**SEMANTIC DATA TECHNOLOGY**

Department of Computing and Information Science



**Modelling for Crime Analysis Dataset**

**Module Code: MOD004979**

**Faculty:**

Module Leader: Arooj Fatima

Module Tutor: Priyanka Sharma

Name: Syed Sarmasth Hussain

Student ID: 2418497

**Table of Contents**

[1. Introduction 3](#_Toc184952893)

[2. Concept and Requirement Analysis 3](#_Toc184952894)

[2.1 Application Concept/Rationale 3](#_Toc184952895)

[2.2 Aim of the project 3](#_Toc184952896)

[2.4 System Usefulness and Achievability 4](#_Toc184952897)

[2.5 Questions related to Crime Data and Management 4](#_Toc184952898)

[3. Design 6](#_Toc184952899)

[3.1 Summary of Ontology 6](#_Toc184952900)

[3.2 Classes 6](#_Toc184952901)

[3.3 Properties 7](#_Toc184952902)

[3.4 Data Properties 8](#_Toc184952903)

[3.5 Individuals 9](#_Toc184952904)

[3.6 Constraints and Axioms 10](#_Toc184952905)

[4. Requirements Mapping 11](#_Toc184952906)

[4.1 Mapping Table 11](#_Toc184952907)

[5. Evaluation and Use 15](#_Toc184952908)

[5.1 Use Case 1 15](#_Toc184952909)

[5.2 Use Case 2 16](#_Toc184952910)

[5.3 Use Case 3 16](#_Toc184952911)

[5.4 Use Case 4 17](#_Toc184952912)

[5.5 Use Case 5 17](#_Toc184952913)

[6. Implementation 18](#_Toc184952914)

[6.1 Understanding the Problem and the Dataset 18](#_Toc184952915)

[6.2 Natural Language Processing (NLP) Techniques 19](#_Toc184952916)

[*6.2.1 Approach to NLP* 19](#_Toc184952917)

[6.3 Overall Code 20](#_Toc184952918)

[*6.3.1 Structure of SPARQL Queries* 20](#_Toc184952919)

[6.4 Code Validation 20](#_Toc184952920)

[7. Conclusion and Future Work 21](#_Toc184952921)

[References 22](#_Toc184952922)

# **Introduction**

The task entails the creation of asemantic-basedapplication. For this application, we will be concentrating on Crime Data Analysis and Search Application. This system will enable the user to navigate through the crime-related data and make the results refined by using parameters including crime type, location, date and time, etc., and to also answer queries. Semantic technology will assist in improving the data storage, and the associated searching and querying using ontologies for defining the relationships and hierarchy among the concepts obtained in the domain of crimes.

# **Concept and Requirement Analysis**

## **2.1 Application Concept/Rationale**

The Crime Data Analysis and Search Application is an application designed to enable users to gain insight into crime statistics and its trends. They enable semantic association of various aspects of the crime data including, crime type, location, and date for formulating intelligent and flexible interrogation of the crime data. Concerning the aspect of functionality, users can query specific information about crime patterns, differences and similarities and format the data in a manner that is different from how data from traditional database systems can be processed.It makes use of SPARQL for analysing the specific types of crime, the locations where these crimes were committed, as well as other correlated metadata. It may be useful to law enforcement officers, researchers, policymakers, as well as the general public providing them with an opportunity to get crime statistics easier.

## **2.2 Aim of the project**

The maintargets of the application are:

* To implement a simple and flexible environment for querying crime data in a natural language.
* To execute a semantic model for crime data using an ontology to facilitatedata linking and intelligent querying.
* To support a variety of queries, such as categorizing crimes based on specific types, dates, and locations.
* To provide insightful and accurate results based on user-defined criteria.

## **2.4 System Usefulness and Achievability**

The goals of the system are achievable and realistic.Through OWL, RDF, and SPARQL, which are semantic web technologies, the application can enable an efficient and extensive method of asking for information with the complexity of the data and is hence ideal for evaluation in crime analysis; urban structure and other related areas of study. Ontology integration facilitates relationships and classification of crime data making the system functional for future extensibility (Keshavarzi and Ghaffary, 2022).

## **2.5 Questions related to Crime Data and Management**

*Questions Related to Data Management*

1. How will crime types be classified in the ontology, such as violent crimes or property crimes?
2. What data sources will provide information that will fill the crime data, such as police data or crime reports?

It is also important for the whole modelling approach to establish,

1. How different locations will be represented in the data model such as the city, the district and the neighborhood.
2. What quantity of crime data, is expected, and how this may influence performance?
3. How will we handle incomplete or missing dataconcerning crime?
4. What interval will be available for the user to plan at (for instance weekly, monthly, daily)?

*Questions Related to Search Functionality*

1. What the system should include as a search option – date, type of crime, district?
2. What will be done when a user makes a query that does not return any data?
3. How does the user input a query and how can this be optimized for intuitive learning?
4. Is it possible for users to search for crime trends at any one period of time against another such as one year against the other?
5. How the system will enable users to narrow down their search (e.g. applying filters)?
6. If the user’s query is quite general, what kind of information or advice will the system give?

