



ARTICLE

Ontology-Based Crime News Semantic Retrieval System

Fiaz Majeed¹, Afzaal Ahmad¹, Muhammad Awais Hassan², Muhammad Shafiq^{3,*}, Jin-Ghoo Choi³ and Habib Hamam^{4,5,6,7}

¹Department of Information Technology, University of Gujrat, Gujrat, 50700, Pakistan

²Department of Computer Science, University of Engineering & Technology, Lahore, 54890, Pakistan

³Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38541, Korea

⁴Faculty of Engineering, Uni de Moncton, Moncton, NB E1A3E9, Canada

⁵International Institute of Technology and Management, Commune d'Akanda, BP, Libreville, 1989, Gabon

⁶School of Electrical Engineering, Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg, 2006, South Africa

⁷Spectrum of Knowledge Production & Skills Development, Sfax, 3027, Tunisia

*Corresponding Author: Muhammad Shafiq. Email: shafiq@ynu.ac.kr

Received: 16 September 2022 Accepted: 18 April 2023 Published: 31 October 2023

ABSTRACT

Every day, the media reports tons of crimes that are considered by a large number of users and accumulate on a regular basis. Crime news exists on the Internet in unstructured formats such as books, websites, documents, and journals. From such homogeneous data, it is very challenging to extract relevant information which is a time-consuming and critical task for the public and law enforcement agencies. Keyword-based Information Retrieval (IR) systems rely on statistics to retrieve results, making it difficult to obtain relevant results. They are unable to understand the user's query and thus face word mismatches due to context changes and the inevitable semantics of a given word. Therefore, such datasets need to be organized in a structured configuration, with the goal of efficiently manipulating the data while respecting the semantics of the data. An ontological semantic IR system is needed that can find the right investigative information and find important clues to solve criminal cases. The semantic system retrieves information in view of the similarity of the semantics among indexed data and user queries. In this paper, we develop an ontology-based semantic IR system that leverages the latest semantic technologies including resource description framework (RDF), semantic protocol and RDF query language (SPARQL), semantic web rule language (SWRL), and web ontology language (OWL). We have conducted two experiments. In the first experiment, we implemented a keyword-based textual IR system using Apache Lucene. In the second experiment, we implemented a semantic system that uses ontology to store the data and retrieve precise results with high accuracy using SPARQL queries. The keyword-based system has filtered results with 51% accuracy, while the semantic system has filtered results with 95% accuracy, leading to significant improvements in the field and opening up new horizons for researchers.

KEYWORDS

Web 3.0; crime ontology; semantic web; knowledge representation



1 Introduction

The spread of terrorism and crime is one of the biggest problems of our time and it is the most pressing problem today. Collecting, storing, and maintaining crime news is a challenge for journalists and reporters. Crime News describes different types of criminal activity, including murder, fraud, robbery, kidnapping, and terrorist attacks. Extracting relevant information from crime news is important and essential for both the public and law enforcement agencies. In the Internet age, crime news spreads through a variety of sources, including web documents, social networks, video archives, etc. Mostly, web documents and traditional database systems are used to store crime news, and then keyword-based search systems are used to retrieve the information. Crime news on the Internet is abundantly available in unstructured formats, so using traditional keyword-based information retrieval systems to organize large amounts of data can lead to problems in obtaining relevant information [1,2].

Keyword-based systems extract data by finding the exact word in the corpus, and so they cannot understand user queries with greater precision and accuracy. The system provides a huge list of documents to reply to the user's query, which contains multiple unnecessary pages containing specific keywords, and these documents are ranked using different algorithms. For properly managing this massive amount of data through these systems and getting relevant results based on user needs is a daunting task. The query can be a keyword or a phrase, and the search system accepts the query and provides a ranking of documents that contain a specific keyword. Most of the time, users are not able to submit complete information and they use different keywords than those used in the database. Furthermore, keyword-based Information Retrieval (IR) systems lack the intelligence to extract hidden knowledge, which also produces irrelevant results.

For example, consider the following crime news data on plane hijacking: Hijacked Libyan plane lands in Malta with 118 people on board. Malta's Prime Minister Joseph Muscat tweeted: "Aware of a potential hijacking...". From the above data, the keyword-based system cannot determine the exact relationship between the Libyan plane (the hijacked object) and Malta (the current existing location), and cannot definitively answer the following questions due to the lack of semantic support and reasoning mechanisms:

- The city where the plane was hijacked (source)?
- What was the destination of the hijacked plane?
- In which country did the hijacking attack affect passengers?

To solve this problem, many information retrieval techniques have been developed, such as vector space model-based [3], probability-based [4], and context-based information retrieval systems to solve this problem, but unfortunately, search still cannot make people satisfied. In 1998, Berners-Lee proposed a concept of the semantic web, in which information is presented in a machine-understandable form with the integration and help of ontologies [5]. The Semantic Web uses semantic technologies and languages such as web ontology language (OWL), resource description framework (RDF), semantic protocol, and RDF query language (SPARQL) among others. Semantic systems take advantage of efficient information retrieval better than traditional keyword-based systems.

