WEBCAT: THE DESIGN AND IMPLEMENTATION OF THE WEB-BASED CRIME ANALYSIS TOOLKIT

Charles M. Hawkins
Justin D. Pittman
Francis B. Prats
William E. Wheeler
Donald Brown
Jason Dalton
Department of Systems and Information Engineering
University of Virginia
brown@virginia.edu

Butch Johnstone
Virginia Department of Criminal Justice Services
805 East Broad Street
Richmond, VA 23219 U.S.A.
bjohnstone@dcjs.state.va.us

ABSTRACT

Based on its recent studies, the National Institute of Justice (NIJ) identifies information exchange as the most vital need of law enforcement agencies around the United States. Research by the Virginia Institute for Justice Information Systems (VIJIS) at the University of Virginia has also revealed the need for statistical analysis of criminal incident data. In order to address the need for information sharing and cross-jurisdictional analysis between different law enforcement agencies in a given region, VIJIS in conjunction with the Systems and Information Engineering (SIE) Department at the University of Virginia authorized the development of the Web Crime Analysis Toolkit (WebCAT). This application integrates an XML-driven database along with a Geographic Information System (GIS) and other analysis tools through a free, Web-based interface.

1 INTRODUCTION

Presently, most law enforcement agencies in the United States use custom-made computer software, systems, and applications to store, display, and analyze crime patterns within their jurisdictions. Due to differences in software and the tools used to develop and run these various applications some files and data from other jurisdictions are incompatible with a particular jurisdiction’s software package. Law enforcement officials also face the challenge of converting large amounts of crime data into relevant information that can be used to solve and deter crimes. Such agencies struggle to analyze criminal incident data effectively and properly allocate their resources. For larger law enforcement agencies problems arise when large amounts of data must be transmitted to analysts and decision makers over antiquated systems and unintuitive interfaces. Smaller departments have the problem of funding for software, analysis training, and infrastructure to support their analysis (Rich, 1995).

Here at the University of Virginia in conjunction with VIJIS, our capstone team has worked to address the need for a crime analysis application that facilitates efficient cross-jurisdictional analysis. In order to accomplish this goal, our team developed the following objectives:

- Provide a system that is accessible over the Internet that allows for convenient data access
- Provide a tool to spatially display criminal incident data through the use of GIS technology
- Provide tools for statistical analysis
- Provide a tool to generate summary reports of analysis
- Provide a secure web environment for the sharing of criminal incident data

WebCAT is designed to overcome shortfalls of previous crime analysis applications and provide agencies with an efficient tool for better resource allocation and crime prevention. The implementation of this system within Mountain Empire, a region in Southwestern Virginia, will help demonstrate the importance of crime analysis and data sharing between multiple jurisdictions within a given region.

2 RELATED RESEARCH

After the initial meeting with our client, the team began to research current problems faced by law enforcement agencies when performing crime analysis, as well as shortfalls of previous crime analysis applications. We used the li-
brary, Internet, and key contact people to gather this information and determine the tools WebCAT needed to be a successful crime analysis application. Jason Dalton, Program Director of VIJIS, familiarized the team with previous crime analysis applications that were designed at the University of Virginia such as ReCAP and CARV.

2.1 National Incident Based Reporting System

The Commonwealth of Virginia consists of 261 different law enforcement jurisdictions. Each jurisdiction collects a vast amount of criminal incident data and stores it into their database. Unfortunately, police officers only have direct access to the data collected by their jurisdiction, yet cannot access data from other jurisdictions that may be relevant to their investigations. The only statewide sharing of this data is through the National Incident Based Reporting System (NIBRS), in which the incident data collected is converted to the NIBRS format, stored on a collection of floppy discs, and mailed to the state NIBRS center monthly. One problem with the NIBRS format is that individual crimes do not contain information on their specific location within the jurisdiction where they occurred. Therefore, the crime data analysis done by a law enforcement agent is restricted to data on incidents that have occurred in his or her jurisdiction. Any cross-jurisdictional patterns might be lost and officers in either jurisdiction may not detect these patterns (VGIN, 2001).