*Questions Related to Semantic Model and Ontology*

1. What properties and classes will be included in the crime ontology (e.g., CrimeType, Location, and Date)?
2. How will ontology define/handle the relations between classes (for instance CrimeType and Location) under the ontology?
3. What can be the benefit of using a semantic model (ontology) over a traditional relational database for this application?
4. For which future developments does the ontology facilitate addition (for example for the expansion of crime kinds or areas)?
5. How will the ontology capture the relationship between crimes, offenders and the victims?
6. How will the credibility of the ontology be established to have the capacity to present accurately the crime domain?

*Questions Related to User Interface and Experience*

1. What kind of environment will the users be operating the application within (web-based, mobile etc)?
2. Which levels of technical skills will the users require to be able to utilize the application?
3. Can the system take natural language queries; and if so, how will it process the natural language?
4. What kind of display will it give when users enter a query (For example, graphs, tables, lists)?
5. In what way are the search results going to be ordered in a way that displays the most relevant data on top?
6. Which control features and elaborate user feedback channels would be incorporated (for example, errors, or a query-suggestion list)?

*Questions Related to Security and Privacy*

1. What measures shall we take so that protected data, for instance, data of the victims, would not be disclosed to unauthorized people?
2. Will the system provide an authentication mechanism for the users of the system and how will the different access levels be implemented?
3. How can we implement close for users requiring the use of assistive technologies such as screen readers?
4. When it comes to data sharing and export functionality, how do we address it & what formats are to be supported?
5. Can users input new informations, if yes, how will data quality will be regulated?
6. Which steps shall be taken to increase the chances of accuracy and reliability in the crime data of the application?

# **Design**

## **3.1 Summary of Ontology**

Crime Data Analysis and Search Application is a semantic-based application that searches crime data through an ontology(Spyropoulos et al., 2023). In the design process, one recognises classes, properties and individuals that should be included and constraints and axioms that should be sought to improve the structure of the data. For creating the ontology for this dataset, Protégé is used. The penalty is used as the basis for the ontology design – the main idea behind the penalty is semantically organizing the crime data to get insights into the patterns and trends. The use of ontology will allow users to carry out intelligent searches and retrieval of archived crime data.The crime domain requires some distinct classes and properties to mirror significant entities and interactions of the ontology. Here is an outline of the key elements in the design:

## **3.2 Classes**

We can define classes as subsets of the domain that can be also referred to as the classes of objects within a certain domain. The crime ontology includes the following main classes:

|  |  |
| --- | --- |
| **Class** | **Description** |
| Crime | Describes a certain crime event such as type of crime, place, and participants. |
| CaseStatus | Describes the status of a criminal case as open, closed or under investigation of the case. |
| Charge | Stands for a legal amount due to a crime or the suspect, such as theft or assault. |
| Suspect | Stands for a client or suspect in a criminal case. |
| CrimeType | Concerns the nature or kind of a crime (for example violence, property, cyber) |
| CrimeStatistics | Crime data which includes the number of crimes, and crime frequency by area. |
| CriminalProfile | Stands for features of a criminal such as their age, gender or prior criminal record. |
| CyberCrime | Indicates offences that use computers or the internet (for instance, hacking, Internet fraud). |
| Evidence | Stands for tangible or electronic items that belong to a crime like finger prints, videos, documents etc. |
| Sentence | It means legal or punitive sanction given to a culprit or offender once he is convicted of a crime. |
| Location | Contains the information as to the location where the crime was committed (address, city, country, etc.). |
| Arrest | Denotes the process of apprehending a suspect, the officer who effected the arrest and the date when it was done. |
| Officer | Stands for a police officer who participates in investigative, arrest, or case capacities |
| Organisation | An actor who is associated with the crime, investigation or case prosecution (for example, police, lawyer). |
| Vehicle | Stands for a car that was used in the commission of the crime/ belonged to the suspect/victim. |
| Weapon | An object used in an attempt to solicit a crime including, a firearm, knife, blunt object, and the like. |
| Witness | A person who was present and observes a criminal activity or was at the time of commission of the criminal activity. |
| CriminalRecord | Holds the record of a particular person, especially the criminal record, and all the prior convictions or any other offences made by the criminal. |
| Trial | Stands for the legal determination of a crime that involves a court case, trial, date of a trial and the judgment. |
| Victim | Depicts someone who has undergone an injury or any loss in any criminal activity that occurred in the society. |

***Table 1.*** *Summary of Classes*

## **3.3 Properties**

Relationships between classes are defined by properties. There are two main types of properties: Object Properties are links between classes and Data Properties are links between classes and literal values.

*Object Properties*

In an ontology, object properties define how objects of particular classes are related. Table 4 below presents the properties of objects that make up the crime ontology.

|  |  |  |
| --- | --- | --- |
| **Property** | **Domain** | **Range** |
| hasSuspect | Crime | Suspect |
| hasVictim | Crime | Victim |
| involvesEvidence | Crime | Evidence |
| witnessedBy | Crime | Witness |
| usedWeapon | Crime | Weapon |
| profileOf | CriminalProfile | Suspect |
| relatedToOrganization | Crime | Organization |

***Table 2.*** *Object Properties in Ontology*

hasSuspect: Embodies the relationship where a crime is linked to a suspect.

hasVictim: Denotes the relationship between a victim and their crime.

involvesEvidence: Corresponds to the relationship between evidenceand crime involved.

witnessedBy: Describes the link between a crime and the witness who observed it.

usedWeapon: Relates a crime to a weapon used in the act.

profileOf: Associates a criminal profile with a suspect.

relatedToOrganization: Describes the connection between an organization involved such as a law firm, police department etc. and crime.

