Spatial and Temporal Crime Analysis using Semantic Methodology

Rohan Mathur, Sumeet Suryawanshi, Akash Rana, Sadhanand Srinivasan, Shreya Thummalapalli
School of Computing and Augmented Intelligence
Arizona State University, Tempe Arizona
$\{rmathu19, ssurvaw5, akrana1, ssrin172, sthumm14\}$@asu.edu

***Abstract*—This paper addresses the complex problem of urban crime by combining comprehensive crime data in New York City and Los Angeles into an integrated ontology-based knowledge graph. This research shows how geographic, demographic, and temporal aspects of crime data can be conceptually modeled using semantic web technologies as a platform for advanced querying and pattern recognition. This will also help stakeholders identify the high-risk areas, demographics of perpetrators and victims, and spatial and temporal trends in crime across the two cities. This will further be enhanced with an OWL-based ontology with SPARQL querying capabilities for more actionable insights to support decisions of law enforcement and policymakers in resource allocation, crime prevention, and addressing social inequities in urban violence. By utilizing SPARQL querying in GraphDB, we enable efficient retrieval of relevant insights to support crime analysis, demonstrating the potential of semantic technologies**

***Index Terms*—crime analysis, shootings, ontology, RDF, Semantic Web, GraphDB, react.js, node.js, knowledge graph, data, spatial analysis, temporal analysis.**

I. Introduction

Gun violence in urban cities remains one of the public safety problems that are persistent to date. Major American cities such as New York and Los Angeles, each have high volumes of shooting and overall crime. These are often bound within neighborhoods and affect specific demographics. While New York and Los Angeles have a large amount of available crime data, the volume and complexity of these datasets[9][10] make certain insights challenging. Deep trends such as demographic disparities or spatial clustering might be lost because of the huge volume and less organized nature of the information by using traditional data analysis methods. This paper provides a semantic, ontology-based solution for integration and analysis of crime data from New York City and Los Angeles. This approach converts the data into RDF triples and an OWL ontology that allows for higher-order querying across geographic, demographic, and temporal dimensions. These queries will allow the user to identify high-risk areas, study demographic patterns, and analyze in detail the relation between the types of crimes and their locations. This will form a basis on which informed decisions can be made, such as law enforcement decisions related to resource allocation and crime prevention.Our application uses React for the front end, providing an interactive interface for exploring crime data insights, and Node.js for the backend, which manages SPARQL queries to GraphDB. The Node.js backend directly interfaces with GraphDB, executing SPARQL queries to retrieve RDF-structured crime data across geographic, demographic, and temporal dimensions.

II. Problem Definition

Shootings and crimes across New York City and Los Angeles which affects a varied number of communities differently reflect concerns about public safety. Crimes are tracked in both cities within dense databases which log incident location, time of incident, and demographic information on victims and suspects. The data, however, is complex at such volume that it often does not enable meaningful insight toward preventing crimes and developing policies. This work intends to bridge this gap by establishing relationships among crime data elements of both cities using ontology-structured models. This application aims to integrate geographic locations, demographic profiles, types of crimes, and temporal aspects into one OWL ontology for deep analytics and querying. It will also allow identification of high-risk areas, and provide better understanding of disparities between neighborhoods, which will help in resource allocation and effective crime prevention strategies.

A. Solution Overview

We will implement a semantic web application that enables users to visualize the shooting and crime data of New York City and Los Angeles on one common knowledge graph. These incidents will be integrated with the geographical, demographical, temporal dimensions and data that are of interest to the stakeholders for comparison of crime trends across cities.

This application will help law enforcement agencies and policy makers by converting raw data to RDF triples and hosting a knowledge graph in the cloud. The application shall further support SPARQL-based advanced querying, enabling users to investigate the relations, patterns, and trends that are not directly evident when raw data are analyzed.

B. Use Cases

- Identification of High-Risk Areas: This would allow the user to visualize areas with high risks in New York City and Los Angeles. Using the semantic model derived, it would bring forward geographic hotspots of shootings and other serious crimes for proper resource allocation by decision-makers in law enforcement.

- Perpetrator and Victim Demographics by Region: This would provide analysis of perpetrator and victim demographics, including race, age, gender, and ethnicity, for different regions in New York City and Los Angeles. It will help determine the demographic inequity within crimes and shooting incidents.
- Temporal Patterns of Crime A temporal analysis feature will provide the user with the capability to understand at what times crimes might be expected to happen in each city. This may include seasonal trends, patterns of high-crime periods that will be very helpful for law enforcement agencies in making predictions and doing proper preparations in advance. Integrating date and time factors, including holidays and weekends, enables the ontology to reveal time-based trends in crime occurrences. By identifying peak times for certain crimes, authorities can optimize resource allocation and anticipate higher-risk periods.
- Correlation Between Crime Types and Locations: It would enable the user to further investigate how these different classes of crimes-assaults and shootings, property crimes-relate to one another in the same location or various locations. By comparing and contrasting the data from New York City to that of Los Angeles, the application user could find patterns in crime repetition that might be similar or different in the two cities.
- Spatial Proximity of Crimes: This dataset will enable the users to query the geographic coordinates of the nearest shooting/crime in both cities. This will give the law enforcement agencies an idea of which areas are hot spots of crime which will enable law enforcement for better resource allocation to these areas.