2.2 National Institute of Justice InfoTech Specification

Recognizing this problem when performing cross-jurisdictional analysis, the National Institute of Justice created a program called InfoTech. This program gives agencies access to multiple data sources through a web-based query. Currently, agencies in Florida, California, and Oregon use this system. In Oregon, the program provides a means of information sharing between the Department of Corrections, the state police laboratory, and the judicial department. More specifically, InfoTech allows correctional facilities and other agencies to share inmate profiles. By doing so, it collects valuable information on an inmate’s medical and criminal history. This information is useful if the inmate is reincarcerated in another facility, or released. Another advantage of the InfoTech program involves the ability of individual users to tailor the interface to their specific needs (Baker, 2000). The InfoTech program offers law enforcement agencies a flexible tool that provides them with a way to access and analyze criminal data from other jurisdictions.

2.3 ReCAP and CARV

The University of Virginia’s Systems Engineering Department under the direction of VIJIS and Professor Donald Brown has begun to address the need for information sharing through the development of various crime analysis programs such as ReCAP and CARV. VIJIS developed ReCAP, a free crime analysis tool that integrates a database management system with GIS capabilities and other analytical tools. One disadvantage of the ReCAP design is that it is not web-based and thus requires installation within each jurisdiction (Francis, 2002).

In 2000, a capstone team developed CARV, a crime analysis tool that gives users access to criminal incident data and allows them to perform various methods of statistical analysis. This system runs on a Citrix Metaframe, a server that has the ability to run applications over any connection. However, one drawback of CARV is that the server contains only 25 ports, which means only a maximum of 25 users can have access to the system at a given time (O’Dell, 2001).

3 WEBCAT DESIGN

After initial research was complete, our capstone team took a systems engineering approach in solving the problem. First, we established the goals and objective of the system. Next, the team established a set of metrics in order to determine the most viable technologies that could be used within WebCAT. Once the desired technologies were chosen, implementation of the system began.

3.1 Metrics

In order to determine what technologies to use when designing and implementing WebCAT, our team formulated a number of metrics that were used to evaluate the performance of the technologies being considered and their feasibility to work together within WebCAT. The metrics for this project are listed below.

- **Cost** – Cost represents the monetary value needed for each goal. The main goal of cost is to minimize the amount of money needed to implement the system in order to ensure a free product for the client. Cost will be measured in dollars.

- **Ease of implementation** – Represents the feasibility of the group completing a complex problem in the time-frame allocated. This metric will be measured by the amount of time it takes to code and implement the given component of the system.

- **Ease of Use** – The group wants to ensure that all users, no matter what their level of expertise, can manipulate and understand the product. This metric will be measured by performing usability tests with people that have various levels of experience using such products.
• **Speed** – Measures the relative speed of the software versus a standard set via surveying and experimentation. For this metric, we will run tests on transfer times to see how long it takes to load and spatially display criminal incident data. We will compare these results with feedback from the VCAN survey responses we receive.

• **Accuracy** – The group wants to ensure that all returned queries and calculations are free of error and accurate while maintaining efficiency. This metric will be measured by testing the number of valid queries or calculations that fail versus the number of invalid queries or calculations that pass.

• **Flexibility** – The group would like to provide components of the system that can be easily integrated with the product as a whole. This will be measured using past research on the various products that the team is trying to implement, and the time it takes to integrate the various components with one another.

### 3.2 Alternative Technologies

Last semester, our team began to research available technologies for the various components of WebCAT. In order to determine the most viable technologies, the team closely studied the capabilities of each technology and determined whether or not it met the metrics described in the previous section.

#### 3.2.1 Extensible Markup Language

WebCAT does not use a typical relational database to store criminal incident data. A relational database is a collection of data items organized as a set of formally-described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables (Lozano, 2001). Instead, this application makes use of the Extensible Markup Language (XML) to simulate the presence of a database. The market for XML storage of data has and will continue to increase dramatically over the next few years. According to the analyst firm ZapThink, the market for XML storage will grow from $75 million in 2000 to over $4 billion by 2005 (Trippe, 2002). It is necessary that current and future applications make use of XML technology due to its increased use just over the last few years.