## **3.4 Data Properties**

Data properties refer to the elements of a class that describe instances within that class(Jia et al., 2023)these are not relationships but attributes that characterise the entity. The following Table 3 showsthe properties of data used in the crime ontology.

|  |  |  |
| --- | --- | --- |
| **Property** | **Domain** | **Range** |
| hasDate | Crime | xsd:dateTime |
| severityLevel | Crime | xsd:string |
| Age | Suspect/Victim | xsd:int |
| Gender | Suspect/Victim | xsd:string |
| Description | Crime/Suspect/Victim | xsd:string |
| Height | Suspect/Victim | xsd:float |
| Weight | Suspect/Victim | xsd:float |
| crimeReportNumber | Crime | xsd:string |
| incidentType | Crime | xsd:string |
| locationName | Location | xsd:string |
| locationCoordinates | Location | xsd:string |
| investigationDate | Investigation | xsd:dateTime |
| trialDate | Trial | xsd:dateTime |
| sentenceDuration | Sentence | xsd:int |

***Table 3.*** *Data Properties*

hasDate: Represents the date when a crime occurred.

severityLevel: Describes the severity of the crime (e.g., low, medium, high).

age: Represents the age of the suspect or victim.

gender: Describes the gender of the suspect or victim.

description: Provides a textual description of the crime or the suspect/victim.

height: Represents the height of the suspect or victim.

weight: Represents the weight of the suspect or victim.

crimeReportNumber: A unique identifier for the crime report.

incidentType: Describes the type of incident (e.g., theft, assault).

locationName: Specifies the name of the location where the crime occurred.

locationCoordinates: Represents the geographic coordinates of the crime location.

investigationDate: The date when the investigation into the crime started.

trialDate: The date when the trial took place for the crime.

sentenceDuration: Represents the duration of the sentence for the convicted individual.

## **3.5 Individuals**

Specific individuals of the society are used in the ontology design to represent the classes developed in the ontology as indicated in Table 4. In crime-related ontological models, they become definite things or objects in a particular domain which may include concrete crimes, criminals, or victims, places, and other characters encountered in the crime scenes.

|  |  |  |
| --- | --- | --- |
| **Individual Name** | **Data Properties** | **Object Property Link** |
| Crime1, Crime2, Crime3 | incidentType, severityLevel | hasSuspect, hasVictim |
| Suspect1, Suspect2 | age, gender, height | hasSuspect |
| Victim1, Victim2 | age, gender | hasVictim |
| Weapon1, Weapon2 | Description | usedWeapon |
| Location1, Location2 | locationCoordinates | occursAt |
| Witness1, Witness2 | age, gender | witnessedBy |

*Table 4. Individuals*

Concerningthe crime ontology, individuals are referred to identify concrete instances of crime, criminals, victims, witnesses, criminals’ tools, and criminals’ places. For instance, Crime1, Crime2 and Crime3 are objects of the Crime class and each of them has its attributes: incidentType, severityLevel, and dateTime. Likewise Suspect1 and Suspect 2 are objects of the Suspect class and their characteristics include age, gender, and height. These attributes aid in explaining the nature of the people as well as the various ways in which they can be distinguished from one another in as much as a highly structured manner allows.

Moreover, interpersonal relations between entities are described in the context of object properties. These properties relate entities in the relations of the ontology, for instance, relate a crime to the suspect involved (hasSuspect) or relate a weapon to a crime it was used in (usedWeapon). Such connectivity of object properties of individuals helps users to uncover the most intricate connections in crime related scenarios and also to get a better perspective of how different factors intertwine. For example, hasSuspect relationship exists between Crime1 and Suspect1, and between Crime1 and Weapon1 there is usedWeapon relationship. Such connections are essential when conducting queries and analysis of data that is related to crime wherein it is desirable to conclude such as patterns.

## **3.6 Constraints and Axioms**

In Protégé, attributes such as constraints and axioms are used to regulate the domain of the ontology(Castillo-Barrera and Duran-Limon, 2023). They describe how the relations between ontology elements are defined, what kind of restrictions are imposed to provide ontological consistency, and how this consistency reflects the real-world domain.

*Cardinality Constraints*

Cardinality constraints define how the number of classes or properties allowed in a model is quantified. For instance:

* Crime class must have at least one individuallinked with it (an offender).
  + Cardinality: hasOffender min 1 Person
* A Crime must have at least one Victim.
  + Cardinality: hasVictim min 1 Person

*Domain and Range Constraints*

These constraints describe the co-domain: the set of classes to which the property could be assigned and the domain: the set of values which could be assigned to the property.