In this paper, we propose a semantic crime knowledge management system that takes news data and identifies crime news from the data by comparing sentences with predefined words, e.g., murder, arrest, and robbery, etc. The crime data is processed and stored in the proposed ontology in semantic structures. Furthermore, rule sets are defined using semantic web rule language (SWRL) to acquire

new knowledge from existing information stored in the ontology. The proposed ontology is then stored in Apache Jena Fuseki. Our semantic system translates user queries into SPARQL queries.

This work has the following key contributions:

- We develop a text unit crime news knowledge management system based on an ontology representation model.
- We derive new knowledge from existing information and store it in the crime ontology.
- We provide ontology-based data access and integrate SPARQL queries to retrieve documents with high precision.
- We validate and compare results produced by keyword-based and ontology-based systems to determine which technique produces the most precise and accurate results.

This paper is divided into five distinct sections. First section deals with an introduction to the criminal intelligence system. [Section 2](#) describes the literature review whereas [Section 3](#) explains the methodology. [Section 4](#) presents the results and discussion. Finally, [Section 5](#) describes the conclusions and future work.

2 Related Works

This section is about some of the research conducted by the researchers related to the current topic under observation.

The authors proposed an ontology-based system using the semantic vector space model (SeVSM) [6]. The SeVSM model is developed based on the principle of the VSM technique while adding semantic functions. SeVSM outperforms vector space model (VSM), and in order to verify the performance of the system, various queries are performed on both systems. VSM retrieves results with 20% accuracy, while SeVSM retrieves results with 80% accuracy. An in-depth understanding of the SPARQL protocol and RDF query language is given [7]. This work has deepened the capabilities of the query language and illustrates with concrete examples, and defines the semantics of these query functions. Hogan introduced the SPARQL participation profile, which enables query results to support inference, including support for RDF, RDF Schema (RDFS), and OWL semantics. This work discusses hypertext transfer protocol (HTTP) based procedures that can be used to send requests to SPARQL services on the Web, and a SPARQL service description vocabulary that can be used to describe and advertise the functionality supported by these services. Finally, it discusses the importance of SPARQL for web data, a major research direction currently being explored and open challenges.

In [8], issues related to keyword-based systems are addressed and the importance of ontology-based systems in information retrieval is shown. Classical keyword-based systems ignore semantic information, and various other problems associated with it, such as word mismatch problems and synonym problems. To retrieve information efficiently, ontology-based systems can help, but their ontologies are based on genetic algorithms. Furthermore, ontologies do not derive new knowledge from existing information. The authors gain insight into the need for ontology-based information retrieval [2]. A large number of users conduct research and use the data available on the Internet, and the data available on the Internet is in a large, unstructured format. Keyword-based systems are based on data statistics and face various problems, such as word mismatch, reasoning knowledge, and the inability of keyword-based systems to understand the semantics of the query. Ontology systems can solve this problem and retrieve information with higher accuracy. It outlines various modern ontology-based information retrieval techniques for text, multimedia and multilingual data types.

The semantic-based search engines have an advantage over keyword search engines in the accuracy of returned results. The search procedure of a semantic search engine is founded on the semantics of the query. It allows users to guarantee more precise and relevant results according to the meaning of the word the user is searching for, rather than the page ranking algorithm and keywords. Users can get relevant results through a semantic-based search engine. This research also lists the fundamental differences between keyword search engines and semantic-based search engines and clearly illustrates the importance of semantic search engines compared to keyword-based systems [9]. The work presented in [10] provided users with a new way to obtain information on the Internet. There are a lot of documents on the Internet, and getting relevant information from them is also a very difficult task. The Semantic Web can be called the extended generation of today's World Wide Web (WWW), which habits semantic markup to retrieve information from web pages. It generates annotations and metadata for raw data by transforming it into a qualified representation document. These old-fashioned search engines have not yet implicit the structure and semantics of semantic web pages. These engines expect documents to be unstructured text, but various documents are involved in the IR process. Ontologies are considered the backbone of the semantic system. We observe that semantic markup in documents results in a large number of documents related to text documents.

In [11], an ontology-based extraction mechanism for news data using the semantic web is provided. Arora's research focuses on extracting news data and analyzing it to identify fake news. However, this system does not acquire knowledge from ontology. Furthermore, it does not provide a mechanism for efficient storage and retrieval purposes. The Arora system does not support data visualization. The work presented in [12] focused on the automation of the UK surveillance system, which is equipped with 5.9 million cameras. Data retrieval of suspect-related information takes a lot of time and involves very high labor costs. To this end, a basic ontology is developed to retrieve complex events. The system classifies incidents, investigates suspicious behavior, regulates the presentation of suspects, assists in forensic analysis, and automatically expands knowledge structures.

