

Review of Applications to assist Legal issues related to Cyber Crimes

Dr. REKHA B VENKATAPUR¹, GANASHREE K², DHANYA SHARANYA SHREE R³, GANASHREE G M⁴

Department of Computer Science and Engineering, K S Institute of Technology, Bengaluru, India¹⁻⁴

Abstract: Cybercrime is becoming an increasingly widespread issue in India, leaving individuals struggling to navigate the complex legal system surrounding these violations. To address this problem, a Cyber Crime Chatbot Generator for Indian Law has been developed. This innovative technology aims to bridge the legal knowledge gap by providing an easy-to-use, readily accessible educational resource for dealing with cybercrime-related problems.

The chatbot generator is built on a robust platform that provides detailed guidance on recognizing, understanding, and addressing cybercrime concerns under Indian legislation. The system promotes transparency and trust in the legal system by facilitating early consultations, connecting users with expert attorneys, and providing ongoing case monitoring. The system offers a wealth of legal resources, articles, and frequently asked questions on cybercrime to enhance users' knowledge and understanding.

The system also encourages customers to share their experiences through a feedback mechanism, which helps maintain high service standards and continuously improve the system. In addition to empowering individuals, this innovative approach aims to enhance the effectiveness and efficiency of India's legal system in combating cybercrime and promote a more equitable and knowledgeable society. The Cyber Crime Chatbot Generator for Indian Law is poised to have a significant impact on the fight against cybercrime in India by utilizing technology to bridge the legal knowledge gap.

Keywords: Agent Communication Language (ACL), Ontology, Ortho Matcher, RP-score, Mean Reciprocal Rank.

I. INTRODUCTION

In today's interconnected world, the digital environment has become a fertile ground for various cybercrimes and malicious activities. Cybercrime encompasses a wide range of illegal activities in which criminals use computers and networks either as the primary means or as a method to achieve their objectives. This pervasive threat has raised numerous issues, demanding scrutiny and effective solutions.

Cybercrime, also known as computer crime, refers to any criminal activity that involves a computer and a network, which can be either the victim or the tool used in the crime.

- Internet fraud: The use of software or services to deceive or take advantage of people, often carried out through fraudulent online scams.
- Bank fraud: The unlawful acquisition of money, assets, or property belonging to a financial institution through potentially illegal means.
- Extortion: Criminal activities involving the use of force or threats to obtain goods, money, or services from individuals or organizations. The technique of "pharming" involves exploiting DNS vulnerabilities to redirect visitors from trustworthy websites to fake ones.
- Phishing: The illicit acquisition of private information, including credit card numbers, usernames, and passwords, using deceptive online methods.
- Cyberterrorism: Cyberterrorism, which often involves technologies like computer viruses, entails intentionally disrupting computer networks, particularly those involving internet-connected personal computers, for terrorist purposes.
- Electronic spamming: Spamming online with unwanted messages, especially advertisements using electronic messaging services and sometimes on the same website.
- Cyberbullying: The deliberate and persistent use of information technology networks to harass or harm others.
- Internet Homicide: Cases in which the perpetrator and victim only knew each other online before the murder or met online.
- Cheque Kiting: The deceptive practice of writing checks from one bank to another with the intention of exploiting the time gap during the clearance process. It constitutes fraudulent activity.