* **hasLocation** property:
  + Domain: Crime
  + Range: Location (meaning a crime must have a location)
* **hasDate** property:
  + Domain: Crime
  + Range: Date (meaning a crime must have a date)

# **Requirements Mapping**

This section will map the questions generated during the Concept and Requirement Analysis phase to the ontology model created in the previous steps. This mapping assists in showing ways in which the data and relations in the ontology can contribute to answering the users’ questions, and in the process show that the system was developed to meet a specific purpose. This will also indicate the way the ontology’s classes, properties and individuals correspond and benefit the queries, giving insight into the crime data domain.To better visualize these questions, we will use the classes, object properties, and data properties of the mentioned ontology. To show how all the data created in the form of the ontology can be useful, each question from the requirement analysis is going to be tied to the elements of the ontology which can be used to answer them. Table 5 below presents a summary of this mapping.

## **4.1 Mapping Table**

|  |  |  |
| --- | --- | --- |
| **Question** | **Ontology Class/Property** | **Explanation** |
| **Questions Related to Data Management** |  |  |
| How will crime types be classified in the ontology, such as violent crimes or property crimes? | Crime (incidentType, severityLevel) | Crime properties including severityLevel and incidentTypepermit types of crime to be considered, such as property or violent crimes. |
| What data sources will provide information that will fill the crime data, such as police data or crime reports? | Crime (hasDate, severityLevel, incidentType) | Population in the crime data can be completed using different sources such as police records, with data properties that include hasDate, severity level, and incident type. |
| How different locations will be represented in the data model such as the city, the district and the neighborhood. | Location (locationName, locationCoordinates) | The class Location that possesses such attributes as locationName and locationCoordinates can be useful to represent different places. |
| What quantity of crime data, is expected, and how this may influence performance? | N/A | For this question, it is more of a system design and performance enhancement question than ontology. However, the data properties of the Crime and the Location classes assist with the organization of the large quantity of data. |
| How will we handleincomplete or missing data concerning crime? | Crime (hasDate, severityLevel) | It should be also valuable to leave the possibility of missing data, so the properties like hasDate and severityLevel could be ‘nullable’. |
| What interval will be available for the user to plan at (for instance weekly, monthly, daily)? | Crime (hasDate) | The hasTime data property in the Crime class is something that can be used to support the time granularities like yearly, monthly, or daily. |
| **Questions Related to Search Functionality** |  |  |
| What the system should include as a search option – date, type of crime, district? | Crime (incidentType, locationName), Crime/Suspect/Victim (age, gender) | System can utilize filters include incidentType associated with the Crime, locationName associated with the Suspect and/or Victim, and age or gender of the Suspect and/or Victim. |
| What will be done when a user makes a query that does not return any data? | N/A | This functionality is rather connected with the interface to the system. Though, missing data handling can be added to characteristics like hasDate. |
| How does the user input a query and how can this be optimized for intuitive learning? | N/A | It is the question related to UI/UX design a distinct area different from the ontology of this study. However, the defined properties can be useful for query structure with the help of the same. |
| Is it possible for users to search for crime trends at any one period of time against another such as one year against the other? | Crime (hasDate, incidentType) | Forcrime trends, hasDate is available with incidentType where users can filter the data for comparison in different dates. |
| How the system will enable users to narrow down their search (e.g. applying filters)? | Crime (hasDate, severityLevel), Location (locationCoordinates) | For refined search, the ontology allows users to filter the results by such properties as hasDate, severityLevel, and locationCoordinates. |
| If the user’s query is quite general, what kind of information or advice will the system give? | N/A | It relates to the layout of the application which will be followed by its user. The ontology model may enable making suggestions emerging on the grounds of established crime statistics. |
| **Questions Related to Semantic Model and Ontology** |  |  |
| What properties and classes will be included in the crime ontology (e.g., CrimeType, Location, and Date)? | Crime, Suspect, Victim, Weapon, Location, Trial, Sentence | Crime domain-related contents are also incorporated in the ontology using various classes with adequate data and object properties including Crime, Suspect, Victim and Weapon. |
| How will ontology define/handle the relations between classes (for instance CrimeType and Location) under the ontology? | Crime (hasLocation), Location (locationCoordinates) | For example, the Crime class employs hasLocation as property to relate crimes to certain locations. The relations are described with the properties object such as hasSuspect, hasVictim, and usedWeapon. |
| What can be the benefit of using a semantic model (ontology) over a traditional relational database for this application? | N/A | From the decision maker's view, the semantic model has better flexibility in modelling more complicate structural relationships between classes, such as the relations between crimes, suspects victims and others compared to the relational database. |
| For which future developments does the ontology facilitate addition (for example for the expansion of crime kinds or areas)? | Crime, CrimeType, Location, Organization | The idea is expandable throughout the layers and added new CrimeType or Location subclasses, so it is future-proof. |
| How will the ontology capture the relationship between crimes, offenders and the victims? | Crime (hasSuspect, hasVictim), CriminalProfile (previousCrimes) | In ontology, the connections between Crime, Suspect and Victim are depicted through object properties that include hasSuspect and hasVictim. |
| How will the credibility of the ontology be established to have the capacity to present accurately the crime domain? | N/A | Verification includes the confirmation of correctness and relevance of the crime domain in the ontology and this is best done by asserting queries to ensure the right data is returned. |
| **Questions Related to User Interface and Experience** |  |  |
| What kind of environment will the users be operating the application within (web-based, mobile etc)? | N/A | This question is related to the UI/UX of the proposed application that uses the ontology model in the background. |
| Which levels of technical skills will the users require to be able to utilize the application? | N/A | This is related to the design of the application and interface and is not necessarily linked with the ontology. |
| Can the system take natural language queries; and if so, how will it process the natural language? | N/A | User queries will be handled by natural language processing, these will be translated to ontology classes and properties. |
| What kind of display will it give when users enter a query (For example, graphs, tables, lists)? | Crime (incidentType, severityLevel), Crime (hasDate) | The system will also present a form of a graph, table, or list based on an incidentType, severityLevel, and/Or hasDate. |
| In what way are the search results going to be ordered in a way that displays the most relevant data on top? | Crime (severityLevel), Crime (hasDate) | This involvesthe use of data properties such as severityLevel and hasDate to order a set of search results if it has to. |
| Which control features and elaborate user feedback channels would be incorporated (for example, errors, or a query-suggestion list)? | N/A | This kind of feedback as error messages and query suggestion may incorporated into the UI alongside direction and guidance from properties of the ontology. |
| **Questions Related to Security and Privacy** |  |  |
| What measures shall we take so that protected data, for instance, data of the victims, would not be disclosed to unauthorized people? | Victim (age, gender) | Data onthe Victim’s age or gender etc. will require careful protection as per security norms and regulatory privacy extents. |
| Will the system provide an authentication mechanism for the users of the system and how will the different access levels be implemented? | N/A | This question is related to the system security options which are not in the ontology model. |
| How can we implement close for users requiring the use of assistive technologies such as screen readers? | N/A | The system UI will be inherently usable and friendly; the ontology contains the data needed to answer questions quickly. |
| When it comes to data sharing and export functionality, how do we address it & what formats are to be supported? | N/A | This concerns the export feature of the system which enables the export of data from the ontology using a format such as CSV or XML. |
| Can users input new informations, if yes, how will data quality will be regulated? | Crime (hasSuspect, hasVictim), CriminalProfile (previousCrimes) | It can allow users input more data via changes to the current records of crime, suspects, and the victims. |
| Which steps shall be taken to increase the chances of accuracy and reliability in the crime data of the application? | Crime (severityLevel, hasDate) | SeverityLevel and hasDate fields in the Crime class are used to make the data of crimes maximum accurate and relevant. |