Reference [13] described the safest route between two locations on the roadmap. Different points are scored based on crime rates in specific locations. To find the lowest path cost, Dijkstra's algorithm is used to count the cumulative number of violations for each node-specific intersection. In [14], a semantic integration system (MOSAIC) was proposed, which uses various tools including multi-model, data analysis, data mining, criminal networks, and decision support systems to understand criminal networks. The author provided an ontology-based sentiment analysis mechanism for the general election news to see whether it receives a positive impact or a negative one. They performed two experiments. The first experiment was based on the lexicon and the second experiment was conducted using the support vector machine (SVM). The SVM algorithm shows better performance than the lexicon experiment. The lexicon-based method shows that the recall score was 28.8% and the precision score was 46.3% and the SVM method shows that the recall score was 82.4% and the precision score was 83.3%. Moreover, it is planned to use a combination of different methods to improve the accuracy of their system [15]. However, the accuracy of a crime news semantic knowledge management systems is still of interest.

3 Materials and Methods

3.1 Architecture of Methodology

The architecture of the methodology is shown in Fig. 1, therein we show the overall system architecture and design of the semantic system. The detailed explanation of each module is discussed in the following sub-sections.

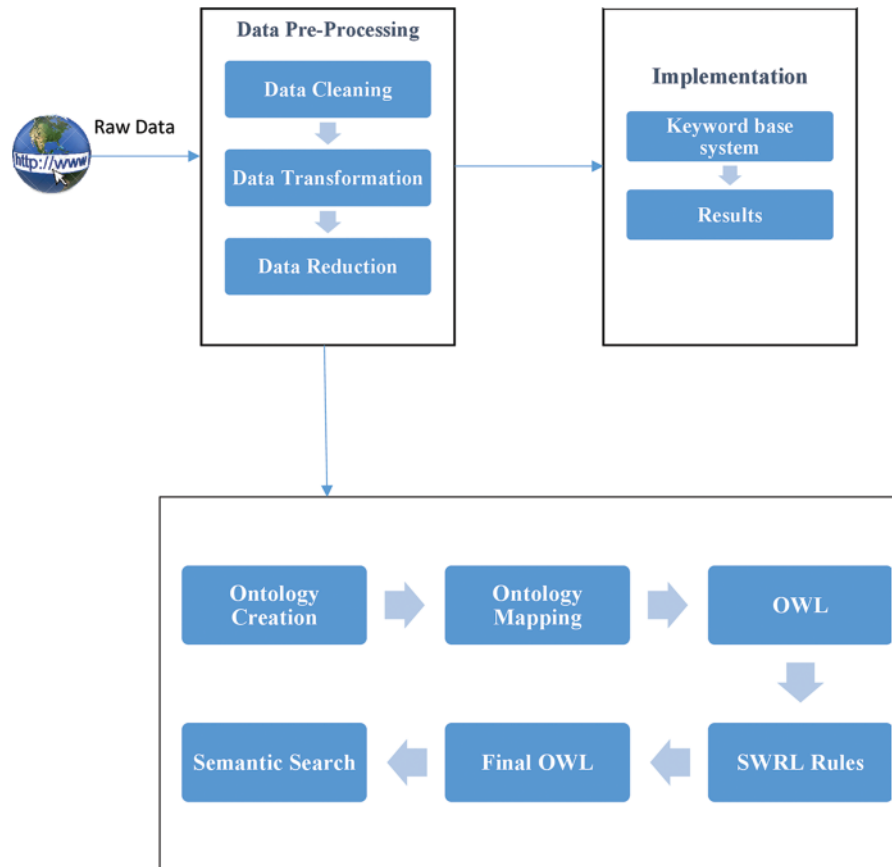


Figure 1: Architecture of the proposed methodology

3.2 Data Description

The dataset is from the Kaggle dataset repository [16]. The dataset used in this work contains headlines from the past few years. The contents of the records include “PageType”, “Link_paper”, (the date the article was published online, in MM-DD-YYYY format), “Link_news” and “headline text” (English title text). It contains various news such as sports, entertainment, political issues and crime news. The datasets used in this study were initially unstructured. Therefore, in order to distinguish crime news from non-criminal news, these statements are compared with predefined words, such as murder, arrest, robbery, etc., describing the nature of the criminal activity. After extracting crime reports, the system stores the data in the proposed ontology called CrimOntology.

3.3 Ontology

Ontology is a body of knowledge that describes a domain of knowledge that humans can understand. In ontology, data is represented in the form of a class, each class delivers an abstract structure, and instances are nodes stored in the class. Data properties are RDF literals that represent node values, and object properties describe the relationship between two instances. Stored data are called individuals in the ontology. The semantic system offers dynamic structure. Hence, new information can flexibly be constructed from any instance at any stage.

3.4 Ontology Development

Ontology 101 technique was used to develop CrimOntology. The steps for developing the instance of ontology are as follows. In the first step, we define the scope, crime news and ontology. This ontology will cover news specific to crime news. We have examined existing ontologies for defining structures. We discovered an open-source ontology of Amhiggin [17] called CrimeGeolocationOntology. The ontology is well defined and carries aspects of the ontology structure, such as cardinality.