- Card Skimming: Illegally capturing data from a payment card's magnetic stripe for fraudulent use.
- Identity Theft: The unlawful use of another person's personal information for fraudulent or other illegal purposes.
- Crypto virology: The study of how cryptography is utilized in malware and ransomware attacks.
- Call Tag Scam: This type of phone scam involves criminals impersonating reputable companies or authoritative figures to deceive victims into divulging their personal information or money.
- Pharming: The practice of redirecting visitors from trustworthy websites to dubious ones.
- Sick Baby Hoax: A scam in which criminals pretend to be in dire need, claiming to have a sick child and seeking money or charity.
- Child Pornography: The illegal and immoral creation, distribution, or possession of pornographic content featuring children.
- Online Predators: Individuals who use the internet to sexually groom and exploit minors.
- Cyberstalking: The unwelcome and ongoing harassment and stalking of individuals online.
- Internet Trolls: Those who deliberately initiate online arguments or disputes for their amusement.
- Drug Trafficking: The unlawful distribution, sale, or transportation of narcotics, sometimes facilitated through Internet networks.
- Hacking: Illicit access to computer networks or systems without authorization, typically to steal confidential data or information.
- Dumpster Diving: Retrieving discarded devices or documents from the trash to gain unauthorized access to private information.
- Spoofing: A common tactic in phishing and other cybercrimes, spoofing involves altering data or identity to appear as something or someone else.
- Salami Theft: A type of financial fraud where victims often remain unaware of small amounts of money being taken from multiple accounts.
- Hijacking: Taking control of a digital system or communication channel without authorization.
- Warez: Unauthorized distribution of software, often involving pirated or cracked versions of commercial software.
- Mucking: The technique of disguising addresses to protect them from spam.
- Bluebugging: Unauthorized access to a mobile device to make calls, send messages, or access data.
- Click Fraud: Use of fake or fraudulent click-to-tone advertisements to generate income.
- Crap Flooding: Inundating internet forums and chat rooms with offensive or irrelevant information.
- Email Spoofing: The act of altering sender addresses to deceive recipients.
- Phreaking: The illegal use of phone networks or systems for purposes such as making free calls or engaging in other criminal activities.

These terms describe various cybercrimes and unethical practices, emphasizing the significance of cybersecurity and law enforcement efforts to combat them.

II. LITERATURE SURVEY

2.1 LawBot:

LawBot is designed and implemented to provide functionality for facilitating legal research over the Internet. Legal research involves the process of locating and organizing a collection of legal documents, including constitutional provisions, statutes, and cases relevant to the matter in question.

LawBot is implemented as a set of agents that operate according to the user's preferences to gather, filter, organize, and recommend pertinent case histories, state statutes, or supreme court cases.

Recognizing the need to simplify research for the public, LawBot has integrated a novel ontology-based search feature. This ontology maps colloquial terms to corresponding legal terminology, enhancing query refinement by searching related words to yield a more comprehensive document set.

2.1.1 LawBot Architecture:

The system architecture comprises the following modules:

- Interface: The interface serves to connect with the user preference database, gather user interactions, forward user queries to the resource managers, and display results provided by the resource manager to the user.
- Resource Manager: This module is responsible for organizing, retrieving, and processing information. It interprets user queries by consulting domain ontologies and deploys the relevant resource agents as needed.

- **Resource Agent:** Resource agents are programmed to access specific web-based sites based on the user's jurisdiction selection. Retrieved documents undergo word filtering and are sorted by relevance. These ordered documents are then sent back to the resource manager using Agent Communication Language (ACL).
- **LawBot Database:** This database stores user preferences, domain ontology, and auxiliary information, such as frequently used links.
- **Proactive Component:** The proactive component uses stored user preferences and past queries to conduct more extensive offline research when the user is not logged in. If new, highly relevant documents are found in this offline search, the user receives notifications about the availability of new documents pertinent to their interests.

2.1.2 Advantages:

- LawBot enables users to define the search domain, ensuring that documents retrieved originate solely from official sources.
- Each returned document link also provides context by including the surrounding text.
- If the number of documents for the user-specified search falls below a threshold, LawBot augments the search by including synonyms, hypernyms, and hyponyms.
- The legal ontology developed is a plug-and-play model and can be re-implemented for different domains, such as a medical bot.

2.1.3 Disadvantages:

- Responding to queries can be slow as it requires fetching the entire document, not just the link, to find the context of the match and rank the document.
- Caching of data locally is not implemented, necessitating internet repository access for each user request.
- The static and handcrafted nature of the ontology is one of its limitations.

2.2 LIRFSS:

In the framework of LIRFSS (Legal Information Retrieval and Focused Semantic Search), the system exclusively focuses on domain-specific legal judgments originating from the Supreme Court of India.