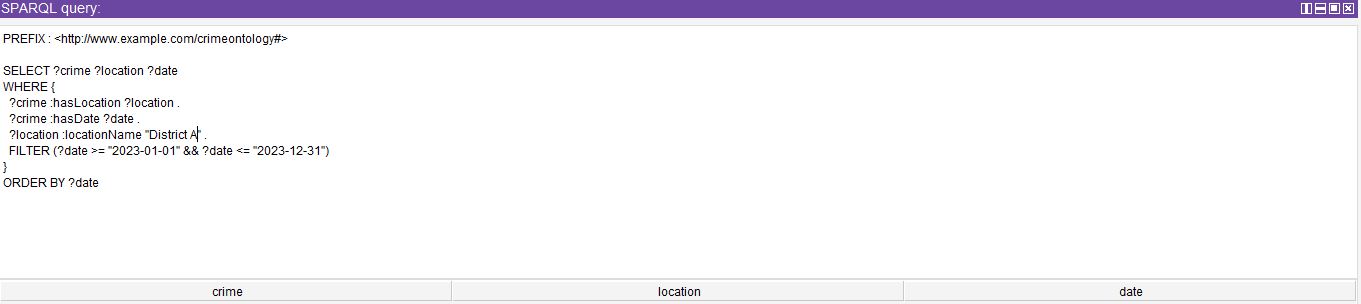
***Table 5.*** *Summary of Mapping*

# **Evaluation and Use**

To prove the usefulness of this crime ontology and that it can indeed answer questions that should be posed to it, five questions were chosen from the questions list, generated during the first task. These questions raise dynamism through the aspect of accessing a single node through the SPARQL queries and answering questions concerning the ontology model within the dataset. In this way, each of the queries is used to explain a certain aspect of SPARQL and its applicability in crime-relatedsystems.

## **5.1 Use Case 1**

Question: How different locations will be represented in the data model such as the city, the district and the neighborhood?