3.4.1 Classes and Class Hierarchy

Classes are the main feature in CrimOntology, they behave like nodes related to domains, range of relationships and represent a hierarchy of system domains. The class and subclass structure describes the class relationship to each other. Below is a list of the classes in CrimOntology and their functions.

- Month: This class contains instances of months, such as March, April, and May. It is used to determine the month in which the crime occurred.
- Location: This class categorizes where the crime occurred (e.g., cities such as Lahore, Karachi, etc.).
- News: The news class defines all kinds of news such as sports, health, and weather. So it has a crime subcategory that only shows crime news.
- Crime: This class focuses on crime news, statistics and types.
- Season: It contains four seasons including summer, winter, autumn and spring. It shows which crime happened in which season.

3.4.2 Object Property

Object properties are used as predicates following first-order logic (FOL) to describe complex structures and show relationships between classes and individuals. For CrimOntology, the object properties are:

- cityCrimes: It has a domain (Location) and a range (Crime) to determine where the crime occurred.
- crime type: There is a domain (crime) and a range (Type) to show the relationship between crime and crime type, e.g., Fraud, Murder, Kidnapping, etc.
- occurOn: It has a domain (Crime) and a range (Month) to determine which crimes have occurred in a given month.
- hasLocation: It has a domain (Crime) and a range (Location). This is the opposite of cityCrimes.
- inSeason: It has domain (Crime) and a range (Season) that will have information about seasons and crimes.
- dateCrime: It has domain (Month) and a range (Crime). It is the inverse of occurOn.
- seasonCrime: It has a domain (Season) and a range (Crime). It is the opposite of the inSeason property.

3.4.3 Data Property

In Fig. 2, the data property describes the dynamic structure of the instance to define the literal value of each node in the class structure. The data properties of the system are defined as follows.

- NewsTitle: The title that describes the crime story.
- Arrested: It shows how many criminals have been arrested in criminal cases.
- CrimeDate: This data property displays the full formatted date.

- Fraud: It shows how many fraud crimes have occurred.
- Injured: This shows the number of persons injured by a crime or accident.
- Kidnap: It shows the number of people kidnapped in crime, especially children.
- Killed: This property shows how many people were killed in the course of a crime, including suspects.
- noOfCriminals: It shows how many criminals are involved in the crime.
- noOfVictims: This property shows how many people are victims of crime.

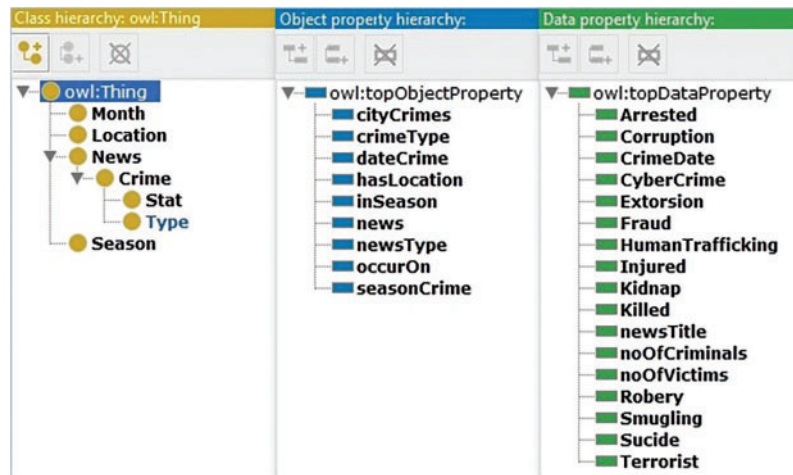


Figure 2: The hierarchical structure of ontology classes, object properties, and data properties

3.5 Ontology Structure

In Fig. 3, we show the final ontology. CrimOntology's classes (representing concepts) are represented by red nodes. Classes provide abstract structure and instances are nodes stored in the class. Each class has its own domain and range, which is used to set the class domain and range of that class. The data properties of the class are indicated by green nodes, and the red text describes the data type literals associated with the nodes. In an ontology instance, data properties are used to describe the relationship between two instances or two individuals of data values. Data properties are RDF literals that represent node values, and object properties describe the relationship between two instances. Finally, blue nodes represent object properties and define conceptual relationships. Object properties are used to describe the relationship between two instances or two individuals of a class.

3.6 Ontology Mapping

The ontology mapping module converts unstructured and semi-structured data into structured form. After parsing all the data, the JENA library maps the data to CrimOntology. It facilitates insertion, deletion, and other operations on data, and creates relationships between classes and nodes. Finally, the connection is safely closed and the JENA writer saves the CrimOntology as an RDF/extensible markup language (XML) file format.