These judgments, pertaining specifically to criminal murder cases, are sourced from the Supreme Court's online repository accessible at <http://judis.nic.in/supremecourt/chejudis.asp>

2.2.1 Information Extraction (IE):

Information Extraction (IE) is the process of deriving organized data from unstructured text. In this scenario, IE is applied to a collection of criminal murder case judgments.

It involves annotating the documents to extract specific details like the date, location of the incident, and pertinent Indian Penal Code (IPC) sections, transforming raw text into structured information.

2.2.2 LIRFSS System Architecture:

- **Online Stored Legal Data:** Judgments are collected in PDF format. These judgments contain metadata, including petitioner names, respondent names, judgment dates, names of judges involved in the case, dates, locations of the crime, relevant IPC sections, and the judgment itself.
- **Information Retrieval:** The initial phase of the system involves basic annotations across the corpus. Various processing modules, including ANNIE, are used for named entity recognition. ANNIE encompasses resources like Ortho Matcher, NE Transducer, Sentence Splitter, Gazetteer, English Tokenizer, and Document Reset PR to recognize named entities within the text. GATE annotates the location of the crime from the text, using the 'Organization' token for the High Court location.
- IPC section numbers are annotated with the 'Identifier' token, and dates are annotated with the 'Date' token. The information extracted from the corpus is supplemented by creating triplets in the format <(location), (IPC-section number), (date)>.

2.2.3 Semantic Search:

In this phase, the extracted information is stored in a database, enabling the execution of relevant queries to obtain the desired results. Triplets are stored in a MySQL database, enabling data retrieval.

2.2.4 Visualization Analysis and Clustering Analysis:

- Visualization Analysis: R, a language and environment for statistical computing and graphics, is used for visualization. R was chosen due to its free and open-source nature.
- Clustering Analysis: Cluster analysis is effective for quickly reviewing data, especially when there are multiple categories. Clustering is applied to a set of judgments, with the chosen variable being the location of the crime mentioned in the text.

2.2.5 Advantages:

- LIRFSS proposes a new hypothesis about a legal system based on a dataset of 20 documents.
- Verification of the hypothesis is conducted through visual data exploration, which presents data visually, allowing users to gain insights, draw conclusions, and interact directly with the data.

2.2.6 Disadvantages:

- The results obtained were unable to locate information about the location of the crime.
- IPC section numbers from the text were misinterpreted as section numbers with dates.
- Calculating recall and precision in the preliminary evaluation is not feasible due to the limited size of the corpus.

2.3 VLA:

2.3.1 Virtual Legal Assistance Components:

- Consists of four components for legal consultations with an AI-based Virtual Assistant.
- Aim to streamline the legal case journey by providing intelligent access to historical case data.
- Reduces time-consuming manual research efforts by legal experts.

2.3.2 User Interaction and Case Refinement:

- Allows users to add case details, narrowing down the search and ranking references.
- Uses scoring algorithms like RP-score and Mean Reciprocal Rank for reference ranking.
- Recalculates scores upon changes in the case text, guiding legal experts in decision-making.

2.3.3 VLA Architecture:

- Text Analytics: Involves tokenization, cleaning, stop word removal, stemming, and part-of-speech tagging. Utilizes algorithms like Word2vec and Doc2vec for text similarities.
- Knowledge Base: Repository storing domain corpus, rules, rank scores, and trained model information.
- Question Generation Engine: Utilizes semantic networks and ontologies to transform contextual text into interrogative sentences post-content determination.
- User Interface with Dialect Converter: Allows speech-to-text and vice-versa conversions, backed by the Question Generation Engine.

2.3.4 Advantages:

- First Line of Support: Acts as the initial support, guiding the way forward for legal experts.
- Dialog-Based Interactions: Engages in auto-way interactive conversations, suggesting pathways for legal cases.

2.3.5 Disadvantages:

- Lack of Human Interaction: This does not provide direct human interaction, potentially limiting nuanced understanding.
- Decoupled Components: The four components operate independently, possibly posing integration challenges.
- Limited Domain Coverage: Not available for all legal domains, restricting its applicability.

