany, Italy and UK. The Author has research alignment towards Cyber Security, IoT, Dark Web Crime analysis and investigation techniques, and working towards tracing illicit anonymous contents of cyber terrorism and criminal activities. He also chaired International Conferences and Hosted several research events including National and International Research Schools, PhD colloquium, Workshops, training programs. He also published several Research Patents.

Anand Rajavat is Deanacademic of Shri Vaishnav Vidyapeeth Vishwavidalaya and Professor & Director of Shri Vaishnav Institute of Information Technology of Shri Vaishnav Vidyapeeth Vishwavidyalaya (Indore). He has been engaged in teaching, research, training and consultancy for the last 22 years. He obtained Ph.D. in the faculty of Computer Engineering in the field of Software Engineering and M. E.in Computer Engineering with specialization in Software Engineering from Devi Ahilya Vishwavidyalaya (DAVV) Indore. Currently he has been also working as Visiting Faculty in St. Cloud State University (SCSU), Minnesota, USA. Besides undertaking consultancy assignments, he has authored/ co-authored more than 110publications. He has been on the Panel of Reviewers of number of National/International Journals. He has supervised Twenty-Eight (32) M. E. / M. Tech. Dissertations. He has attended more than 102 Certification Course/Training/Workshops and faculty development programs conducted by leading multinational and national IT companies. Dr. Anand Rajavat was awarded with Excellence in Academics Award in 2021 and Dronacharya Award by IBM in 2008 and 2009 for development and deployment of best research project. Dr. Anand Rajavat has been associated with various professional bodies like Computer Society of India, Indian Society of Internet, ACM and IEEE. He has been the Chairman of Board of Studies of Computer Science &Engineering, Information Technology and Computer Applications of Shri Vaishnav Vidyapeeth Vishwavidyalaya (Indore). He had been the member of Governing Body of Shri Vaishnav Institute of Technology & Science, Indore.