There are many important distinctions between XML and relational data. Relational data is about rows and columns of previously defined chunks of information that is very well structured within a table. While XML data is somewhat structured, there are two fundamental differences between it and relational data. First, XML can embed hierarchies of parent-child relationships in ways that relational data cannot. Second, XML does not care how long or complicated a given field is. XML also offers a very flexible way to manage data. Users can define their own tags according to their specific needs (Trippe, 2002). In WebCAT’s case, our team created the following XML structure for criminal incident files:

```xml
<?xml version="1.0" encoding="ISO-8859-1" ?>
<crimes>
  <crime id="1">
    <cid>Crime ID</cid>
    <type>Crime Type</type>
    <address>
      <street>Street</street>
      <city>City</city>
      <state>State</state>
      <zip>Zip</zip>
    </address>
    <criminal>
      <c_firstname>Criminal's First Name</c_firstname>
      <c_lastname>Criminal's Last Name</c_lastname>
      <c_age>Criminal's Age</c_age>
      <c_gender>Criminal's Sex</c_gender>
      <c_race>Criminal's Race</c_race>
    </criminal>
    <victim>
      <v_firstname>Victim's First Name</v_firstname>
      <v_lastname>Victim's Last Name</v_lastname>
      <v_age>Victim's Age</v_age>
      <v_gender>Victim's Sex</v_gender>
      <v_race>Victim's Race</v_race>
    </victim>
    <date>mm-dd-yyyy</date>
    <time>0000</time>
    <longitude>99.999999</longitude>
    <latitude>99.999999</latitude>
  </crime>
</crimes>
```

![Figure 1: WebCAT's XML Format](image)

#### 3.2.2 Geographic Information System

Last year, the WebCAT team implemented the Geographic Information System (GIS) component using ESRI’s ArcIMS. This year’s team decided that using ArcIMS for the GIS module conflicted with the refined goals of the project. While the software is available for use by the University of Virginia, ArcIMS software license does not allow the team to distribute the technology for use by third parties. As a result, every law enforcement agency using WebCAT would have to purchase a license agreement for ArcIMS. Such conditions conflict with the team’s goal of providing the toolkit to agencies at a low cost. As a result, the team felt compelled to remove ArcIMS from the application and select different GIS software.
Our team researched various GIS applications during the fall. As a result of research, the team chose to use MapServer, an open-source GIS application that was developed by the University of Minnesota. MapServer was chosen mainly due to the fact that it was free and thus helped to fulfill the team’s goal of providing an inexpensive crime analysis application to law enforcement agencies.

3.2.3 Scalable Vector Graphics

We chose to use SVG to design the control and time charts for WebCAT. SVG is a type of Extensible Markup Language (XML) that allows for straightforward and efficient two-dimensional graphics on the web. With SVG, plain text is used to create graphics features that would usually be associated with bitmap images (jpg, gif). This greatly decreases processing and downloading time. SVG also works well across all platforms, output resolutions, color spaces, ranges of available bandwidths, and memory capacity. Finally, scripts can be embedded into SVG allowing for dynamic graphics (Sall, 1999).

4 OVERVIEW OF SYSTEM

WebCAT allows users to upload or create their own data files. Users can then use their file with all of the analysis tools within WebCAT. WebCAT offers crime analysts several temporal visualization tools for analyzing criminal data such as the time chart and control chart. Users can also spatially display and cluster criminal incident data using the GIS component. After performing analysis, a user can generate summary reports. Users also have the option to query the NIBRS database for crime information from across the state of Virginia.

4.1 Administration and Security Controls

WebCAT must recognize unique users and their system privileges in order to maintain a secure application. The motivation here is to restrict the access to the application to registered users. Upon entering the uniform resource locator (URL) to access WebCAT, users are directed to the Login Module, which prompts them for their system user name and password.

The design of the Management Module includes a graphical user interface for setting the general and specific parameters involved with the user accounts domain. Such functionality is conducive to managing who uses the interface and in what capacity. Tracking such information lends itself to improving the interface and keeping information secure. Also, administrators can maintain accountability as far as who is responsible for sending and receiving data shared with other jurisdictions. The application may support different privilege levels defined by the system administrator. The main purpose of the Management Module is to offer agencies control over the interaction between their analysts and the application.