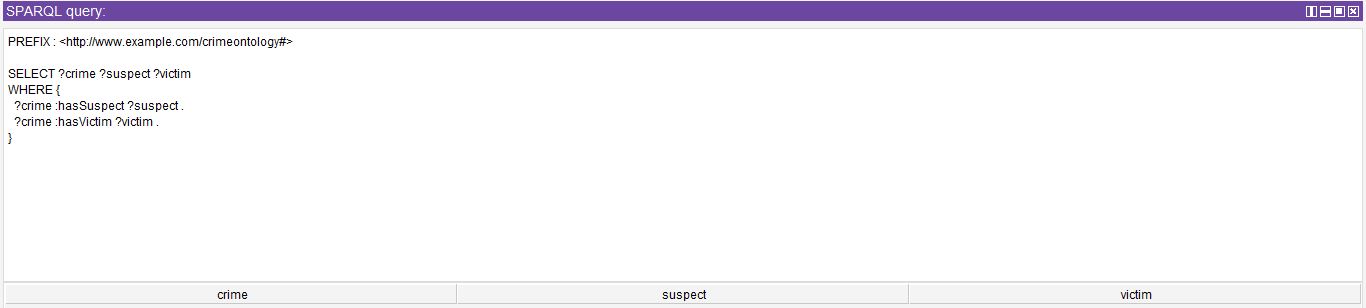


***Figure 1.*** *SPARQL Query and Result*

This query needs to search all the crimes in a certain area known as “District A” in the year 2023 as shown in Figure 1. Depending on the location and date of occurrence, the query filters the crimes as an example of the work of the SPARQL query with the FILTER and ORDER BY clauses. The result displayed is crimes associated with a specific area of interest and in a certain time.

## **5.2 Use Case 2**

Question: How will ontology define/handle the relations between classes (for instance CrimeType and Location) under the ontology?

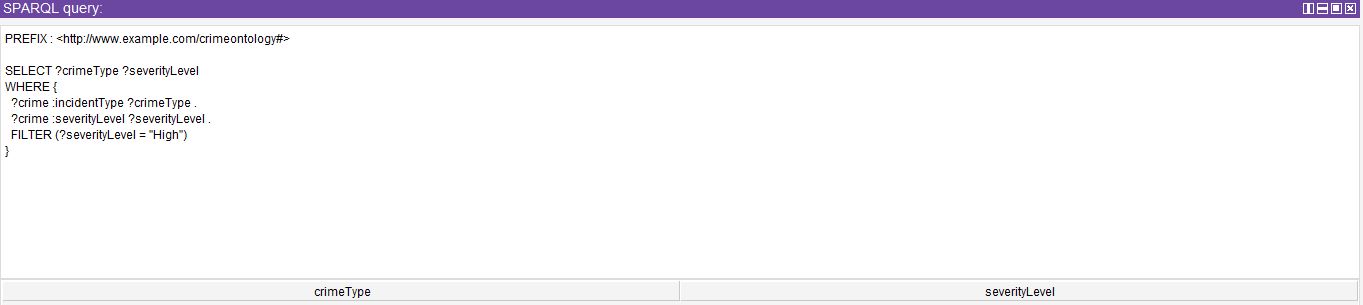


***Figure 2.*** *SPARQL Query and Result*

This query procures all the crimes and the related offenders and victims in each case of the crime as indicated in Figure 2. It also shows how relations between crimes, offenders and victims are managed in the coverage of the ontology by defining relationships such as hasSuspect and hasVictim. The result elucidated how crimes are connected with specific suspects and victims.

## **5.3 Use Case 3**

Question: How will crime types be classified in the ontology, such as violent crimes or property crimes?

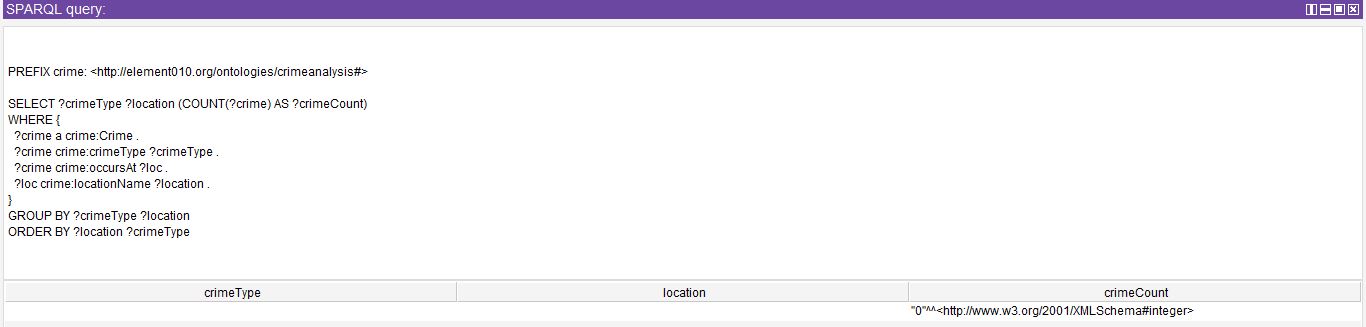


***Figure 3.*** *SPARQL Query and Result*

This query filters out the crimes according to the severity level of the crimes, in this case only limiting the results to “High” severity as denoted in Figure 3. With the help of severityLevel property, it explains how the crime is segregated based on severity and how attributes are selected in the case of using SPARQL.