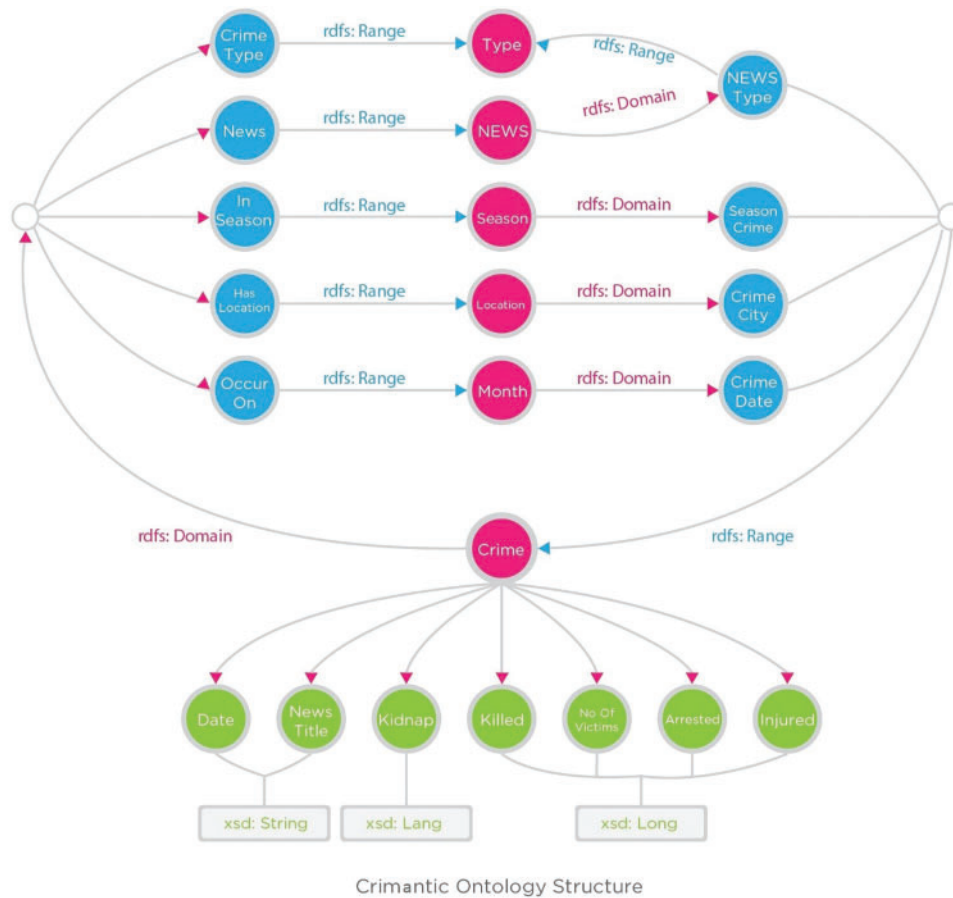


Figure 3: Ontology structure in the system

3.7 Semantic Inferencing Rules

Inference engines and rules bring non-explicit knowledge from existing information. A set of rules is defined in semantic web rules language (SWRL) in Fig. 4. These rules add new relationships between classes and nodes. Using Rule 1, the system is able to indicate the season of crime, while the news indicates the month. Information about seasons is automatically derived from this rule. This type of reasoning is a powerful feature of semantic systems.

If crime news CrimeNews1 occurs in a particular month of April, the system can identify when the crime occurred. Information about the season is not stored in CrimOntology, but this rule adds new information to the system that CrimeNews1 occurs in the spring of the season. Some other rules regarding natural language query (NLP) query and semantic inferencing are explained in Table 1.

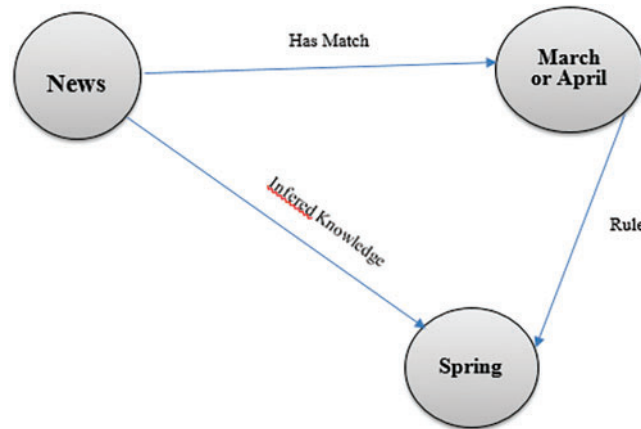


Figure 4: $\text{Crime}(\text{?crime}) \wedge \text{occurOn}(\text{?crime}, \text{March}) \wedge \text{occurOn}(\text{?crime}, \text{April}) \rightarrow \text{inSeason}(\text{?crime}, \text{Spring})$

Table 1: SWRL rules and their conclusions

Sr. No.	NLP query	Semantic inferencing
1	hasLocation(?crime, ?place)	cityCrimes(?place, ?crime)
2	dateCrime(?date, ?crime)	occurOn(?crime, ?date)
3	Crime(?crime) ^ occurOn(?crime, ?date)	Month(?date)
4	Crime(?crime)	News(?crime)