4.2 Data Import

After logging in to the system, users can search the client computer for specific crime files to be used by WebCAT. Currently, users can only upload text files. These files must be either comma, space, or semicolon delimited. Another option is to simply select the fields with which the analyst would like to work. After submitting these types, the user may select the order in which he would like the fields to appear and either manually enter the data or simply cut and paste from another file. Once a user uploads or creates a file, it is then converted into WebCAT’s XML format.

4.3 Querying Component

After WebCAT converts an uploaded dataset to XML, the user has the option to narrow the dataset via querying. Using an HTML form:

- The user is prompted for which data fields they would like to enter values
- The user is prompted for which data fields they would like returned
- A file is written to extract data from the XML document
- The user is redirected to a page tabulating the query results
- The user may save the results to his or her user directory, or start a different query

The team is still optimizing the querying procedure, and is close to finishing it. For applications tailored to agencies in Virginia, WebCAT’s Querying Module will also support queries to the Virginia Incident-Based Reporting Database. The database serves as a state repository for criminal incident data.

4.4 Statistical Analysis

There are two types of control charts available in WebCAT. The yearly control chart displays the number of crimes per week for each of the past 52 weeks for one or more years. In the underlying code, a mean and standard deviation are computed based on the values of the underlying data set. From these figures, an upper control limit and a lower control limit are calculated. It allows users to spot crime trends over the year. Weeks that are “out of control,” in which the number of criminal incidents is higher than three standard deviations above the mean, are displayed below the graph under “Significantly High Weeks:”
Figure 2: Weekly Control Chart

A second type of control chart allows for a 90-day analysis, showing the days out of the most recent three month period where crime was out of control. Research indicates that the identical control chart feature in the ReCAP system, when used on an everyday basis, was able to alert police departments and improve the reassignment times in some cases by more than two months (Draim, 2001).

The time chart allows users to quickly view at what times along a 24 hour period a certain crime set is “out of control.” The crimes are separated into bins of 10 minutes in length. Just as in the control chart, a mean and a value one standard deviation greater are displayed in order to give an analyst a sense of when in the day more crimes are occurring than normal. In studying the time chart below, any 10 minute intervals where more than 5.76 crimes have occurred are deemed to be out of control. Based on the graph, a user could conclude that more crimes are happening late night and early morning between about 9:00 pm. and 1:00 am. The analysts could then track this trend in future periods and if it holds true make it known to superiors who might then take steps to allocate resources differently, for example adding extra patrol cars at these times.

4.5 Spatial Analysis

WebCAT provides analysts with a tool to spatially display criminal incident data in order to visualize crime patterns. Users can upload crime data sets that include geographic information, such as latitude and longitude. The GIS system then plots these points by extracting these coordinates and reading them into a map file (.map). The map file contains the various layers of the GIS, such as roads, waterways, and other boundaries. The GIS allows the user to zoom in or out, select which layers are displayed, and view a layer of crime points and a cluster image for the purpose of crime analysis.

One of the most useful procedures available in WebCAT’s GIS component involves the clustering of criminal incident data using a Kernel Density Estimation. Essentially, the KDE examines every pixel on the map and determines its aggregated distance to every incident point on the map. After calculating the distance to every point, the algorithm compares the number to a normal Gaussian distribution, essentially determining the probability of a crime occurring on that point in relation to the mapped crime points. The KDE would then assign shades of a color to each value of the Gaussian distribution returned. Darker colors represent areas with higher concentrations of crimes. The figure below shows WebCAT’s GIS interface displaying a map of Fairfax County, Virginia along with a layer of clustered crime points.

Figure 4: WebCAT’s GIS Interface with Clustering Algorithm

4.6 Crime Reports

After performing the desired analysis, analysts can generate a series of crime reports. The report module provides the user with a way to get information out of the system. The report module makes available to the user:
obtain all robberies that occurred in Tazewell County in Virginia in the past five years, they must contact the Virginia Department of Criminal Justice Services (DCJS). DCJS would then query the NIBRS database and then send the results to the agency via email. According to Deborah Roberts, a program analyst at DCJS, this request may take as long as 20 minutes. However, WebCAT offers a built-in tool that allows agencies to query the NIBRS database from their own office. Using WebCAT, agencies could obtain all robberies that occurred in Tazewell County in less than 30 seconds.