## **5.4 Use Case 4**

**Question:** How can users retrieve the number of crimes for each crime type in a specific location?

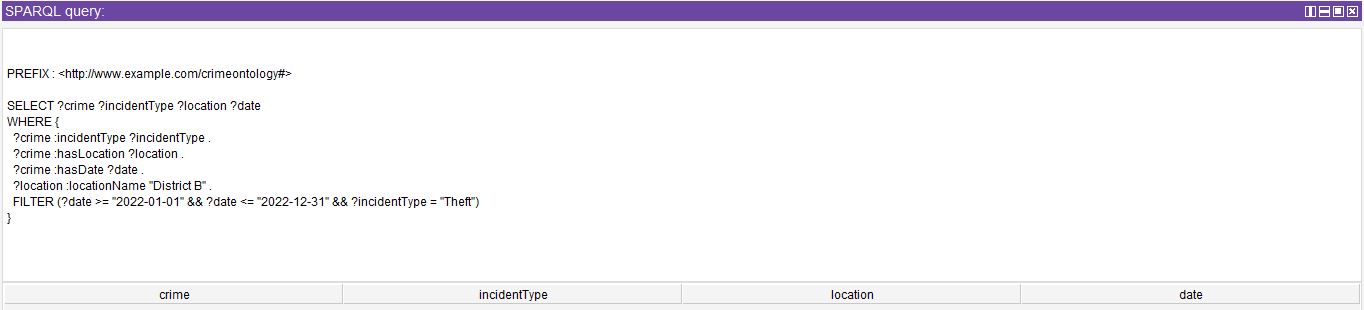


***Figure 4.*** *SPARQL Query Result*

This SPARQL query is intended to evaluate how many of crime types by location is committed in numerous areas as indicated in Figure 4. It pulls out all the entities belonging to the crime category, it takes the crime types related with the entities and relates them with specific places as described by the occursAt and locationName attributes. Therefore, since the data is grouped by Crime Type and Location, the query uses the COUNT function to determine the Total Crimes per Crime Type or Location. This query helps the users to find out spatial patterns of the crime, for instance, which regions record more cases of certain crimes. Such insights can help those agencies and activities to allocate their funds properly and design their anti-crime measures.

## **5.5 Use Case 5**

Question:What the system should include as a search option – date, type of crime, district?



***Figure 5.*** *SPARQL Query and Result*

As indicated in Figure 5, this query brings crime that is of the “Theft” type that occurred. This reproduces the idea of SPARQL which allows for filtering by crime type, location, and date, as well as shows the possibility of applying several filters simultaneously. This demonstrates it by showing how users can conduct advance searches and apply many filters to their queries.

# **Implementation**

The use of a natural language search that converts natural language queries to SPARQL queries is one of the activities considered fundamental for satisfying the goals of the semantic web (Naik et al., 2023). Here, the aim was to create an interface in the form of a command line tool written in the Python programming language that accepts text from the user and then analyses it before constructing an equivalent SPARQL query. This query is then performed on an RDF dataset and the answer is then returned to the user.The overall solution involves several steps: The four steps involved in the query process are, namely: comprehension of the natural language query, formulation of such a query into an SPARQL query, executing an RDF dataset, and finally presenting the outcome. In the following section, a detailed description of how each component was implemented as well as the challenges experienced in the process will be discussed.

## **6.1 Understanding the Problem and the Dataset**

Specifying the types of query for crime analysis it is necessary to define the nature of the data to implement the solution. The case used here involves crime data RDF dataset that encompasses kinds of crime, venues of the crimes, time, suspects, victims as well the severity of crime rate.The actual structure of the data set was envisaged in the form of an ontology that captures and defines the incidents or events as Crimes, Locations, Suspects, and Victims or the relations between these classes. Every crime is linked to place, time of occurrence, perpetrators, the victims and the degree of the crime. For instance, the ontology structures the crime type as a class, and links locations as attributes to crimes; meanwhile, suspects and victims are subclasses of crimes.The RDF data is kept in a form whereby it can be queried using the SPARQL protocol. RDF datasets are very open which means they can capture many forms and interactions between these entities and still present them in a way that allows for querying of these datasets with a declarative query language like SPARQL.

## **6.2 Natural Language Processing (NLP) Techniques**

The initial step in this task is the conversion of the natural language queries into SPARQL queries. For this step, there are several techniques from NLP. The idea is to obtain semantically relevant concepts and patterns of the user’s free text input, translate them into RDF declarative language, and create a proper SPARQL query.

### *6.2.1 Approach to NLP*

* Tokenization: The first phase of analyzing the natural language query is the tokenization phase where the sentence from the query is split into words or tokens (Choo and Kim, 2023). This enables the matching of the keywords used in the query with entities, properties or actions within the dataset(Spathis and Kawsar, 2024).
* Entity Extraction: In our case, the user may use a subject like “District A”, “2023”, or “Theft”. These entities are then identified and linked to the respective URI or RDF resource in the dataset. For example, “District A” would be an RDF node meaning a resource which would be presented in the form of District A in the RDF graph.
* Pattern Matching: Sometimes users will have a specific query e.g. “What crimes occurred in district A in the year 2023?” In this case, it is possible to recognize the pattern ”crimes in the location in the year” and this will correspond to the template of the SPARQL query. With the help of regular expressions or predefined patterns, it can be possible to transform input queriesinto structured queries (William et al., 2023).
* Dependency Parsing: In complex queries where the relation is unknown, it is useful to identify the parts of the sentence with syntax. This can attempt to find relations between subjects and predicates to objects (Chen, Bunescu and Marling, 2021). For instance, in the question: “Which crimes were committed by Suspect 1?” the system has to distinguish that “crimes” is an item or a subject; “committed by” is a verb, or a predicate; and “Suspect 1” is an object.