3.8 Semantic Searching

After building the complete ontology, the proposed ontology is stored on the Apache Jena Fuseki server, and various queries are run to see the efficiency of the retrieval results. The SPARQL query language is used to manage data stored in RDF/XML format, including subject, predicate, and object parts. SPARQL [18] is a standard query language that allows users to query data from linked data applications and RDF databases. The “RDF Query Language and SPARQL Protocol” is another name of the SPARQL which allows RDF-based retrieval of data by the users from any data source or database. SPARQL queries provide the ability to search complex types of relationships from a variety of data. It can also extract hidden information stored in various sources and formats. It is specifically designed to retrieve data stored in RDF format, whereas structured query language (SQL) and NoSQL are designed to retrieve data from relational databases.

With this basic structure, the system can call up complex information according to user requirements. We explain the different searches and their associated SPARQL queries in Table 2.

These queries count the number of accidents or fatalities in the news and get the exact results clearly.

Table 2: SPARQL queries

Sr. No.	NLP query	SPARQL
1	No. crimes occurred in Lahore	?CrimeOntology crime:cityCrime crime:Lahore.
2	Find all types of frauds in all the cities	?x test : newsTitle ?y FILTER(regex(str(?y), "fraud")). ?x test : hasLocation ?l
3	No. of people killed in crimes	?CrimeOntology crime:NewsType crime:Killed.
4	No. of crime occurred in the winter season	(count(?x)as ?count) ?x test : inSeason ?y FILTER(regex(str(?y), "Winter"))
5	Find all types of terrorist attacks	?CrimeOntology crime:NewsType crime:Terrorist.

4 Results and Discussion

System performance is validated using precision and recall parameters. The system of Smeaton TREC [19] explains the development and benchmarking of retrieval system evaluation.

We conducted two experiments. In the first experiment, we used Apache Lucene [20] to implement a keyword-based text search engine. Apache Lucene is open-source software for developing powerful, full-featured keyword-based text search engines. A second experiment was conducted to evaluate the proposed system. To implement the semantic system, we used Protégé, a free and open-source framework and ontology editor for creating intelligent systems. Three metrics are used to evaluate the system: Recall, Precision and F-Measure. Precision describes how many relevant results are extracted from a given output result. Mathematically it can be written as:

$$Precision = \frac{No. of retrieved relevant results}{Total number of retrieved results} \quad (1)$$

The recall describes how many related results were fetched by the total number of related results. Recall can be mathematically written as:

$$Recall = \frac{No. of relevant retrieved results}{Total relevant results} \quad (2)$$

F-Measure is a combination of precision and recall. The F-Measure can be written mathematically as:

$$F - Measure = \frac{2 * precision * Recall}{Precision + Recall} \quad (3)$$

4.1 Comparison between Results of Two Techniques

We show different queries applied to the traditional system and our proposed semantic system, and validate the performance using precision and recall parameters. Whereas traditional systems use Apache Lucene to filter out results, semantic systems filter out results by mapping user queries to SPARQL queries. The main component of the proposed system is semantic search, which allows users to query data from semantic stores. We applied the same requirements to both the traditional system and our semantic system and analyzed the results. We show the results obtained with our semantic system, while in the second part we show the results obtained with the traditional system.

First, we will discuss the precision of these two systems. To calculate the precision, we perform various queries on the keyword-based and semantic system. We calculated the precision value for each query by dividing the extracted relevant results by the specified output results against the user's search in Fig. 5. Afterward, we calculated the average value of precision based on the results for both the keyword- and semantic-based systems. The precision of the keyword-based information retrieval system is low because most of the search results returned do not match the user's search preferences. The proposed semantic system has good precision because all the results found are semantically matched to the user's search. The keyword-based system filtered out the results with 66.2% precision. On the other hand, the semantic system filtered out the results with 96.6% accuracy.

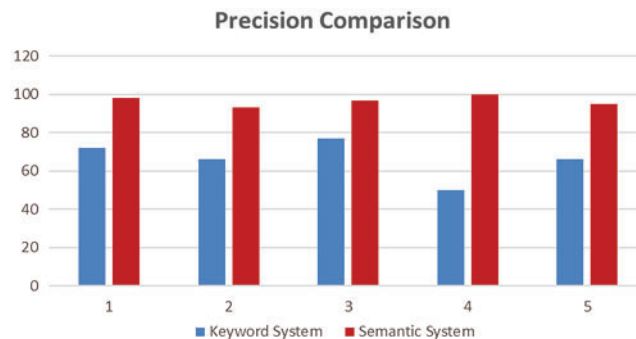


Figure 5: Precision comparison of keyword and semantic system

Now, we will discuss the recall of both systems. To calculate the recall, we perform various queries on the keyword-based and semantic system. We calculated the recall value for each query by dividing the extracted relevant results by the total number of relevant results searched by the user as shown in Fig. 6. Afterward, we calculated the average value of recall based on the results for both the keyword-based and semantic system. The recall rate of a keyword-based information retrieval system is low because only a small percentage is relevant to the user's search. The semantic system has good recall value because all retrieval results semantically match the user's search preferences. The keyword-based system filtered out the results with a recall rate of 48%. On the other hand, the semantic system filtered out the results with 94% recall.