WebCAT generates some of the reports automatically for the user, while other reports provide some level of customization through the solicitation of user inputs. Generally, much of the files available for synthesis in a report are stored in the user’s personal directory within WebCAT over the course of the user’s web session.

5 RESULTS

Currently, many GIS systems can cost thousands of dollars to purchase and maintain. This high cost can be prohibitive to smaller departments, which already lack the necessary funds and resources to purchase such applications. However, WebCAT provides a GIS component for spatial analysis as well as other analysis tools at minimal cost. All of the software used in WebCAT is available for free so no licensing and purchasing costs will be incurred. Minimal costs may be involved in purchasing the necessary software to set up a server within an agency.

One of the main advantages of WebCAT over other crime analysis applications is the fact that it is Web-based. Other crime analysis applications such as ReCAP must be installed on every computer within a given jurisdiction. Installation and configuring time on each computer may take as long as eight hours. If a given agency has as little as ten computers, total installation time for all of the computers would still take at least 80 hours. However, since WebCAT can be accessed from the Internet, the system only needs to be installed on a single server instead of multiple computers. While installation time for WebCAT will also take approximately eight hours, a given agency will be able to access and use the system after only eight hours. This is a dramatic reduction in implementation time and thus will allow agencies to perform crime analysis shortly after installing WebCAT.

WebCAT also provides faster analysis than previous applications. For example, if a given agency would like to obtain all robberies that occurred in Tazewell County, Virginia in the past five years, they must contact the Virginia Department of Criminal Justice Services (DCJS). DCJS would then query the NIBRS database and then send the results to the agency via email. According to Deborah Roberts, a program analyst at DCJS, this request may take as long as 20 minutes. However, WebCAT offers a built-in tool that allows agencies to query the NIBRS database from their own office. Using WebCAT, agencies could obtain all robberies that occurred in Tazewell County in less than 30 seconds.

As a result of this project the group has met its goal of providing law enforcement agencies with a low cost and efficient crime analysis application that will facilitate crime pattern detection and improved resource allocation.

6 CONCLUSION

WebCAT will be useful to crime analysts across the state of Virginia in several ways. First, it will provide smaller police departments on the county level a source of free crime analysis and incident pattern recognition. Second, it will allow larger jurisdictions to efficiently handle large amounts of crime incident data through the querying tools and use the analytical tools to supplement their analysis methods. Third, it can be used by officials and policy makers on the state level to work with the NIBRS system and perform relevant analysis for cross-jurisdictional crime incident data. Finally, it will hopefully provide a means for crime analysts across Virginia to begin to work together to improve the way crime analysis is performed. Such a tool can facilitate the drive toward a more universal reporting standard, sensible ways of storing crime data, and the development of further tools to make crime analysis easier and more cost effective. Hopefully, the use of WebCAT in the Mountain Empire Region of Virginia will demonstrate its effectiveness in crime prevention and promote its use throughout the entire state.

REFERENCES


AUTHOR BIOGRAPHIES

CHARLES M. HAWKINS is a fourth year Systems and Information Engineering (SIE) student at the University of Virginia from New Orleans, Louisiana. He can be contacted at <cmh7v@virginia.edu>

JUSTIN D. PITTMAN is a fourth year SIE student at the University of Virginia from Richmond, Virginia. He is still in the process of interviewing for a job after graduation. You can reach him by email at <jdp3k@virginia.edu>

FRANCIS B. PRATS is a fourth year SIE student at the University of Virginia. He has been accepted for graduate studies at the University of Virginia through the SIE Masters Exploratory Research Program. He can be reached at <prats@virginia.edu>

WILLIAM E. WHEELER is a fourth year SIE student at the University of Virginia from Philadelphia, Pennsylvania. He is still deciding whether to enter the work force or attend graduate school upon graduation. He can be reach at <wew4d@virginia.edu>