III. Related Literature

Recent work in crime analysis has noted the increasing need for data-driven approaches in order to address urban violence in complex environments like New York City and Los Angeles.

One study introduced an ontology-based system using the Resource Description Framework for performing hate crimes and shooting analysis [2]. The approach has placed special emphasis on the strengths of ontology-based systems for actionable insight analysis in law enforcement and policy and improves crime pattern visualization in New York City, focusing on gender-based crime patterns, temporal distribution, and demographic disparities in shootings.

Another significant contribution is an agent-based modeling framework that compares the impact of several interventions on firearm-related homicide in New York City [3]. This study compares high-risk interventions. Results suggest that a combination of targeted and population-wide interventions yields the most significant reduction in firearm homicides, highlighting the potential of simulation modeling as a valuable tool in evaluating crime prevention strategies.

This study applied machine learning to analyze borough-specific drivers of gun violence in NYC and identified a set of socio-economic variables, such as poverty levels, that varied across districts with different influences on patterns of violence [4]. It found that violence does not spread evenly over the city, which means that borough-specific trends are necessary in crime prevention. This study uses machine learning to yield district-level facts that provide evidence to guide targeted interventions at the grassroots level.

A space-time analytics approach to NYC shooting incidents further identifies spatiotemporal clusters that correlate with demographics and infrastructure, such as the presence of Black populations and vacant housing units [1]. Regression and network analyses show that gun violence tends to rise with distance from police stations, revealing gaps in policing coverage. These findings indicate that social and spatial factors play an essential role in predicting shooting patterns and could help authorities enhance resource allocation and policing strategies to address violence hotspots.

Another analysis examines how the COVID-19 pandemic and associated social-distancing mandates influenced NYC violence patterns, particularly during the period of unrest following George Floyd's death [5]. Using time-series analysis, it shows that boroughs with varied socio-economic and racial demographics experienced different impacts, indicating that crises can deepen existing inequalities in urban violence. This paper highlights the potential of spatial analysis for understanding how crises can cause disproportional rises in violence across diverse urban areas, with implications for future crisis management and resource allocation.

Spatial and statistical analysis applied in one study on gun violence in Philadelphia, New York, and Los Angeles during the COVID-19 pandemic found concentrated increases within "hot spots" with disproportionate impacts within Black and Hispanic communities [6]. These accounted for 36% of Philadelphia's, 47% of New York's, and 55% of Los Angeles's shooting increases. Although poverty and segregation are contributing factors, neither thoroughly explains the block-level clustering that underlies these patterns, highlighting a role for focused, place-based intervention.

While existing studies have employed a variety of techniques such as ontology-based RDF frameworks for trend analysis, machine learning for district-specific violence patterns, and space-time analytics for identifying crime clusters, our approach builds upon these by integrating both New York City and Los Angeles crime data into a unified RDF-driven OWL ontology. This enables the comparison between cities with deeper semantic integration for the opening of more advanced querying and insight on different urban contexts. We help stakeholders utilize a knowledge graph for pinpointing high-risk areas and demographic disparities in real-time [7].

IV. Approach and High-level System Design

The project workflow aims to integrate and analyze crime data from multiple sources, such as NYPD Shooting Incident Data and Los Angeles Crime Data. Each dataset undergoes cleaning and transformation to build a consistent and unified dataset that can enable meaningful analysis [8]. After that, the integrated crime dataset is modeled as an OWL ontology by utilizing Protégé. This ontology will represent the entities, relationships, and concepts related to crime data in a structured way. Real-world instances are populated into the ontology, resulting in an RDF file that will allow semantic queries. The RDF data is stored in graphDB, which is SPARQL-compatible, to enable efficient querying. SPARQL is a query language designed for RDF data. This enables data filtering, enhancing the system's analytical capabilities. The graphDB is made available through RESTful API. This API serves as an interface between the frontend and the backend logic where data is processed and SPARQL queries are executed. The frontend consists of a React.js application, whereas the backend is implemented using Node.js [9]. A user can request data, and the application will filter information and present the results intuitively. The React application calls the back-end API, requesting data. Meanwhile, the backend processes these requests for the data retrieves relevant data from the graphDB and sends those back to the frontend. Integrating data processing with ontology modeling, graph database, and API-driven communication gives users an integrated system with a seamless experience to explore and query crime data from many sources [10].