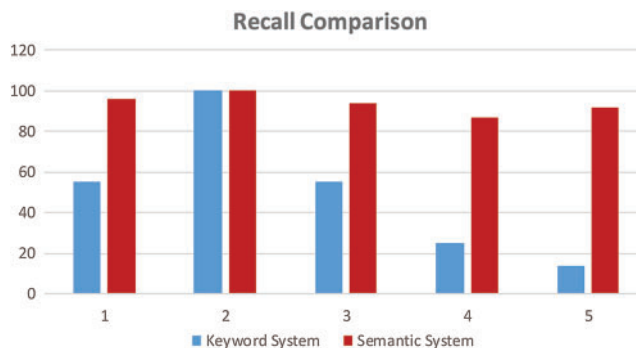


Figure 6: Recall comparison of keyword and semantic system

Finally, we compute the F-measure based on precision and recall and verify the accuracy as shown in Fig. 7. The F-measure is calculated based on the above formula. We calculated the average value of the F-measure for both keyword-based and semantic system. The keyword-based system filtered out

the results with 51% accuracy. On the other hand, the semantic system filters out the results with 95% accuracy, which is a considerable improvement over the keyword-based system.

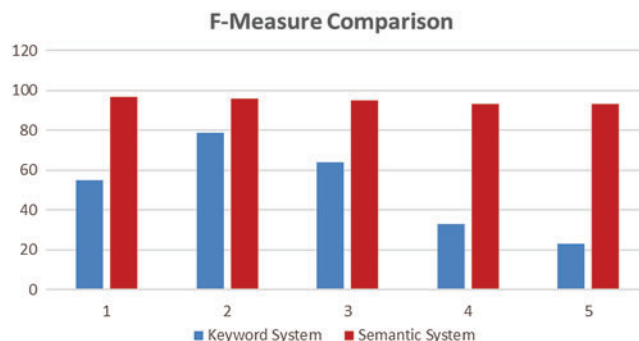


Figure 7: F-Measure comparison of keyword and semantic system

4.2 Discussion

We have successfully demonstrated that the system is able to find hidden knowledge from existing data. By discovering new knowledge, the system can work independently of the availability of complete information. For example, the system does not contain information about the season in which the crime occurred, but it can be easily deduced from the month. The system automatically determines from the date of the message that if the crime happened in December, then the crime happened in the winter. The system writes new rules using the SWRL rules language. These rules derive new propositions and add them to the ontology.

The structure of CrimOntology is dynamic, which makes it suitable for information retrieval through NLP interfaces. The user writes a query in English. The query text is tokenized, and words like “Lahore” are classified as “city” to explain their meaning. Using categorization, we convert the English query to a SPARQL query and run the SPARQL query in GraphDB to display the results. Our system uses existing information, and when a new crime ontology is imported into GraphDB, all information and metadata (vocabulary) are updated. This allows users to query new crimes without changing the source code, making them more scalable with new data.

This work proposes a dynamic knowledge management system integrated with GraphDB. It stores data in the form of nodes, edges, and relationships, and provides a REST API interface for retrieving data from semantic storage using SPARQL. The REST API interface also allows users to perform various operations using curl commands, such as GET, POST, and DELETE. Additionally, GraphDB provides a security layer, query optimization, and ontology management, making it the best choice for crime ontologies. The traditional system used Apache Lucene to filter out the results with 51% accuracy. On the other hand, the semantic system, which filters the results by mapping user queries to SPARQL queries, filtered results with 95% accuracy, showing improvement. Keyword-based systems cannot adapt to changing data, but semantic systems can perform data manipulation and consistency checks for efficient storage and retrieval.

5 Conclusions

We propose a crime news semantic knowledge management system, which is more powerful than traditional RDBMS-based systems. It can identify and correlate different types of information and

is able to understand complex English user queries. Our semantic system uses SPARQL queries to retrieve precise results with 95% accuracy, a major improvement over traditional systems. In the future, CrimOntology can expand the breadth and depth of more spatiotemporal information about crime dynamics, crime types, and victims to expand the thinking capabilities of the system. Similarly, hidden and implicit information about criminal methods and suspect preferences for specific crime types can be extracted by analyzing their past behavior. Moreover, SPARQL query interfaces for the image-based data can be integrated with the current system to retrieve the information more effectively from the knowledge base. This work discusses various forms of crime reporting, such as seasonal crime, accidents, and urban crime. It can also expand to other types of crime, including spy agencies, mafia, and cybercrime.