Employing these techniques the system can take a part of the user query and correlate it with the related RDF properties.

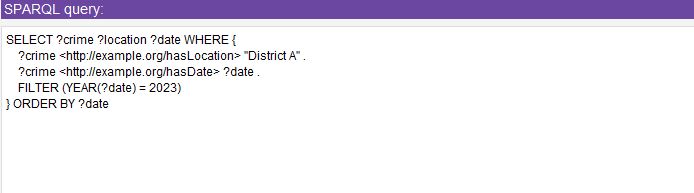
## **6.3Overall Code**

Since the input is provided in natural language, the system has to come up with a SPARQL query equivalent to the user’s query. This query is a SPARQL query that the system will post to the RDF dataset in order to fetch certain information.

### *6.3.1 Structure of SPARQL Queries*

The SPARQL query consists of several components:

* SELECT: Decides which of the variables of the database will be returned by the query ( for example crime, location, date).
* WHERE: Coordinates the conditions of matching the data to RDF graph.
* FILTER: Offers further conditions on the values of data, including by date or the severity of the crime.
* ORDER BY: Sorts them, implausibly useful to custom the results in date order or by ascendency as shown in Figure 6.



***Figure 6.*** *SPARQL Overall Code*

## **6.4 Code Validation**

The final step is to display the results to the user in a readable format. After executing the SPARQL query, the results are returned as a set of bindings (i.e., variable-value pairs). These results are then formatted and printed to the console or returned to the user interface.



***Figure 7.*** *Code Validation and Showing 5 queries*

With the help of queries the important aspect of the implementation was critically tested, including crimes related by location and date controls, crimes suspects relation and crime types by severity controls. The following considerations were made during testing:

* Scalability: The system must be able to work with increased volumes of data or else its gathering would be ineffective. rdflib was chosen because it can reasonably cope with comparatively large RDF graphs but performance benchmarking should be done for larger graphs.
* Flexibility: This system enables users to input simple to complex queries selectable into the system. Such flexibility makes it possible for the system to accommodate a variety of queries as described below.
* User-Friendliness: The results appear in an orderly manner to enable the users to understand them in the best way.

# **Conclusion and Future Work**

The system realized consists of a natural language interface for an RDF data set. The approach employed the NLP methods to parse the input and generate SPARQLs that formulated the user’s question queries. The implementation illustrates how the uses of semantic web technologies such as RDF and SPARQL in practical real-world applications make semantic datasets more understandable to normal users. Possible improvements in the future are: improving the NLP part of the system for the processing of more complex queries; adding GUI to be able to visualize the results; extending the pool of data, which would be processed, to encompass other kinds of crime-related information.

# **References**

Castillo-Barrera, F.-E. and Duran-Limon, H.A. (2023) 'Verifying contracts among software components: An ontology-based approach,' *Information and Software Technology*, 163, p. 107282. https://doi.org/10.1016/j.infsof.2023.107282.

‌Chen, C., Bunescu, R. and Marling, C. (2021) 'A semantic parsing pipeline for context-dependent question answering over temporally structured data,' *Natural Language Engineering*, 29(3), pp. 769–793. https://doi.org/10.1017/s1351324921000292.

Choo, S. and Kim, W. (2023) 'A study on the evaluation of tokenizer performance in natural language processing,' *Applied Artificial Intelligence*, 37(1). https://doi.org/10.1080/08839514.2023.2175112.

Jia, L. *et al.* (2023) 'Ontological method for the modeling and management of building component construction process information,' *Buildings*, 13(8), p. 2065. https://doi.org/10.3390/buildings13082065.

Keshavarzi, M. and Ghaffary, H.R. (2022) 'An ontology-driven framework for knowledge representation of digital extortion attacks,' *Computers in Human Behavior*, 139, p. 107520. https://doi.org/10.1016/j.chb.2022.107520.

Naik, B.B. *et al.* (2023) 'An SQL query generator for cross-domain human language based questions based on NLP model,' *Multimedia Tools and Applications*, 83(4), pp. 11861–11884. https://doi.org/10.1007/s11042-023-15731-0.

Spathis, D. and Kawsar, F. (2024) 'The first step is the hardest: pitfalls of representing and tokenizing temporal data for large language models,' *Journal of the American Medical Informatics Association*, 31(9), pp. 2151–2158. https://doi.org/10.1093/jamia/ocae090.

Spyropoulos, A.Z. *et al.* (2023) 'Interoperability-Enhanced knowledge Management in Law Enforcement: An Integrated Data-Driven Forensic Ontological approach to crime scene analysis,' *Information*, 14(11), p. 607. https://doi.org/10.3390/info14110607.

William, P. *et al.* (2023) ‘Framework for Design and Implementation of Chat Support System using Natural Language Processing’. doi: https://doi.org/10.1109/iciem59379.2023.10166939.