V. Ontology design and a visualization

The ontology of this project provides a structured way to capture key information related to crime incidents, together with their contextual information, that supports the analytical requirements of our project. The major classes in the ontology include 'Crime', 'Location', 'Person', 'Jurisdiction', and 'PremiseType'. Every crime incident has been uniquely identified with 'incidentID' and contains specific temporal and spatial information. The Location class contains subclasses such as Jurisdiction and PremiseType. This allows the model to capture the exact classification of the crime scene. For example, the Jurisdiction class contains properties like 'inZone' and 'hasCity', which describes the city and the zone of the crime location, while under PremiseType, there is the description of the location type, which could be a 'residential', 'commercial', 'street', etc. This enables a detailed analysis of the type of locations where crime occurs frequently. The ontology defines object properties, such as 'involvesVictim' and 'involvesPerpetrator', relating 'Crime' class to individuals involved in incidents. The Person class (superclass of Victim and Perpetrator) has demographic attributes, such as 'Sex', 'AgeGroup', and 'Race', providing in-depth crime demographic analysis. The Time class encapsulates temporal details with attributes such as 'Hour', 'Minute', and 'isHoliday', providing details of when the crime incident occurs. Object properties such as 'occurredDuring', 'hasLocation', 'hasJurisdiction', and 'hasPremiseType' link Crime to its particular context, thus providing a semantic framework.

VI. About the Dataset

The data for our use cases have been acquired from two different datasets. These give us critical insights into crime-related incidents. The datasets will allow a comprehensive and proper understanding of shooting incidents and broader crime trends in different geographical areas.

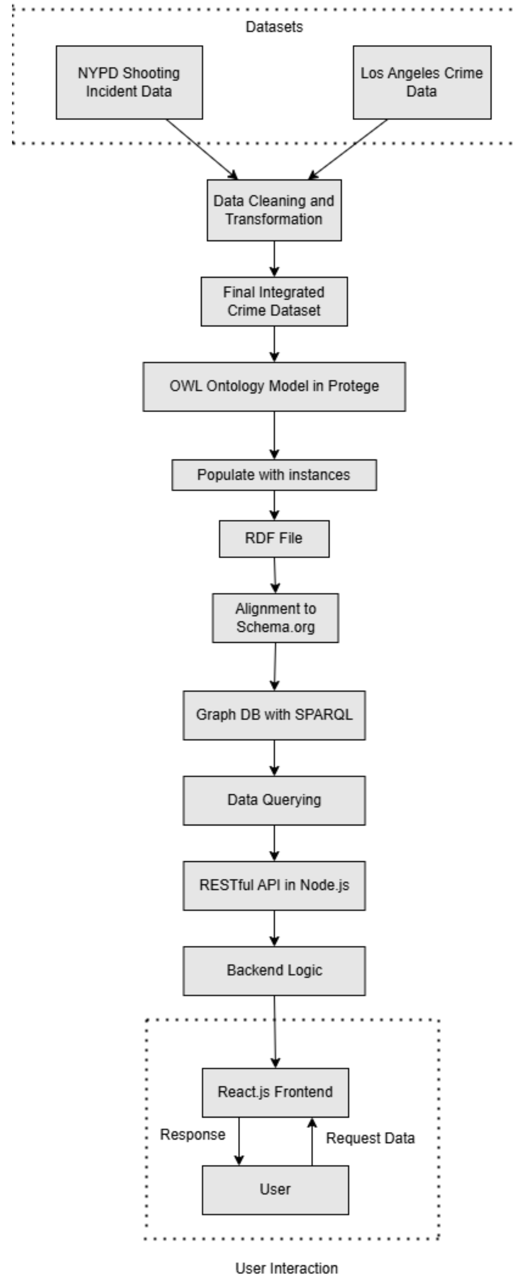

Fig. 1. Workflow Diagram

The first dataset is the NYPD Shooting Incident Data [13], which presents detailed information on shooting incidents across the New York City. The dataset contains 21 columns and 28.6k rows, which include all the necessary attributes, with the Incident Key being a unique identifier generated randomly for each shooting incident. Key columns also include 'OCCUR_DATE' and 'OCCUR_TIME' for the exact date and time of the incident, 'BORO' for the borough of occurrence, and 'PRECINCT' for the exact location and jurisdiction of the event. Further categories in this data include 'PERP_AGE_GROUP', 'PERP_SEX', and 'PERP_RACE' for perpetrators, with corresponding age and sex information for victims, 'VIC_AGE_GROUP' and 'VIC_SEX'. Another critical parameter is the 'STATISTICAL_MURDER_FLAG', indicates that a fatality occurred in the incident. All these provide important contexts useful in analyzing shooting incident patterns in various aspects.

The second dataset comes from the LAPD Crime Data [14], giving a wider perspective on the broad categories of crimes recorded in Los Angeles. Each crime is identified with a 'DR_NO', which stands for Division of Records Number, serves as the unique identifier for the incident. Other parameters also include 'DATE OCC', representing the date of occurrence, and 'TIME OCC', indicating time in 24-hour format. The geographical details are derived through 'AREA' and 'AREA NAME', which describe the division and location in Los Angeles where the crime was committed. 'Rpt Dist No' allows further regional breakdowns. Demographic details of victims include 'Vict Age', 'Vict Sex', and 'Vict Descent', among other specifics such as 'Crm Cd Desc' (description of the code for crime) and 'Mocodes' (methods of operation associated with the crime). This dataset adds depth to our analysis by providing a comparative perspective of the crime trends in a major city by location, demographics, and classification of crime.