Acknowledgement: The author acknowledges Natural Sciences and Engineering Research Council of Canada (NSERC) and New Brunswick Innovation Foundation (NBIF) for the financial support of the global project.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: F. Majeed, A. Ahmad, M. A. Hassan, M. Sahfiq; data collection: M. Shafiq, J.-G. Choi, H. Hamam; Data curation, analysis and interpretation of results: M. Shafiq, F. Majeed, A. Ahmad, M. A. Hassan; draft manuscript preparation: M. Shafiq, J.-G. Choi, H. Hamam. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data underlying this article will be shared (after the patent of the underlying research is filled) upon reasonable request to the corresponding author.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] M. Maree, "Multimedia context interpretation: A semantics-based cooperative indexing approach," *New Review of Hypermedia and Multimedia*, vol. 26, no. 1–2, pp. 24–54, 2020.
- [2] M. N. Asim, M. Wasim, M. U. Ghani Khan, N. Mahmood and W. Mahmood, "The use of ontology in retrieval: A study on textual, multilingual, and multimedia retrieval," *IEEE Access*, vol. 7, pp. 21662–21686, 2019.
- [3] T. Dasgupta, A. Naskar, R. Saha and L. Dey, "CrimeProfiler: Crime information extraction and visualization from news media," in *Proc. of Int. Conf. on Web Intelligence*, Leipzig, Germany, vol. 1, pp. 541–549, 2017.
- [4] A. Malve and P. M. Chawan, "A comparative study of keyword and semantic based search engine," *International Journal of Innovative Research in Science Engineering and Technology*, vol. 4, no. 11, pp. 11156–11161, 2015.
- [5] T. Berners-Lee, "Semantic web road map," 1998. [Online]. Available: <https://www.emse.fr/~beaune/websem/SWRoadmapLee.pdf>
- [6] M. Tang, J. Chen and H. Chen, "SemOIR: An ontology-based semantic information retrieval system," in *Proc. of Int. Conf. on Software Quality, Reliability and Security Companion*, Macau, China, pp. 204–208, 2020.
- [7] E. Kalemi, S. Y. Yayilgan, E. Domnori and O. Elezaj, "SMONT: An ontology for crime solving through social media," *International Journal of Metadata Semantics and Ontologies*, vol. 12, no. 2, pp. 71–81, 2017.

- [8] B. Yu, "Research on information retrieval model based on ontology," *EURASIP Journal on Wireless Communications and Networking*, vol. 1, no. 2019, pp. 1–8, 2019.
- [9] A. Malve and P. M. Chawan, "A comparative study of keyword and semantic based search engine," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 4, no. 11, pp. 11156–11161, 2015.
- [10] A. Gómez-Pérez and O. Corcho, "Ontology languages for the semantic web," *IEEE Intelligent Systems*, vol. 17, no. 1, pp. 54–60, 2002.
- [11] S. Arora and N. Baliyan, "Extraction and analysis of information in news domain using semantic web," in *Proc. of Int. Conf. on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, Ghaziabad, India, pp. 1–6, 2019.
- [12] C. Allen, U. Nodelman and E. N. Zalta, "The Stanford encyclopedia of philosophy: A developed dynamic reference work," *Metaphilosophy*, vol. 33, no. 1, pp. 210–228, 2002.
- [13] V. Sharma, R. Kulshreshtha, P. Singh, N. Agrawal and A. Kumar, "Analyzing newspaper crime reports for identification of safe transit paths," in *Proc. of Conf. of the North American Chapter of the Association for Computational Linguistics*, Seattle, Washington, pp. 17–24, 2015.
- [14] P. Seidler, "R. Adderley criminal network analysis inside law enforcement agencies: A data-mining system approach under the national intelligence model," *International Journal of Police Science and Management*, vol. 15, no. 4, pp. 323–337, 2013.
- [15] E. A. Kardinata, N. A. Rakhmawati and N. A. Zuhroh, "Ontology-based sentiment analysis on news title," in *Proc. of Conf. on Computer and Information Technology (EIconCIT)*, Surabaya, Indonesia, pp. 360–364, 2021.
- [16] I. U. Rahman, "Dawn news headlines, Kaggle," 2019. [Online]. Available: <https://www.kaggle.com/datasets/ibadia/dawn-news-headlines>
- [17] P. Y. Vandenbussche, J. Umbrich, L. Matteis, A. Hogan and C. Buil-Aranda, "SPARQLES: Monitoring public SPARQL endpoints," *Semantic Web*, vol. 8, no. 6, pp. 1049–1065, 2017.
- [18] J. Hosseinkhani, H. Taherdoost and S. Keikhaee, "ANTON framework based on semantic focused crawler to support web crime mining using SVM," *Annals of Data Science*, vol. 8, no. 2, pp. 227–240, 2021.
- [19] F. A. Smeaton, O. Paul Over and K. Wessel, "Evaluation campaigns and TRECVID," in *Proc. of ACM Int. Workshop on Multimedia Information Retrieval*, California, USA, pp. 321–330, 2006.
- [20] A. Lucene, "A high-performance, full-featured text search engine library, The Apache Software Foundation," 2021. [Online]. Available: <https://lucene.apache.org/>