III. COMPARISON

Table 1 Comparative illustration of various applications

Published Year	Author Name	Solution	Methodology	Technology Used	Features	Challenges	Input	Output
2022	Mickel Beal	JP Morgan	Contract Intelligence, AI program	AI, Data extraction, automation	Extracts 150 attributes from commercial credit agreements	N/A	Commercial Credit agreements, contracts	Extracted attributes from agreements
2020	Sandip Sen	LawBot	Agent-based Legal research	AI, Ontology-Based search, data gathering	Gathers, filters, organizes and recommends legal documents	Slow response, ontology limitations	Legal Research Preferences user queries	Filtered and Organized Legal documents
2020	Gaurav Goel, Nishan Jain	Virtual Legal Assistance	Text Analytics, ranking	Text analytics, ranking algorithms, semantic networks	Narrows search, ranks references, guides legal experts	Lack of Human interaction	Legal Case details, historical case data	Ranked References based on winning probability
2015	Dan Rubins, Megan Satterfield-smith	Legal Robot	Contract analysis	AI, machine learning, text analysis, document conversion	Converts legal content into a numerical format	Still in beta, advanced analytics	Legal content, documents	Analytical insights
2014	Ilan Admon, Noory Bechor	LawGeex	Machine learning, text analytics	AI, Text analysis, legal expertise	Validates contracts based on predefined policies	N/A	Legal contracts, predefined policies	Edited contracts, approvals, cost reduction
2014	Akash Venkat, Andrew Arruda	Ross Intelligence	Natural language search	AI, machine learning, legal data analysis	Supports legal research, provides recommended reading	N/A	Legal queries, research tasks	Relevant information, recommended readings, case law
2013	Dominic Hudson, Tim Pullan	Thought River	Fathom Contextual Interpretation Engine	AI, machine learning, content summarization, risk analysis	Automates high volume contract reviews to flag risky contracts	N/A	Contract reviews, portfolios, investigation	Content extracts, clause insights, risk flags

2013	Jake Heller, Pablo AI Tendonodo	Casetext	Legal opinions analysis	Legal opinions analysis	Anticipates opposing counsel's arguments, identifies cases	N/A	Legal opinions, cases	Case analysis and identification
2012	Emilio Matthei, Florian Kuhlmann	LEVERTON	AI-powered data extraction	AI, cloud-based platform	Extracts rent, maintenance costs, expiration dates	N/A	Real estate contracts, documents	Data organized in a spreadsheet format
2012	Adam Nguyen, Ned Gannon	eBrevia	Natural language processing	NLP, machine learning, customizable software	Extracts specific information and creates comprehensive reports Accuracy, customization	Accuracy, Customization	Legal contracts, documents	Summarized documents, searchable text
2011	Alexander Hudek, Noah Waisberg	Kira System	Advanced Algorithms, AI contract review	AI, Machine Learning, Text Analysis	Accurate due diligence, content extraction, cross-referencing	N/A	Legal contracts, documents	Extracted and highlighted content
2011	Dhruv Gaur	LIRFSS	Information Extraction	data mining, visualization	Extracts structured data from unstructured legal text	The limited size of the corpus	Legal judgments, criminal case documents	Structured information from unstructured text

IV. CONCLUSION

In conclusion, this present study has provided a comprehensive overview of various AI-driven solutions that are revolutionizing the legal industry. From contract analysis to legal research and information extraction, these technologies are enhancing efficiency, accuracy, and decision-making for legal professionals. While each solution offers unique features and benefits, they also face their respective challenges. The legal landscape is evolving rapidly, and the integration of AI in legal processes continues to shape the future of the legal profession. As the field advances, it will be essential for legal practitioners to stay informed and adapt to these innovative tools to remain competitive and provide better legal services.

REFERENCES

- [1]. Sandip Debnath & Sandip SenB, "LawBot: A multiagent assistant for legal research", Article in IEEE Internet Computing - December 2000
- [2]. Dhruv Gaur, " Data Mining and Visualization on Legal Documents", International Conference on Recent Trends in Information Systems 2011.
- [3]. Nishant Jain, Gaurav Goel, " An Approach to Get Legal Assistance Using Artificial Intelligence," 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) Amity University, Noida, India. June 4-5, 2020.