Data preprocessing is crucial because these datasets contain a lot of missing or inconsistent values. Data Cleaning is necessary. Therefore, we removed unnecessary columns or duplicates that did not add value to our use case, dropping the rows with missing values in those critical columns to maintain data integrity. For example, the categorical variables of age groups were coded using integers for efficient querying and data linkage across datasets. A Python script was utilized to perform all these tasks and libraries such as Pandas and NumPy were used to improve the datasets. Apart from all the above steps, more preprocessing will be done to match the datasets and improve the analysis. We added a 'holiday' field to both datasets, which can identify if the crime incident falls on a holiday. We standardized the race categories for both to ensure consistency in demographic analysis.

Similarly, premise types and time formats were standardized to provide consistent location and time information. 'Age group' had to be normalized to match the datasets for easy integration and comparison. These additional steps ensure the optimization of both datasets for the most accurate and consistent analysis of all use cases.

VII. Implementation

A. Ontology Design and Enrichment:

We have designed a comprehensive ontology using the Protégé platform, which will accurately model crime-related data. We have defined those classes, properties, and relationships representing real-world entities and their relations. We ensured that the ontology was in conformance with the needs of the project and, at the same time is capable of doing various representations of data to provide semantics. We also included domain-specific enrichments necessary for capturing subtlety in the crime data to offer more refined analysis.

B. Data Cleaning and Standardization:

After the development of the ontology, extensive cleaning and preparation of data in Python was performed: irrelevant column removals, handling missing values, and standardization of time formats, age groups, and racial classifications to have homogeneous data. Addition of 'holiday' field in both datasets to trace whether a crime happened during a holiday or not. Pre-processing is done to get the data to align with the ontology and prepare it for its conversion to RDF.

C. RDF Conversion:

After refining the datasets, we need to convert the data into RDF format using Protege. We have to perform these RDF transformations, where tabular data will be transformed into RDF triples to adhere to our ontology structure. During this process, we can exploit transformation rules to create semantic linkages between perpetrators, victims, locations, and time information. It will result in RDF triples capturing these associations, which is key to linking data analytics.

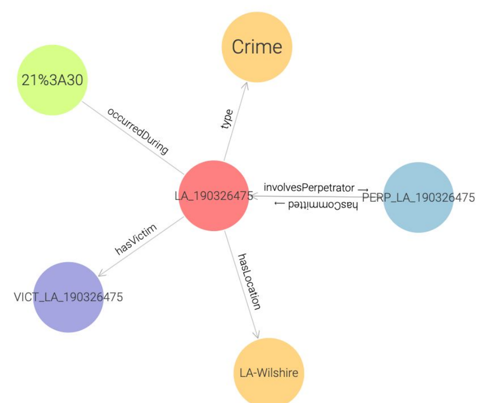

Fig. 2. GraphDB Representation

D. Storing in GraphDB:

GraphDB will be used as the triple store to manage and query the RDF data. This setup will provide a well-founded platform for semantic data storage that shall provide smooth querying and retrieval. Structured RDF data will be loaded into GraphDB, making it accessible to SPARQL queries, as shown in Fig. 2. This setup ensures an optimum performance output in data storage, retrieval, and management to address the analysis requirements of the project effectively.

E. Backend API Development with Node.js:

We plan to develop a Node.js backend API that can interface with the triple store provided by GraphDB. In turn, it would facilitate querying via SPARQL for data retrieval in an efficient manner. All these endpoints should be organized to access the data in a consumable manner. Further, we plan to add relevant API documentation so that users understand how to craft queries for retrieving data relevant to specific use cases. This provides the backbone of data access and interacts seamlessly with the triple store.

F. Frontend Application using React.js:

The final step is to create a frontend web application to visualize and explore the data efficiently. In that respect, a modern JavaScript framework will be utilized. It will be an easy-to-use interface for the end users who can easily explore some of the predefined use cases and present the results of the analytics in various forms, such as charts, graphs, or maps. This application will transform comprehensive data into understandable and actionable insights to support informed decision-making by stakeholders.

VIII. Data Collection and Processing

The analysis of spatial and temporal crime patterns required the use of two distinct datasets: the NYC shooting incidents dataset, which contained approximately one million rows, and the LA crime dataset, with 28,500 rows. We used extensive preprocessing and augmentation to enable effective comparison and analysis to these datasets.

A. Data Collection

The NYC dataset provided detailed records of shooting incidents, while the LA dataset offered general crime data. Both datasets were modified by adding a unique identifier field, ”crimeDrNo”, which includes the prefixes ”NY” for NYC and ”LA” for Los Angeles. This ensured clear differentiation between the two sources of data.

B. Data Processing

1) LA Crime Dataset: Significant modifications were made to the LA crime dataset to enhance its utility for the analysis. A new crimeDrNo field was added to uniquely identify each record and align it with the NYC data structure. Fields for victimId and perpId were derived from the crimeDrNo, as the original dataset lacked explicit victim and perpetrator identifiers. Additionally, a ”city” column was added to specify the location. Time data field was refined by adding hours and minutes columns to extract and display the occurrence time of crimes more clearly. A new location field was derived from the areaName column, grouping crime sites into categories such as ”street,” ”transit,” ”commercial,” and ”dwelling.” To align with the NYC dataset, a victimAgeGroup field was created based on the existing age column. Furthermore, a holiday column with boolean values (true or false) was added to indicate whether the crime occurred on a holiday.

2) NYC Shooting Dataset: The NYC shooting dataset also required extensive preprocessing. An incidentKey field was introduced for consistency with the LA dataset, and fields for victimId and perpId were derived from the crimeDrNo. The time of occurrence was divided into separate occurTime and occurHour columns to facilitate detailed temporal analysis. A location field was created based on borough information, categorizing incidents by area, and a city column was added to specify New York City as the location. To include additional spatial granularity, a precinct field was introduced. A victimDissent field was added to align with the victimDissent information available in the LA dataset. Similar to the LA crime dataset, a holiday column was included to denote crimes occurring on holidays.

C. Data Cleaning and Enrichment

Extensive data cleaning was conducted across both datasets to ensure high-quality and reliable analysis. Missing or inconsistent values were addressed, and new columns were derived to standardize temporal and spatial data. For example, columns for hours and minutes made it easy to categorise the time data based on specific times of the day, while adding location field helped classify crime sites effectively. Aligning fields such as victimAgeGroup and victimDissent ensured compatibility between datasets, making cross-city comparisons meaningful.

The enriched datasets were thus prepared to facilitate detailed spatial and temporal crime analysis, leveraging a semantic methodology to uncover patterns and insights.

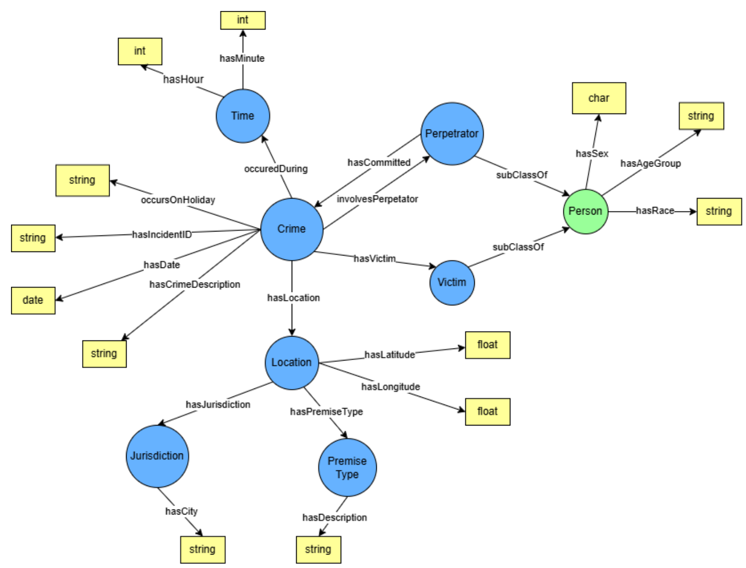

Fig. 3. Knowledge Graph

IX. Ontology Design

The ontology design for this paper was developed using the Protégé tool to construct a comprehensive schema for the representation of crime data. It incorporates an organized hierarchy of classes, object properties, and data properties to allow detailed semantic relationships among entities that pertain to crimes, as ahown in Fig. 3. Further, it is aligned with schema.org standards to make it compatible and extensible.

A. Classes

The ontology defines several key classes to represent different aspects of the crime data. These include Crime, Victim, and Perpetrator, where both Victim and Perpetrator are modeled as subclasses of the Person class. Other classes include Location, Time, PremiseType, and Jurisdiction. Each class encapsulates specific attributes and relationships relevant to its domain, contributing to a holistic representation of the crime schema.

B. Object Properties

Object properties establish the semantic relationships between classes, enabling the ontology to represent complex associations within the crime dataset. For instance, the Crime class is linked to Location via the hasLocation property and to Perpetrator and Victim through the involvesPerpetrator and hasVictim properties, respectively. The relationship between Crime and Time is captured using the OccurredDuring property. Additionally, the Location class connects to Jurisdiction and PremiseType through the hasJurisdiction and hasPremiseType properties, respectively. The Perpetrator class is associated with Crime through the hasCommitted property. These object properties provide a clear and consistent framework for describing interactions between entities, enhancing the ontology's ability to model real-world crime scenarios.

C. Data Properties

Data properties are utilized to capture granular details about entities within the ontology. For the Crime class, attributes such as hasDate, hasIncidentID, and occursOnHoliday provide essential temporal and identification details. The Location class includes hasLatitude and hasLongitude to specify geographic coordinates, ensuring spatial accuracy. For demographic and descriptive purposes, the Perpetrator class is described with the hasAgeGroup property, while the Victim class includes hasRace, hasSex, and hasAgeGroup to provide detailed victim profiles. The Precinct class includes the hasCity property to associate it with its respective urban area, while the PremiseType class includes a hasDescription property to categorize crime locations. The Time class incorporates hasHour and hasMinute properties, enabling precise temporal analysis. These data properties facilitate the representation of specific, measurable characteristics, ensuring robust and detailed data modeling.

```
<!-- http://schema.org/AdministrativeArea -->

<owl:Class rdf:about="http://schema.org/AdministrativeArea"/>


<!-- http://schema.org/Event -->

<owl:Class rdf:about="http://schema.org/Event"/>


<!-- http://schema.org/Person -->

<owl:Class rdf:about="http://schema.org/Person"/>


<!-- http://schema.org/Place -->

<owl:Class rdf:about="http://schema.org/Place"/>
```

Fig. 4. Alignemnt with schema.org

D. Alignment with Schema.org Standards

Alignment with Schema.org Standards To ensure compatibility with widely recognized semantic web standards, the ontology aligns its structure with schema.org, as shown in Fig 4. and Fig. 5 Specific mappings include

```
<!-- http://www.semanticweb.org/ontologies/2024/10/crime-ontology#hasJurisdiction -->

<owl:ObjectProperty rdf:about="http://www.semanticweb.org/ontologies/2024/10/crime-ontology#hasJurisdiction">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/ontologies/2024/10/crime-ontology#Location"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/ontologies/2024/10/crime-ontology#Jurisdiction"/>
    <owl:equivalentProperty rdf:resource="http://schema.org/administrativeArea"/>
</owl:ObjectProperty>


<!-- http://www.semanticweb.org/ontologies/2024/10/crime-ontology#hasLocation -->

<owl:ObjectProperty rdf:about="http://www.semanticweb.org/ontologies/2024/10/crime-ontology#hasLocation">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/ontologies/2024/10/crime-ontology#Crime"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/ontologies/2024/10/crime-ontology#Location"/>
    <owl:equivalentProperty rdf:resource="http://schema.org/location"/>
</owl:ObjectProperty>


<!-- http://www.semanticweb.org/ontologies/2024/10/crime-ontology#hasPremiseType -->

<owl:ObjectProperty rdf:about="http://www.semanticweb.org/ontologies/2024/10/crime-ontology#hasPremiseType">
    <rdfs:domain rdf:resource="http://www.semanticweb.org/ontologies/2024/10/crime-ontology#Location"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/ontologies/2024/10/crime-ontology#PremiseType"/>
</owl:ObjectProperty>
```

Fig. 5. Properties Aligned

Crime aligned with http://schema.org/Event and Location with http://schema.org/Place. Similarly, Jurisdiction corresponds to http://schema.org/AdministrativeArea, and Time to http://schema.org/Time. These mappings ensure a standard interpretation of core entities and their roles within the schema.

Object properties such as hasLocation and OccurredDuring are mapped to http://schema.org/location and http://schema.org/startTime, respectively, while data properties like hasDate, hasIncidentID, and hasLatitude align with http://schema.org/datePublished, http://schema.org/identifier, and http://schema.org/latitude. This alignment guarantees semantic interoperability and supports meaningful integration of diverse datasets.

Demographic attributes such as hasRace, hasSex, and hasAgeGroup are aligned with http://schema.org/ethnicity, http://schema.org/gender, and http://schema.org/age, respectively. These mappings enhance the ontology's capacity to represent detailed and accurate demographic information. By adhering to schema.org best practices, the resulting ontology is a flexible, extensible framework capable of supporting advanced spatial and temporal crime analysis.

X. Querying with SPARQL

The SPARQL queries used in our project address various analytical use cases, providing insights into spatial and temporal crime trends. These queries enable the identification of high-risk areas by analyzing crime frequency within specific regions, while offering demographic analysis of victims based on race, age group, and sex, wihch is segmented by geographic regions. Temporal patterns are explored by comparing the incidence of crimes on holidays versus non-holidays and analyzing the distribution of crimes during specific hours of the day, divided into four six-hour intervals. Queries also categorize crimes by premises type, allowing for the assessment of crime distribution across residential, commercial, and transit areas. By leveraging these SPARQL queries, the project extracts meaningful patterns and trends from RDF data, forming a strong foundation for understanding the dynamics of crime. This query mechanism supports the core use cases and demonstrates the utility of semantic methodologies in addressing complex real-world challenges.

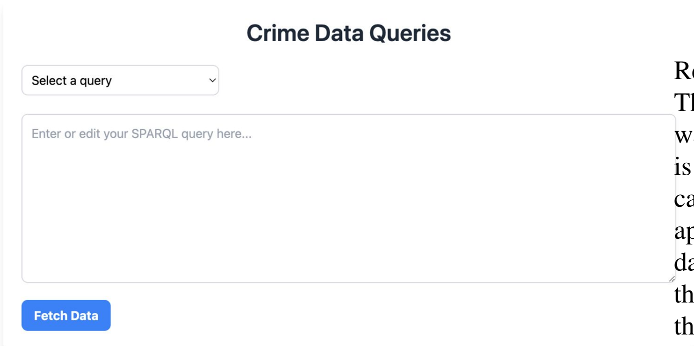

Fig. 6. Application UI

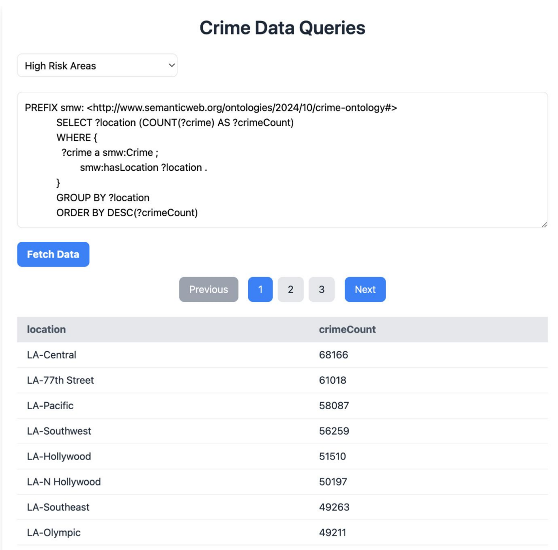

Fig. 7. Query showing results of High Risk Areas

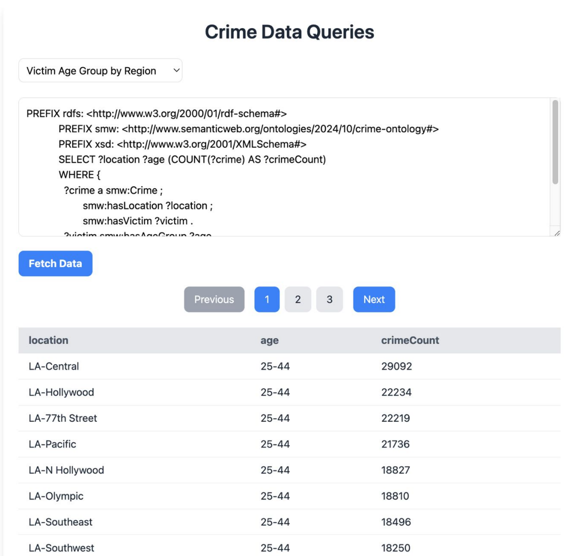

Fig. 8. Query showing victims by age

XI. UI Design

The user interface for the application was developed using React for the front end and Node.js for backend integration. The application is designed to provide an intuitive and efficient way for users to interact with crime data. A central UI feature is a dropdown menu that allows users to select predefined use cases, as shown in Fig. 6. Based on the selected use case, the application dynamically generates a SPARQL query, retrieves data from the GraphDB database, and displays the results on the front end. The following use cases are implemented within the application:

- Identifying High-Risk Areas: This feature enables users to pinpoint locations with a high density of crimes, aiding in spatial analysis and planning, as shown in Fig. 5.
- Analyzing Crimes Against Particular Demographics: Users can examine crime data specific to race, age group, or sex within a particular region, providing valuable insights into targeted crime patterns.
- Temporal Pattern Analysis: This use case differentiates crime occurrences between holidays and non-holidays, highlighting seasonal trends and temporal deviations.
- Crime by Hour of the Day: The day is divided into four six-hour time periods, allowing users to observe crime patterns during different times, such as morning, afternoon, evening, and night.
- Crimes by Premise Type: Users can explore crimes categorized by premises, such as residential dwellings, commercial areas, transit points, and streets.

The application seamlessly integrates the data processing and query generation with a user-friendly design. The retrieved results are displayed in an interactive and visually accessible format, ensuring users can derive actionable insights with minimal effort. Screenshots for all the queries and their results have been provided at this google drive link: https://docs.google.com/document/d/1Nq2CfT3c6Yv-6VVl6qQJsQZxC26UAMEMRdPlC9aQOsQ/edit?usp=sharing [15].

XII. Evaluation of the Application

Our application leverages insights from existing platforms, such as the NYPD and LAPD portals. Although these website primarily represents data through textual reports and dashboards, our approach integrates data from both NYPD and LAPD sources to offer a comprehensive perspective on crime trends. We do this by making the information in multiple datasets combined into one easy-to-handle platform. We make a wider analysis of criminal activities available, giving our users useful insights across many dimensions of public safety. One of the main advantages of our application is flexibility and individualization. Among other things, users can choose high-risk areas, victim demographics such as age, sex, and race, holiday patterns of crime, and the distribution of crime over time and by type of premises. That would also allow the user to get the desired level of analysis for their specific needs or regions of interest, thus making this a flexible tool in conducting a crime assessment and analysis. Our approach is future-ready, since it can be further extended to integrate more external resources for deeper insights, such as social media data or further datasets on crime patterns.

One of the main limitations of our application is that the data is not available in visual format. Unlike other dashboards, such as the one available in the NYPD's hate crime portal, our application presents the results of SPARQL queries in tabular format. Whereas this table does the job in representing data, it may not be as intuitive or interesting compared to charts, graphs, or maps. Future work will involve integrating interactive visualization.

XIII. Challenges

This project has several limitations that were developed in the course of making and implementing this project. The use cases changed because of data limitations; both of these datasets had some missing fields that were added later in order to align the datasets. Besides, non-standardized data required multiple adjustments to be performed to make the datasets compatible for analysis. Challenges also arose in aligning the existing schema with the higher-level schema.org framework, requiring iterative refinements to ensure proper integration and semantic consistency. Despite these challenges, the project successfully established a robust ontology and analysis framework for spatial and temporal crime data.

XIV. Future Work

Future work, building on the existing application, shall continue to enhance the application's usability and analytical capability. One important enhancement is to incorporate a heat map into the application to visualize the crime data on the city map. This will help the user to locate exact places of crime on a map, thereby intuitively showing the distribution of crimes. It also plans to extend its scope towards advanced data visualization, including bar graphs, pie charts, and line charts. These visual elements will enable users to analyze trends and patterns more effectively. Further development will be placed on seamlessly integrating these features into the application's user interface to meet the greater goal of providing a robust and user-friendly analytical platform.

XV. Conclusion

The application is a crucial tool in the fight against some very complex crime-related issues in both Los Angeles and New York City. Using these advanced technologies, such as the Semantic Web, ontologies, and RDF, the project transforms raw crime data into an interconnected and structured framework. It integrates comprehensive data analysis with the user-centered approach, allowing the stakeholders to gain actionable insights from crime patterns, high-risk areas, demographic trends, and temporal crime distributions across two major cities. The application successfully bridges gaps in crime data analysis by providing a unified platform for the NYPD and LAPD datasets. Its ability to support customizable queries and generate detailed tabular reports offers users significant flexibility in exploring crime trends. Our application is limited as visualizations using graphical representations and charts are limited. Further development will be towards bringing the visualizations into an interactive style, feature of updating data in real time, and involvement of other features of crime prevention as well as prediction analytics. A significant contribution of this project is the inclusion of additional fields, such as "crimeAroundHolidays" and "premiseType," which were not explored in the existing literature we reviewed. This tool marks quite great progress towards achieving informed decision-making and preventive responses in cities like New York and Los Angeles.

XVI. Roles and Responsibilities

Akash Rana & Sadanand Srinivasan: Performed comprehensive research by analyzing academic papers and relevant references to identify and validate project use cases, ensuring both theoretical robustness and practical applicability. Leveraged Ontotext Refine to align dataset columns with ontology classes and relationships. Imported data into GraphDB and developed SPARQL queries for effective data retrieval and analysis.

Rohan Mathur: Structured and defined the foundational components of the ontology, including classes, data properties, and object properties, to establish a robust and coherent framework for the knowledge graph. Authored sections of the paper focusing on high-level design, ontology development, and visualization techniques.

Sumeet Suryavanshi: Constructed the knowledge graph by refining and optimizing the ontology to enable seamless data integration, enriched semantic relationships, and enhanced query performance. Conducted data cleaning and designed both the frontend and backend of the web application.

Shreya Thummalapalli: Designed the initial ontology structure using Protégé, integrating domain-specific elements and adhering to established ontology best practices to support seamless refinement and deployment. Conducted ontology alignment with Schema.org by identifying and mapping matching classes and properties, ensuring compatibility and interoperability.

References

[1] Ang, X.J., Wee, H.L., Goh, C.Y., Zhang, Y. (2022). Space Time Analytics of New York City Shooting Incidents. In: Kundu, S.N. (eds) Geospatial Data Analytics and Urban Applications. Advances in 21st Century Human Settlements. Springer, Singapore. https://doi.org/10.1007/978-981-16-7649-9-6

[2] Singh A, Kumar S, Bansal S. NYC CrimeWatch: Crime Analysis Tool. In2024 IEEE 18th International Conference on Semantic Computing (ICSC) 2024 Feb 5 (pp. 311-316). IEEE

[3] Keyes KM, Hamilton A, Tracy M, Kagawa RM, Pear VA, Fink D, Branas CC, Cerdá M. Simulating the bounds of plausibility: Estimating the impact of high-risk versus population-based approaches to prevent firearm injury. PLoS one. 2022 Jun 2;17(6):e0269372.

[4] Sarasa-Cabezuelo A. Analysis of Gun Crimes in New York City. Sci. 2023 Apr 20;5(2):18.

[5] Wolff KT, Intravia J, Baglivio MT, Piquero AR. Violence in the Big Apple throughout the COVID-19 pandemic: A borough-specific analysis. Journal of criminal justice. 2022 Jul 1;81:101929.

[6] Brantingham PJ, Quintana-Navarrete M, Iliff C, Uchida CD, Tita GE. Situational and Victim Correlates of Increased Case Fatality Rates in Los Angeles Shootings, 2005–2021. Journal of urban health. 2024 Apr;101(2):272-9.

[7] David Paradice, D. (2017) AN ANALYSIS OF US SCHOOL SHOOTING DATA (1840-2015), Education(Vol. 138, Issue 2).

[8] CEM: an Ontology for Crime Events in Newspaper Articles Authors: Federica Rollo, Laura Po, Alessandro Castellucci

[9] Data Analyzing Mass Shooting in US, Samriddhi Shakya, Posted on Jul 20, 2016 https://nycdatascience.com/blog/student-works/r-visualization/data-analyzing-mass-shooting-us/
[10] Knowledge Representation Model for Crime Analysis Masila Abdul Jalil, Chia Pui Ling, Noor Maizura Mohamad Noor, Fatihah Mohd. School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia
[11] https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic
[12] https://catalog.data.gov/dataset/crime-data-from-2020-to-present
[13] https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Historic-/833y-fsy8/about_data
[14] https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about_data
[15] https://docs.google.com/document/d/1Nq2CfT3c6Yv-6VVl6qQJsQZxC26UAMEMRdPlC9aQOsQ/edit?usp=sharing